METHODS AND COMPOSITIONS FOR RNA-DIRECTED TARGET DNA MODIFICATION AND FOR RNA-DIRECTED MODULATION OF TRANSCRIPTION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application Nos. 61/652,086 filed May 25, 2012, 61/716,256 filed October 19, 2012, 61/757,640 filed January 28, 2013, and 61/765,576, filed February 15, 2013, each of which applications is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant No. GM081879 awarded by the National Institutes of Health. The government has certain rights in the invention.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING PROVIDED AS A TEXT FILE

[0003] A Sequence Listing is provided herewith as a text file, "BERK-187WO-SeqList_ST25.txt" created on March 13, 2013 and having a size of 7645 KB. The contents of the text file are incorporated by reference herein in their entirety.

BACKGROUND

[0004] About 60% of bacteria and 90% of archaea possess CRISPR (clustered regularly interspaced short palindromic repeats)/CRISPR-associated (Cas) system systems to confer resistance to foreign DNA elements. Type II CRISPR system from *Streptococcus pyogenes* involves only a single gene encoding the Cas9 protein and two RNAs - a mature CRISPR RNA (crRNA) and a partially complementary trans-acting RNA (tracrRNA) - which are necessary and sufficient for RNA-guided silencing of foreign DNAs.

[0005] In recent years, engineered nuclease enzymes designed to target specific DNA sequences have attracted considerable attention as powerful tools for the genetic manipulation of cells and whole organisms, allowing targeted gene deletion, replacement and repair, as well as the insertion of exogenous sequences (transgenes) into the genome. Two major technologies for engineering site-specific DNA nucleases have emerged, both of which are based on the construction of chimeric endonuclease enzymes in which a sequence non-specific DNA endonuclease domain is fused to an engineered DNA binding domain. However, targeting each new genomic locus requires the design of a novel nuclease enzyme, making these approaches both time consuming

and costly. In addition, both technologies suffer from limited precision, which can lead to unpredictable off-target effects.

[0006] The systematic interrogation of genomes and genetic reprogramming of cells involves targeting sets of genes for expression or repression. Currently the most common approach for targeting arbitrary genes for regulation is to use RNA interference (RNAi). This approach has limitations. For example, RNAi can exhibit significant off-target effects and toxicity.

[0007] There is need in the field for a technology that allows precise targeting of nuclease activity (or other protein activities) to distinct locations within a target DNA in a manner that does not require the design of a new protein for each new target sequence. In addition, there is a need in the art for methods of controlling gene expression with minimal off-target effects.

SUMMARY

[0008] The present disclosure provides a DNA-targeting RNA that comprises a targeting sequence and, together with a modifying polypeptide, provides for site-specific modification of a target DNA and/or a polypeptide associated with the target DNA. The present disclosure further provides site-specific modifying polypeptides. The present disclosure further provides methods of site-specific modification of a target DNA and/or a polypeptide associated with the target DNA The present disclosure provides methods of modulating transcription of a target nucleic acid in a target cell, generally involving contacting the target nucleic acid with an enzymatically inactive Cas9 polypeptide and a DNA-targeting RNA. Kits and compositions for carrying out the methods are also provided. The present disclosure provides genetically modified cells that produce Cas9; and Cas9 transgenic non-human multicellular organisms.

FEATURES

[0009] Features of the present disclosure include a DNA-targeting RNA comprising: (i) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) a second segment that interacts with a site-directed modifying polypeptide. In some cases, the first segment comprises 8 nucleotides that have 100% complementarity to a sequence in the target DNA. In some cases, the second segment comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-682 (e.g., 431-562). In some cases, the second segment comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:563-682. In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1

amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

[0010] Features of the present disclosure include a DNA polynucleotide comprising a nucleotide sequence that encodes the DNA-targeting RNA. In some cases, a recombinant expression vector comprises the DNA polynucleotide. In some cases, the nucleotide sequence encoding the DNA-targeting RNA is operably linked to a promoter. In some cases, the promoter is an inducible promoter. In some cases, the nucleotide sequence encoding the DNA-targeting RNA further comprises a multiple cloning site. Features of the present disclosure include an *in vitro* genetically modified host cell comprising the DNA polynucleotide.

[0011] Features of the present disclosure include a recombinant expression vector comprising: (i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a nucleotide sequence encoding the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[0012] Features of the present disclosure include a recombinant expression vector comprising: (i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a nucleotide sequence encoding the site-directed modifying polypeptide, where the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[0013] Features of the present disclosure include a variant site-directed modifying polypeptide comprising: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that exhibits reduced site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some cases, the variant site-directed modifying polypeptide comprises an H840A mutation of the S. pyogenes sequence SEQ ID NO:8 or the corresponding mutation in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the variant site-directed modifying polypeptide comprises a D10A mutation of the S. pyogenes sequence SEQ ID NO:8

or the corresponding mutation in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases,the variant site-directed modifying polypeptide comprises both (i) a D10A mutation of the S. pyogenes sequence SEQ ID NO:8 or the corresponding mutation in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346; and (ii) an H840A mutation of the S. pyogenes sequence SEQ ID NO:8 or the corresponding mutation in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

[0014] Features of the present disclosure include a chimeric site-directed modifying polypeptide comprising: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some cases, the chimeric site-directed modifying polypeptide of comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the DNA-targeting RNA further comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-682 (e.g., SEQ ID NOs:563-682). In some cases, the DNA-targeting RNA further comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-562. In some cases,the enzymatic activity of the chimeric site-directed modifying polypeptide modifies the target DNA. In some cases, the enzymatic activity of the chimeric site-directed modifying polypeptide is nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity. In some cases, the enzymatic activity of the chimeric site-directed modifying polypeptide is nuclease activity. In some cases, the nuclease activity introduces a double strand break in the target DNA. In some cases, the enzymatic activity of the chimeric site-directed modifying polypeptide modifies a target polypeptide associated with the target DNA. In some cases, the enzymatic activity of the chimeric site-directed modifying polypeptide is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity,

SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity.

[0015] Features of the present disclosure include a polynucleotide comprising a nucleotide sequence encoding a chimeric site-directed modifying polypeptide. In some cases, the polynucleotide is an RNA polynucleotide. In some cases, the polynucleotide is a DNA polynucleotide. Features of the present disclosure include a recombinant expression vector comprising the polynucleotide. In some cases, the polynucleotide is operably linked to a promoter. In some cases, the promoter is an inducible promoter. Features of the present disclosure include an *in vitro* genetically modified host cell comprising the polynucleotide.

[0016] Features of the present disclosure include a chimeric site-directed modifying polypeptide comprising: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA. In some cases, the activity portion increases transcription within the target DNA. In some cases, the activity portion decreases transcription within the target DNA.

[0017] Features of the present disclosure include a genetically modified cell comprising a recombinant site-directed modifying polypeptide comprising an RNA-binding portion that interacts with a DNA-targeting RNA; and an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell.

[0018] Features of the present disclosure include a transgenic non-human organism whose genome comprises a transgene comprising a nucleotide sequence encoding a recombinant site-directed modifying polypeptide comprising: (i) an RNA-binding portion that interacts with a DNA-targeting RNA; and (ii) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%

amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the organism is selected from the group consisting of: an archaea, a bacterium, a eukaryotic single-cell organism, an algae, a plant, an animal, an invertebrate, a fly, a worm, a cnidarian, a vertebrate, a fish, a frog, a bird, a mammal, an ungulate, a rodent, a rat, a mouse, and a non-human primate.

[0019] Features of the present disclosure include a composition comprising: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some cases, the first segment of the DNA-targeting RNA comprises 8 nucleotides that have at least 100% complementarity to a sequence in the target DNA. In some cases, the second segment of the DNA-targeting RNA comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-682 (e.g., SEQ ID NOs:563-682). In some cases, the second segment of the DNA-targeting RNA comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-562. In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the enzymatic activity modifies the target DNA. In some cases, the enzymatic activity is nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity. In some cases, the enzymatic activity is nuclease activity. In some cases, the nuclease activity introduces a double strand break in the target DNA. In some cases, the enzymatic activity modifies a target polypeptide associated with the target DNA. In some cases, the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin

ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity. In some cases, the target polypeptide is a histone and the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity or deubiquitinating activity. In some cases, the DNA-targeting RNA is a double-molecule DNA-targeting RNA and the composition comprises both a targeter-RNA and an activator-RNA, the duplex-forming segments of which are complementary and hybridize to form the second segment of the DNA-targeting RNA. In some cases, the duplex-forming segment of the activator-RNA comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NO:SEQ ID NOs:431-682.

[0020] Features of the present disclosure include a composition comprising: (i) a DNA-targeting RNA of the present disclosure, or a DNA polynucleotide encoding the same; and (ii) a buffer for stabilizing nucleic acids. Features of the present disclosure include a composition comprising: (i) a site-directed modifying polypeptide of the present disclosure, or a polynucleotide encoding the same; and (ii) a buffer for stabilizing nucleic acids and/or proteins. Features of the present disclosure include a composition comprising: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA. In some cases, the activity portion increases transcription within the target DNA. In some cases, the activity portion decreases transcription within the target DNA. Features of the present disclosure include a composition comprising: (i) a site-directed modifying polypeptide, or a polynucleotide encoding the same; and (ii) a buffer for stabilizing nucleic acids and/or proteins.

[0021] Features of the present disclosure include a method of site-specific modification of a target DNA, the method comprising: contacting the target DNA with: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity. In some cases, the target DNA is extrachromosomal. In some cases, the target DNA comprises a PAM sequence of the complementary strand that is 5'-CCY-3', wherein Y is any DNA nucleotide and Y is immediately 5' of the target sequence of the complementary strand of the target DNA. In some cases, the target DNA is part of a chromosome *in vitro.* In some cases, the target DNA is part of a chromosome *in vivo.* In some cases, the target DNA is part of a chromosome in a cell. In some cases, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell. In some cases, the DNA-targeting RNA comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-682 (e.g., SEQ ID NOs:563-682). In some cases, the DNA-targeting RNA comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth SEQ ID NOs:431-562. In some cases, the DNA-modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the enzymatic activity modifies the target DNA. In some cases, the enzymatic activity is nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity. In some cases, the DNA-modifying enzymatic activity is nuclease activity. In some cases, the nuclease activity introduces a double strand break in the target DNA. In some cases, the contacting occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair. In some cases, the method further comprises contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA. In some cases, the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that

nucleotides within the target DNA are deleted. In some cases, the enzymatic activity modifies a target polypeptide associated with the target DNA. In some cases, the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity. In some cases, the target polypeptide is a histone and the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity or deubiquitinating activity. In some cases, the complex further comprises an activator-RNA. In some cases, the activator-RNA comprises a nucleotide sequence with at least 60% identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-682.

[0022] Features of the present disclosure include a method of modulating site-specific transcription within a target DNA, the method comprising contacting the target DNA with: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription, wherein said contacting results in modulating transcription within the target DNA. In some cases, transcription within the target DNA is increased. In some cases, transcription within the target DNA is decreased.

[0023] Features of the present disclosure include a method of site-specific modification at target DNA, the method comprising: contacting the target DNA with: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA. In some cases, the site-directed modifying polypeptide increases transcription within the target DNA. In some cases, the site-directed modifying polypeptide decreases transcription within the target DNA.

[0024] Features of the present disclosure include a method of promoting site-specific cleavage and modification of a target DNA in a cell, the method comprising introducing into the cell: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits nuclease activity that creates a double strand break in the target DNA; wherein the site of the double strand break is determined by the DNA-targeting RNA, the contacting occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair, and the target DNA is cleaved and rejoined to produce a modified DNA sequence. In some cases, the method further comprises contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA. In some cases, the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted. In some cases, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell. In some cases, the cell is *in vitro*. In some cases, the cell is *in vivo*.

[0025] Features of the present disclosure include a method of producing a genetically modified cell in a subject, the method comprising: (I) introducing into a cell: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits nuclease activity that creates a double strand break in the target DNA; wherein the site of the double strand break is determined by the DNA-targeting RNA, the contacting occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair, and the target DNA is cleaved and rejoined to produce a modified DNA sequence; thereby producing the genetically modified cell;

and (II) transplanting the genetically modified cell into the subject. In some cases, the method further comprises contacting the cell with a donor polynucleotide, wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA. In some cases, the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted. In some cases, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, an amphibian cell, a bird cell, a mammalian cell, an ungulate cell, a rodent cell, a non-human primate cell, and a human cell.

[0026] Features of the present disclosure include a method of modifying target DNA in a genetically modified cell that comprises a nucleotide sequence encoding an exogenous site-directed modifying polypeptide, the method comprising introducing into the genetically modified cell a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein: (i) the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits nuclease activity. In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, an amphibian cell, a bird cell, a mammalian cell, an ungulate cell, a rodent cell, a non-human primate cell, and a human cell. In some cases, the cell is in vivo. In some cases, the cell is in vitro. In some cases, the expression of the site-directed modifying polypeptide is under the control of an inducible promoter. In some cases, the expression of the site-directed modifying polypeptide is under the control of a cell type-specific promoter.

[0027] Features of the present disclosure include a kit comprising: the DNA-targeting RNA, or a DNA polynucleotide encoding the same; and a reagent for reconstitution and/or dilution. In some cases, the kit further comprises a reagent selected from the group consisting of: a buffer for introducing into cells the DNA-targeting RNA, a wash buffer, a control reagent, a control

expression vector or RNA polynucleotide, a reagent for transcribing the DNA-targeting RNA from DNA, and combinations thereof.

[0028] Features of the present disclosure include a kit comprising: a site-directed modifying polypeptide of the present disclosure, or a polynucleotide encoding the same; and a reagent for reconstitution and/or dilution. In some cases, the kit further comprises a reagent selected from the group consisting of: a buffer for introducing into cells the site-directed modifying polypeptide, a wash buffer, a control reagent, a control expression vector or RNA polynucleotide, a reagent for *in vitro* production of the site-directed modifying polypeptide from DNA, and combinations thereof.

[0029] Features of the present disclosure include a kit comprising: a site-directed modifying polypeptide of the present disclosure, or a polynucleotide encoding the same; and a reagent for reconstitution and/or dilution. Features of the present disclosure include a kit comprising: a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[0030] Features of the present disclosure include a kit comprising: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[0031] Features of the present disclosure include a kit comprising: (i) any of the recombinant expression vectors above; and (ii) a reagent for reconstitution and/or dilution. Features of the present disclosure include a kit comprising: (i) any of the recombinant expression vectors above; and (ii) a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined

by the DNA-targeting RNA. Features of the present disclosure include a kit comprising: (i) any of the recombinant expression vectors above; and (ii) a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[0032] Features of the present disclosure include a kit for targeting target DNA comprising: two or more DNA-targeting RNAs, or DNA polynucleotides encoding the same, wherein the first segment of at least one of the two or more DNA-targeting RNAs differs by at least one nucleotide from the first segment of at least one other of the two or more DNA-targeting RNAs.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] **Figures 1A-B** provide a schematic drawing of two exemplary subject DNA-targeting RNAs, each associated with a site-directed modifying polypeptide and with a target DNA.

[0034] **Figure 2** depicts target DNA editing through double-stranded DNA breaks introduced using a Cas9/Csn1 site-directed modifying polypeptide and a DNA-targeting RNA.

[0035] **Figures 3A-B** depict the amino acid sequence of a Cas9/Csn1 protein from *Streptococcus pyogenes* (SEQ ID NO:8). Cas9 has domains homologous to both HNH and RuvC endonucleases. (**A**) Motifs 1-4 are overlined (**B**) Domains 1 and 2 are overlined.

[0036] **Figures 4A-B** depict the percent identity between the Cas9/Csn1 proteins from multiple species. (A) Sequence identity relative to *Streptococcus pyogenes.* For Example, Domain 1 is amino acids 7-166 and Domain 2 is amino acids 731-1003 of Cas9/Csn1 from *Streptococcus pyogenes* as depicted in Figure 3B. (B) Sequence identity relative to *Neisseria meningitidis*. For example, Domain 1 is amino acids 13-139 and Domain 2 is amino acids 475-750 of Cas9/Csn1 from *Neisseria meningitidis* (SEQ ID NO:79).

[0037] **Figure 5** depicts a multiple sequence alignment of motifs 1-4 of Cas9/Csn1 proteins from various diverse species slected from the phylogenetic table in Figure 32 (see Figure 32, Figure 3A, and Table 1) (*Streptococcus pyogenes* (SEQ ID NO:8), *Legionella pneumophila* (SEQ ID NO:17), *Gamma proteobacterium* (SEQ ID NO:107), *Listeria innocua* (SEQ ID NO:3), *Lactobacillus gasseri* (SEQ ID NO:152), *Eubacterium rectale* (SEQ ID NO:99), *Staphylococcus lugdunensis* (SEQ ID NO:185), *Mycoplasma synoviae* (SEQ ID NO:22), *Mycoplasma mobile* (SEQ ID NO:16), *Wolinella succinogenes* (SEQ ID NO:10), *Flavobacterium columnare* (SEQ ID NO:235), *Fibrobacter succinogenes* (SEQ ID NO:121), *Bacteroides fragilis* (SEQ ID

NO:21), *Acidothermus cellulolyticus* (SEQ ID NO:42), and *Bifidobacterium dentium* (SEQ ID NO:131).

[0038] **Figures 6A-B** provide alignments of naturally occurring tracrRNA ("activator-RNA") sequences from various species (L. innocua (SEQ ID NO:268); S. pyogenes (SEQ ID NO:267); S. mutans (SEQ ID NO:269); S. thermophilus1 (SEQ ID NO:270); M. mobile (SEQ ID NO:274); N. meningitides (SEQ ID NO:272); P. multocida (SEQ ID NO:273); S. thermophilus2 (SEQ ID NO:271); and S. pyogenes (SEQ ID NO:267). (**A**) multiple sequence alignment of selected tracrRNA orthologues (AlignX, VectorNTI package, Invitrogen) associated with CRISPR/Cas loci of similar architecture and highly similar Cas9/Csn1 sequences. Black boxes represent shared nucleotides (**B**) multiple sequence alignment of selected tracrRNA orthologues (AlignX, VectorNTI package, Invitrogen) associated with CRISPR/Cas loci of different architecture and non-closely related Cas9/Csn1 sequences. Note the sequence similarity of N. meningitidis and P. multocida tracrRNA orthologues. Black boxes represent shared nucleotides. For more exemplary activator-RNA sequences, see SEQ ID NOs:431-562.

[0039] **Figures 7A-B** provide alignments of naturally occurring duplex-forming segments of crRNA ("targeter-RNA") sequences from various species (L. innocua (SEQ ID NO://); S. pyogenes (SEQ ID NO://); S. mutans (SEQ ID NO://); S. thermophilus1 (SEQ ID NO://); C. jejuni (SEQ ID NO://); S. pyogenes (SEQ ID NO://); F. novicida (SEQ ID NO://); M. mobile (SEQ ID NO://); N. meningitides (SEQ ID NO://); P. multocida (SEQ ID NO://); and S. thermophilus2 (SEQ ID NO://). (**A**) multiple sequence alignments of exemplary duplex-forming segment of targeter-RNA sequences (AlignX, VectorNTI package, Invitrogen) associated with the loci of similar architecture and highly similar Cas9/Csn1 sequences. (**B**) multiple sequence alignments of exemplary duplex-forming segment of targeter-RNA sequences (AlignX, VectorNTI package, Invitrogen) associated with the loci of different architecture and diverse Cas9 sequences. Black boxes represent shared nucleotides. For more exemplary duplex-forming segments targeter-RNA sequences, see SEQ ID NOs:563-679.

[0040] **Figure 8** provides a schematic of hybridization for naturally occurring duplex-forming segments of the crRNA ("targeter-RNA") with the duplex-forming segment of the corresponding tracrRNA orthologue ("activator-RNA"). Upper sequence, targeter-RNA; lower sequence, duplex-forming segment of the corresponding activator-RNA. The CRISPR loci belong to the Type II (Nmeni/CASS4) CRISPR/Cas system. Nomenclature is according to the CRISPR database (CRISPR DB). S. pyogenes (SEQ ID NO:// and //); S. mutans (SEQ ID NO:// and //); S. thermophilus1 (SEQ ID NO:// and //); S. thermophilus2 (SEQ ID NO:// and //); L. innocua (SEQ ID NO:// and //); T. denticola (SEQ ID NO:// and //); N. meningitides (SEQ ID NO:// and //); S. gordonii (SEQ ID NO:// and //); B. bifidum (SEQ ID NO:// and //); L. salivarius (SEQ ID NO://

and //); F. tularensis (SEQ ID NO:// and //); and L. pneumophila (SEQ ID NO:// and //). Note that some species contain each two Type II CRISPR loci. For more exemplary activator-RNA sequences, see SEQ ID NOs:431-562. For more exemplary duplex-forming segments targeter-RNA sequences, see SEQ ID NOs:563-679.

[0041] **Figure 9** depicts example tracrRNA (activator-RNA) and crRNA (targeter-RNA) sequences from two species. A degree of interchangeability exists; for example, the *S.pyogenes* Cas9/Csn1 protein is functional with tracrRNA and crRNA derived from *L.innocua*. (|) denotes a canonical Watson-Crick base pair while (•) denotes a G-U wobble base pair. "Variable 20nt" or "20nt" represents the DNA-targeting segment that is complementary to a target DNA (this region can be up to about 100nt in length). Also shown is the design of single-molecule DNA-targeting RNA that incorporates features of the targeter-RNA and the activator-RNA. (Cas9/Csn1 protein sequences from a wide variety of species are depicted in Figure 3 and set forth as SEQ ID NOs:1-256 and 795-1346) *Streptococcus pyogenes*: top to bottom: (SEQ ID NO://, //, //); *Listeria innocua:* top to bottom: (SEQ ID NO://, //, //). The sequences provided are non-limiting examples and are meant to illustrate how single-molecule DNA-targeting RNAs and two-molecule DNA-targeting RNAs can be designed based on naturally existing sequences from a wide variety of species. Various examples of suitable seuqences from a wide variety of species are set forth as follows (Cas9 protein: SEQ ID NOs:1-259; tracrRNAs: SEQ ID NOs:431-562, or the complements thereof; crRNAs: SEQ ID NOs:563-679, or the complements thereof; and example single-molecule DNA-targeting RNAs: SEQ ID NOs:680-682).

[0042] **Figures 10A-E** show that Cas9 is a DNA endonuclease guided by two RNA molecules. Figure E (top to bottom, SEQ ID NOs: 278-280, and //).

[0043] **Figures 11A-B** demonstrate that Cas9 uses two nuclease domains to cleave the two strands in the target DNA.

[0044] **Figures 12A-E** illustrate that Cas9-catalyzed cleavage of target DNA requires an activating domain in tracrRNA and is governed by a seed sequence in the crRNA. Figure 12C (top to bottom, SEQ ID NO:278-280, and //); Figure 12D (top to bottom, SEQ ID NOs: 281-290); and Figure 12E (top to bottom, SEQ ID NOs: 291-292, 283, 293-298).

[0045] **Figures 13A-C** show that a PAM is required to license target DNA cleavage by the Cas9-tracrRNA:crRNA complex.

[0046] **Figures 14A-C** illustrate that Cas9 can be programmed using a single engineered RNA molecule combining tracrRNA and crRNA features. Chimera A (SEQ ID NO:299); Chimera B (SEQ ID NO:300).

[0047] **Figure 15** depicts the type II RNA-mediated CRISPR/Cas immune pathway.

[0048] **Figures 16A-B** depict purification of Cas9 nucleases.

[0049] **Figures 17A-C** show that Cas9 guided by dual-tracrRNA:crRNA cleaves protospacer plasmid and oligonucleotide DNA. Figure 17B (top to bottom, SEQ ID NOs: 301-303, and //); and Figure 17C (top to bottom, SEQ ID NO:304-306, and //).

[0050] **Figures 18A-B** show that Cas9 is a Mg2+-dependent endonuclease with 3'-5' exonuclease activity.

[0051] **Figures 19A-C** illustrate that dual-tracrRNA:crRNA directed Cas9 cleavage of target DNA is site specific. Figure 19C (top to bottom, SEQ ID NOs: 307-309, //, 337-339, and //).

[0052] **Figures 20A-B** show that dual-tracrRNA:crRNA directed Cas9 cleavage of target DNA is fast and efficient.

[0053] **Figures 21A-B** show that the HNH and RuvC-like domains of Cas9 direct cleavage of the complementary and noncomplementary DNA strand, respectively.

[0054] **Figure 22** demonstrates that tracrRNA is required for target DNA recognition.

[0055] **Figures 23A-B** show that a minimal region of tracrRNA is capable of guiding dualtracrRNA: crRNA directed cleavage of target DNA.

[0056] **Figures 24A-D** demonstrate that dual-tracrRNA:crRNA guided target DNA cleavage by Cas9 can be species specific.

[0057] **Figures 25A-C** show that a seed sequence in the crRNA governs dual tracrRNA:crRNA directed cleavage of target DNA by Cas9. Figure 25A: target DNA probe 1 (SEQ ID NO:310); spacer 4 crRNA (1-42) (SEQ ID NO:311); tracrRNA (15-89) (SEQ ID NO://). Figure 25B left panel (SEQ ID NO:310).

[0058] **Figures 26A-C** demonstrate that the PAM sequence is essential for protospacer plasmid DNA cleavage by Cas9-tracrRNA:crRNA and for Cas9-mediated plasmid DNA interference in bacterial cells. Figure 26B (top to bottom, SEQ ID NOs:312-314); and Figure 26C (top to bottom, SEQ ID NO:315-320).

[0059] **Figures 27A-C** show that Cas9 guided by a single chimeric RNA mimicking dual tracrRNA:crRNA cleaves protospacer DNA. Figure 27C (top to bottom, SEQ ID NO:321-324).

[0060] **Figures 28A-D** depict de novo design of chimeric RNAs targeting the Green Fluorescent Protein (GFP) gene sequence. Figure 28B (top to bottom, SEQ ID NOs:325-326). Figure 28C: GFP1 target sequence (SEQ ID NO:327); GFP2 target sequence (SEQ ID NO:328); GFP3 target sequence (SEQ ID NO:329); GFP4 target sequence (SEQ ID NO:330); GFP5 target sequence (SEQ ID NO:331); GFP1 chimeric RNA (SEQ ID NO:332); GFP2 chimeric RNA (SEQ ID NO:333); GFP3 chimeric RNA (SEQ ID NO:334); GFP4 chimeric RNA (SEQ ID NO:335); GFP5 chimeric RNA (SEQ ID NO:336).

[0061] **Figures 29A-E** demonstrate that co-expression of Cas9 and guide RNA in human cells generates double-strand DNA breaks at the target locus. Figure 29C (top to bottom, SEQ ID NO:425-428).

[0062] **Figures 30A-B** demonstrate that cell lysates contain active Cas9:sgRNA and support site-specific DNA cleavage.

[0063] **Figures 31A-B** demonstrate that 3' extension of sgRNA constructs enhances site-specific NHEJ-mediated mutagenesis. Figure 31A (top to bottom, SEQ ID NO:428-430).

[0064] **Figures 32A-B** depict a phylogenetic tree of representative Cas9 sequences from various organisms (A) as well as Cas9 locus architectures for the main groups of the tree (B).

[0065] **Figures 33A-E** depict the architecture of type II CRISPR-Cas from selected bacterial species.

[0066] **Figures 34A-B** depict tracrRNA and pre-crRNA co-processing in selected type II CRISPR Cas systems. Figure 34A (top to bottom, SEQ ID NO://,//,//,//,//,//,//,//); Figure 34B (top to bottom, SEQ ID NO://,//,//,//).

[0067] **Figure 35** depicts a sequence alignment of tracrRNA orthologues demonstrating the diversity of tracrRNA sequences.

[0068] **Figures 36A-F** depict the expression of bacterial tracrRNA orthologues and crRNAs revealed by deep RNA sequencing.

[0069] **Figures 37A-O** list all tracrRNA orthologues and mature crRNAs retrieved by sequencing for the bacterial species studied, including coordinates (region of interest) and corresponding cDNA sequences (5' to 3').

[0070] **Figures 38 A-B** present a table of bacterial species containing type II CRISPR-Cas loci characterized by the presence of the signature gene *cas9*. These sequences were used for phylogenetic analyses.

[0071] **Figures 39 A-B** depict the design of the CRISPR interference (CRISPRi) system.

[0072] **Figures 40 A-E** demonstrate that CRISPRi effectively silences transcription elongation and initiation.

[0073] **Figures 41 A-B** demonstrate that CRISPRi functions by blocking transcription elongation.

[0074] **Figures 42 A-C** demonstrate the targeting specificity of the CRISPRi system.

[0075] **Figures 43 A-F** depict the characterization of factors that affect silencing efficiency.

[0076] **Figures 44 A-C** depict functional profiling of a complex regulatory network using CRISPRi gene knockdown.

[0077] **Figures 45 A-B** demonstrates gene silencing using CRISPRi in mammalian cells.

[0078] **Figure 46** depicts the mechanism of the type II CRISPR system from *S. pyogenes*.

[0079] **Figures 47 A-B** depict the growth curves of *E. coli* cell cultures co-transformed with dCas9 and sgRNA.

[0080] **Figure 48** shows that CRISPRi could silence expression of a reporter gene on a multiple-copy plasmid.

[0081] **Figures 49 A-C** depict the RNA-seq data of cells with sgRNAs that target different genes.

[0082] **Figures 50 A-E** depict the silencing effects of sgRNAs with adjacent double mismatches.

[0083] **Figures 51 A-C** depict the combinatorial silencing effects of using two sgRNAs to regulate a single gene.

[0084] **Figure 52** shows that sgRNA repression is dependent on the target loci and relatively distance from the transcription start.

[0085] **Figures 53 A-C** depict experimental results demonstrating that a variant Cas9 site-directed polypeptide (dCas9) is works for the subject methods when dCas9 has reduced activity in the RuvC1 domain only (e.g., D10A), the HNH domain only (e.g., H840A), or both domains (e.g, D10A and H840A).

[0086] **Figures 54 A-C** list examples of suitable fusion partners (or fragments thereof) for a subject variant Cas9 site-directed polypeptide. Examples include, but are not limited to those listed.

[0087] **Figures 55 A-D** demonstrate that a chimeric site-directed polypeptide can be used to activate (increase) transcription in human cells.

[0088] **Figure 56** demonstrates that a chimeric site-directed polypeptide can be used to repress (decrease) transcription in human cells.

[0089] **Figures 57A-B** demonstrate that artificial sequences that share roughly 50% identity with naturally occurring a tracrRNAs and crRNAs can function with Cas9 to cleave target DNA as long as the structure of the protein-binding domain of the DNA-targeting RNA is conserved.

DEFINITIONS – PART I

[0090] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. "Oligonucleotide" generally refers to polynucleotides of between about 5 and about 100 nucleotides of single- or double-stranded DNA. However, for the purposes of this disclosure, there is no upper limit to the length of an oligonucleotide. Oligonucleotides are also known as "oligomers" or "oligos" and may be isolated from genes, or chemically synthesized by methods

known in the art. The terms “polynucleotide" and "nucleic acid" should be understood to include, as applicable to the embodiments being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

[0091] A "stem-loop structure" refers to a nucleic acid having a secondary structure that includes a region of nucleotides which are known or predicted to form a double strand (step portion) that is linked on one side by a region of predominantly single-stranded nucleotides (loop portion). The terms "hairpin" and "fold-back" structures are also used herein to refer to stem-loop structures. Such structures are well known in the art and these terms are used consistently with their known meanings in the art. As is known in the art, a stem-loop structure does not require exact base-pairing. Thus, the stem may include one or more base mismatches. Alternatively, the base-pairing may be exact, i.e. not include any mismatches.

[0092] By "hybridizable" or “complementary” or “substantially complementary" it is meant that a nucleic acid (e.g. RNA) comprises a sequence of nucleotides that enables it to non-covalently bind, i.e. form Watson-Crick base pairs and/or G/U base pairs, “anneal”, or “hybridize,” to another nucleic acid in a sequence-specific, antiparallel, manner (i.e., a nucleic acid specifically binds to a complementary nucleic acid) under the appropriate in vitro and/or in vivo conditions of temperature and solution ionic strength. As is known in the art, standard Watson-Crick base-pairing includes: adenine (A) pairing with thymidine (T), adenine (A) pairing with uracil (U), and guanine (G) pairing with cytosine (C) [DNA, RNA]. In addition, it is also known in the art that for hybridization between two RNA molecules (e.g., dsRNA), guanine (G) base pairs with uracil (U). For example, G/U base-pairing is partially responsible for the degeneracy (i.e., redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. In the context of this disclosure, a guanine (G) of a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule is considered complementary to a uracil (U), and vice versa. As such, when a G/U base-pair can be made at a given nucleotide position a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule, the position is not considered to be non-complementary, but is instead considered to be complementary.

[0093] Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein; and Sambrook, J. and Russell, W., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

[0094] Hybridization requires that the two nucleic acids contain complementary sequences, although mismatches between bases are possible. The conditions appropriate for hybridization between two nucleic acids depend on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of complementation between two nucleotide sequences, the greater the value of the melting temperature (Tm) for hybrids of nucleic acids having those sequences. For hybridizations between nucleic acids with short stretches of complementarity (e.g. complementarity over 35 or less, 30 or less, 25 or less, 22 or less, 20 or less, or 18 or less nucleotides) the position of mismatches becomes important (see Sambrook et al., supra, 11.7-11.8). Typically, the length for a hybridizable nucleic acid is at least about 10 nucleotides. Illustrative minimum lengths for a hybridizable nucleic acid are: at least about 15 nucleotides; at least about 20 nucleotides; at least about 22 nucleotides; at least about 25 nucleotides; and at least about 30 nucleotides). Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the region of complementation and the degree of complementation.

[0095] It is understood in the art that the sequence of polynucleotide need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable or hybridizable. Moreover, a polynucleotide may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a loop structure or hairpin structure). A polynucleotide can comprise at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence complementarity to a target region within the target nucleic acid sequence to which they are targeted. For example, an antisense nucleic acid in which 18 of 20 nucleotides of the antisense compound are complementary to a target region, and would therefore specifically hybridize, would represent 90 percent complementarity. In this example, the remaining noncomplementary nucleotides may be clustered or interspersed with complementary nucleotides and need not be contiguous to each other or to complementary nucleotides. Percent complementarity between particular stretches of nucleic acid sequences within nucleic acids can be determined routinely using BLAST programs (basic local alignment search tools) and PowerBLAST programs known in the art (Altschul et al., J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656) or by using the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489).

[0096] The terms "peptide," "polypeptide," and "protein" are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino

acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[0097] "Binding" as used herein (e.g. with reference to an RNA-binding domain of a polypeptide) refers to a non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). While in a state of non-covalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" (e.g., when a molecule X is said to interact with a molecule Y, it is meant the molecule X binds to molecule Y in a non-covalent manner). Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence-specific. Binding interactions are generally characterized by a dissociation constant (Kd) of less than 10^{-6} M, less than 10^{-7} M, less than 10^{-8} M, less than 10^{-9} M, less than 10^{-10} M, less than 10^{-11} M, less than 10^{-12} M, less than 10^{-13} M, less than 10^{-14} M, or less than 10^{-15} M. "Affinity" refers to the strength of binding, increased binding affinity being correlated with a lower Kd.

[0098] By "binding domain" it is meant a protein domain that is able to bind non-covalently to another molecule. A binding domain can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein domain-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins.

[0099] The term "conservative amino acid substitution" refers to the interchangeability in proteins of amino acid residues having similar side chains. For example, a group of amino acids having aliphatic side chains consists of glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains consists of serine and threonine; a group of amino acids having amide containing side chains consisting of asparagine and glutamine; a group of amino acids having aromatic side chains consists of phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains consists of lysine, arginine, and histidine; a group of amino acids having acidic side chains consists of glutamate and aspartate; and a group of amino acids having sulfur containing side chains consists of cysteine and methionine. Exemplary conservative amino acid substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[00100] A polynucleotide or polypeptide has a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same, and in the same relative position, when comparing the two sequences. Sequence identity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using various methods and computer programs (e.g., BLAST,

T-COFFEE, MUSCLE, MAFFT, etc.), available over the world wide web at sites including ncbi.nlm.nili.gov/BLAST, ebi.ac.uk/Tools/msa/tcoffee/, ebi.ac.uk/Tools/msa/muscle/, mafft.cbrc.jp/alignment/software/. See, e.g., Altschul et al. (1990), J. Mol. Bioi. 215:403-10.

[00101] A DNA sequence that "encodes" a particular RNA is a DNA nucleic acid sequence that is transcribed into RNA. A DNA polynucleotide may encode an RNA (mRNA) that is translated into protein, or a DNA polynucleotide may encode an RNA that is not translated into protein (e.g. tRNA, rRNA, or a DNA-targeting RNA; also called "non-coding" RNA or "ncRNA").

[00102] A "protein coding sequence" or a sequence that encodes a particular protein or polypeptide, is a nucleic acid sequence that is transcribed into mRNA (in the case of DNA) and is 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 (N-terminus) and a translation stop nonsense codon at the 3' terminus (C-terminus). A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic nucleic acids. A transcription termination sequence will usually be located 3' to the coding sequence.

[00103] As used herein, a "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding or non-coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the various vectors of the present invention.

[00104] A promoter can be a constitutively active promoter (i.e., a promoter that is constitutively in an active/"ON" state), it may be an inducible promoter (i.e., a promoter whose state, active/"ON" or inactive/"OFF", is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein.), it may be a spatially restricted promoter (i.e., transcriptional control element, enhancer, etc.)(e.g., tissue specific promoter, cell type specific promoter, etc.), and it may be a temporally restricted promoter (i.e., the promoter is in the "ON" state or "OFF" state during specific stages of embryonic development or during specific stages of a biological process, e.g., hair follicle cycle in mice).

[00105] Suitable promoters can be derived from viruses and can therefore be referred to as viral promoters, or they can be derived from any organism, including prokaryotic or eukaryotic organisms. Suitable promoters can be used to drive expression by any RNA polymerase (e.g., pol I, pol II, pol III). Exemplary promoters include, but are not limited to the SV40 early promoter, mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), a rous sarcoma virus (RSV) promoter, a human U6 small nuclear promoter (U6) (Miyagishi et al. , Nature Biotechnology 20, 497 - 500 (2002)), an enhanced U6 promoter (e.g., Xia et al., Nucleic Acids Res. 2003 Sep 1;31(17)), a human H1 promoter (H1), and the like.

[00106] Examples of inducible promoters include, but are not limited toT7 RNA polymerase promoter, T3 RNA polymerase promoter, Isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, lactose induced promoter, heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc. Inducible promoters can therefore be regulated by molecules including, but not limited to, doxycycline; RNA polymerase, e.g., T7 RNA polymerase; an estrogen receptor; an estrogen receptor fusion; etc.

[00107] In some embodiments, the promoter is a spatially restricted promoter (i.e., cell type specific promoter, tissue specific promoter, etc.) such that in a multi-cellular organism, the promoter is active (i.e., "ON") in a subset of specific cells. Spatially restricted promoters may also be referred to as enhancers, transcriptional control elements, control sequences, etc. Any convenient spatially restricted promoter may be used and the choice of suitable promoter (e.g., a brain specific promoter, a promoter that drives expression in a subset of neurons, a promoter that drives expression in the germline, a promoter that drives expression in the lungs, a promoter that drives expression in muscles, a promoter that drives expression in islet cells of the pancreas, etc.) will depend on the organism. For example, various spatially restricted promoters are known for plants, flies, worms, mammals, mice, etc. Thus, a spatially restricted promoter can be used to regulate the expression of a nucleic acid encoding a subject site-directed modifying polypeptide in a wide variety of different tissues and cell types, depending on the organism. Some spatially restricted promoters are also temporally restricted such that the promoter is in the "ON" state or "OFF" state during specific stages of embryonic development or during specific stages of a biological process (e.g., hair follicle cycle in mice).

[00108] For illustration purposes, examples of spatially restricted promoters include, but are not limited to, neuron-specific promoters, adipocyte-specific promoters, cardiomyocyte-specific promoters, smooth muscle-specific promoters, photoreceptor-specific promoters, etc. Neuron-specific

spatially restricted promoters include, but are not limited to, a neuron-specific enolase (NSE) promoter (see, e.g., EMBL HSENO2, X51956); an aromatic amino acid decarboxylase (AADC) promoter; a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al. (1987) Cell 51:7-19; and Llewellyn, et al. (2010) Nat. Med. 16(10):1161-1166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., Oh et al. (2009) Gene Ther 16:437; Sasaoka et al. (1992) Mol. Brain Res. 16:274; Boundy et al. (1998) J. Neurosci. 18:9989; and Kaneda et al. (1991) Neuron 6:583-594); a GnRH promoter (see, e.g., Radovick et al. (1991) Proc. Natl. Acad. Sci. USA 88:3402-3406); an L7 promoter (see, e.g., Oberdick et al. (1990) Science 248:223-226); a DNMT promoter (see, e.g., Bartge et al. (1988) Proc. Natl. Acad. Sci. USA 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al. (1988) EMBO J. 17:3793-3805); a myelin basic protein (MBP) promoter; a Ca2+-calmodulin-dependent protein kinase II-alpha (CamKIIα) promoter (see, e.g., Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250; and Casanova et al. (2001) Genesis 31:37); a CMV enhancer/platelet-derived growth factor-β promoter (see, e.g., Liu et al. (2004) Gene Therapy 11:52-60); and the like.

[00109] Adipocyte-specific spatially restricted promoters include, but are not limited to aP2 gene promoter/enhancer, e.g., a region from -5.4 kb to +21 bp of a human aP2 gene (see, e.g., Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA 87:9590; and Pavjani et al. (2005) Nat. Med. 11:797); a glucose transporter-4 (GLUT4) promoter (see, e.g., Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725); a fatty acid translocase (FAT/CD36) promoter (see, e.g., Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476; and Sato et al. (2002) J. Biol. Chem. 277:15703); a stearoyl-CoA desaturase-1 (SCD1) promoter (Tabor et al. (1999) J. Biol. Chem. 274:20603); a leptin promoter (see, e.g., Mason et al. (1998) Endocrinol. 139:1013; and Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187); an adiponectin promoter (see, e.g., Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484; and Chakrabarti (2010) Endocrinol. 151:2408); an adipsin promoter (see, e.g., Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490); a resistin promoter (see, e.g., Seo et al. (2003) Molec. Endocrinol. 17:1522); and the like.

[00110] Cardiomyocyte-specific spatially restricted promoters include, but are not limited to control sequences derived from the following genes: myosin light chain-2, α-myosin heavy chain, AE3, cardiac troponin C, cardiac actin, and the like. Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617; and Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

[00111] Smooth muscle-specific spatially restricted promoters include, but are not limited to an SM22α promoter (see, e.g., Akyürek et al. (2000) Mol. Med. 6:983; and U.S. Patent No. 7,169,874); a smoothelin promoter (see, e.g., WO 2001/018048); an α-smooth muscle actin promoter; and the like. For example, a 0.4 kb region of the SM22α promoter, within which lie two CArG elements, has been shown to mediate vascular smooth muscle cell-specific expression (see, e.g., Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; and Moessler, et al. (1996) Development 122, 2415-2425).

[00112] Photoreceptor-specific spatially restricted promoters include, but are not limited to, a rhodopsin promoter; a rhodopsin kinase promoter (Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); a beta phosphodiesterase gene promoter (Nicoud et al. (2007) J. Gene Med. 9:1015); a retinitis pigmentosa gene promoter (Nicoud et al. (2007) supra); an interphotoreceptor retinoid-binding protein (IRBP) gene enhancer (Nicoud et al. (2007) supra); an IRBP gene promoter (Yokoyama et al. (1992) Exp Eye Res. 55:225); and the like.

[00113] The terms "DNA regulatory sequences," "control elements," and "regulatory elements," used interchangeably herein, refer to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate transcription of a non-coding sequence (e.g., DNA-targeting RNA) or a coding sequence (e.g., site-directed modifying polypeptide, or Cas9/Csn1 polypeptide) and/or regulate translation of an encoded polypeptide.

[00114] The term "naturally-occurring" or "unmodified" as used herein as applied to a nucleic acid, a polypeptide, a cell, or an organism, refers to a nucleic acid, polypeptide, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by a human in the laboratory is naturally occurring.

[00115] The term "chimeric" as used herein as applied to a nucleic acid or polypeptide refers to two components that are defined by structures derived from different sources. For example, where "chimeric" is used in the context of a chimeric polypeptide (e.g., a chimeric Cas9/Csn1 protein), the chimeric polypeptide includes amino acid sequences that are derived from different polypeptides. A chimeric polypeptide may comprise either modified or naturally-occurring polypeptide sequences (e.g., a first amino acid sequence from a modified or unmodified Cas9/Csn1 protein; and a second amino acid sequence other than the Cas9/Csn1 protein). Similarly, "chimeric" in the context of a polynucleotide encoding a chimeric polypeptide includes nucleotide sequences derived from different coding regions (e.g., a first nucleotide

sequence encoding a modified or unmodified Cas9/Csn1 protein; and a second nucleotide sequence encoding a polypeptide other than a Cas9/Csn1 protein).

[00116] The term "chimeric polypeptide" refers to a polypeptide which is made by the combination (i.e., "fusion") of two otherwise separated segments of amino sequence, usually through human intervention. A polypeptide that comprises a chimeric amino acid sequence is a chimeric polypeptide. Some chimeric polypeptides can be referred to as "fusion variants."

[00117] "Heterologous," as used herein, means a nucleotide or polypeptide sequence that is not found in the native nucleic acid or protein, respectively. For example, in a chimeric Cas9/Csn1 protein, the RNA-binding domain of a naturally-occurring bacterial Cas9/Csn1 polypeptide (or a variant thereof) may be fused to a heterologous polypeptide sequence (i.e. a polypeptide sequence from a protein other than Cas9/Csn1 or a polypeptide sequence from another organism). The heterologous polypeptide sequence may exhibit an activity (e.g., enzymatic activity) that will also be exhibited by the chimeric Cas9/Csn1 protein (e.g., methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.). A heterologous nucleic acid sequence may be linked to a naturally-occurring nucleic acid sequence (or a variant thereof) (e.g., by genetic engineering) to generate a chimeric nucleotide sequence encoding a chimeric polypeptide. As another example, in a fusion variant Cas9 site-directed polypeptide, a variant Cas9 site-directed polypeptide may be fused to a heterologous polypeptide (i.e. a polypeptide other than Cas9), which exhibits an activity that will also be exhibited by the fusion variant Cas9 site-directed polypeptide. A heterologous nucleic acid sequence may be linked to a variant Cas9 site-directed polypeptide (e.g., by genetic engineering) to generate a nucleotide sequence encoding a fusion variant Cas9 site-directed polypeptide.

[00118] "Recombinant," as used herein, means that a particular nucleic acid (DNA or RNA) is the product of various combinations of cloning, restriction, polymerase chain reaction (PCR) and/or ligation steps resulting in a construct having a structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems. DNA sequences encoding polypeptides can be assembled from cDNA fragments or from a series of synthetic oligonucleotides, to provide a synthetic nucleic acid which is capable of being expressed from a recombinant transcriptional unit contained in a cell or in a cell-free transcription and translation system. Genomic DNA comprising the relevant sequences can also be used in the formation of a recombinant gene or transcriptional unit. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame, where such sequences do not interfere with manipulation or expression of the coding regions, and may indeed act to modulate production of a desired product by various mechanisms (see "DNA regulatory sequences", below). Alternatively, DNA sequences encoding RNA (e.g., DNA-targeting RNA) that is not translated may also be

considered recombinant. Thus, e.g., the term "recombinant" nucleic acid refers to one which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a codon encoding the same amino acid, a conservative amino acid, or a non-conservative amino acid. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. When a recombinant polynucleotide encodes a polypeptide, the sequence of the encoded polypeptide can be naturally occurring ("wild type") or can be a variant (e.g., a mutant) of the naturally occurring sequence. Thus, the term "recombinant" polypeptide does not necessarily refer to a polypeptide whose sequence does not naturally occur. Instead, a "recombinant" polypeptide is encoded by a recombinant DNA sequence, but the sequence of the polypeptide can be naturally occurring ("wild type") or non-naturally occurring (e.g., a variant, a mutant, etc.). Thus, a "recombinant" polypeptide is the result of human intervention, but may be a naturally occurring amino acid sequence.

[00119] A "vector" or "expression vector" is a replicon, such as plasmid, phage, virus, or cosmid, to which another DNA segment, i.e. an "insert", may be attached so as to bring about the replication of the attached segment in a cell.

[00120] An "expression cassette" comprises a DNA coding sequence operably linked to a promoter. "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

[00121] The terms "recombinant expression vector," or "DNA construct" are used interchangeably herein to refer to a DNA molecule comprising a vector and at least one insert. Recombinant expression vectors are usually generated for the purpose of expressing and/or propagating the insert(s), or for the construction of other recombinant nucleotide sequences. The insert(s) may or may not be operably linked to a promoter sequence and may or may not be operably linked to DNA regulatory sequences.

[00122] A cell has been "genetically modified" or "transformed" or "transfected" by exogenous DNA, e.g. a recombinant expression vector, when such DNA has been introduced inside the cell. The presence of the exogenous DNA results in permanent or transient genetic change. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell.

In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones that comprise a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

[00123] Suitable methods of genetic modification (also referred to as "transformation") include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, e.g., Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like.

[00124] The choice of method of genetic modification is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (e.g., in vitro, ex vivo, or in vivo). A general discussion of these methods can be found in Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

[00125] A "target DNA" as used herein is a DNA polynucleotide that comprises a "target site" or "target sequence." The terms "target site" or "target sequence" or "target protospacer DNA" are used interchangeably herein to refer to a nucleic acid sequence present in a target DNA to which a DNA-targeting segment of a subject DNA-targeting RNA will bind (see Figure 1 and Figure 39), provided sufficient conditions for binding exist. For example, the target site (or target sequence) 5'-GAGCATATC-3' (SEQ ID NO:///) within a target DNA is targeted by (or is bound by, or hybridizes with, or is complementary to) the RNA sequence 5'-GAUAUGCUC-3' (SEQ ID NO:///). Suitable DNA/RNA binding conditions include physiological conditions normally present in a cell. Other suitable DNA/RNA binding conditions (e.g., conditions in a cell-free system) are known in the art; see, e.g., Sambrook, supra. The strand of the target DNA that is complementary to and hybridizes with the DNA-targeting RNA is referred to as the "complementary strand" and the strand of the target DNA that is complementary to the "complementary strand" (and is therefore not complementary to the DNA-targeting RNA) is referred to as the "noncomplementary strand" or "non-complementary strand" (see Figure 12).

[00126] By "site-directed modifying polypeptide" or "RNA-binding site-directed polypeptide" or "RNA-binding site-directed modifying polypeptide" or "site-directed polypeptide" it is meant a

polypeptide that binds RNA and is targeted to a specific DNA sequence. A site-directed modifying polypeptide as described herein is targeted to a specific DNA sequence by the RNA molecule to which it is bound. The RNA molecule comprises a sequence that is complementary to a target sequence within the target DNA, thus targeting the bound polypeptide to a specific location within the target DNA (the target sequence).

[00127] By "cleavage" it is meant the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, a complex comprising a DNA-targeting RNA and a site-directed modifying polypeptide is used for targeted double-stranded DNA cleavage.

[00128] "Nuclease" and "endonuclease" are used interchangeably herein to mean an enzyme which possesses catalytic activity for DNA cleavage.

[00129] By "cleavage domain" or "active domain" or "nuclease domain" of a nuclease it is meant the polypeptide sequence or domain within the nuclease which possesses the catalytic activity for DNA cleavage. A cleavage domain can be contained in a single polypeptide chain or cleavage activity can result from the association of two (or more) polypeptides. A single nuclease domain may consist of more than one isolated stretch of amino acids within a given polypeptide.

[00130] The RNA molecule that binds to the site-directed modifying polypeptide and targets the polypeptide to a specific location within the target DNA is referred to herein as the "DNA-targeting RNA" or "DNA-targeting RNA polynucleotide" (also referred to herein as a "guide RNA" or "gRNA"). A subject DNA-targeting RNA comprises two segments, a "DNA-targeting segment" and a "protein-binding segment." By "segment" it is meant a segment/section/region of a molecule, e.g., a contiguous stretch of nucleotides in an RNA. A segment can also mean a region/section of a complex such that a segment may comprise regions of more than one molecule. For example, in some cases the protein-binding segment (described below) of a DNA-targeting RNA is one RNA molecule and the protein-binding segment therefore comprises a region of that RNA molecule. In other cases, the protein-binding segment (described below) of a DNA-targeting RNA comprises two separate molecules that are hybridized along a region of complementarity. As an illustrative, non-limiting example, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of "segment," unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs,

is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and may include regions of RNA molecules that are of any total length and may or may not include regions with complementarity to other molecules.

[00131] The DNA-targeting segment (or "DNA-targeting sequence") comprises a nucleotide sequence that is complementary to a specific sequence within a target DNA (the complementary strand of the target DNA). The protein-binding segment (or "protein-binding sequence") interacts with a site-directed modifying polypeptide. When the site-directed modifying polypeptide is a Cas9 or Cas9 related polypeptide (described in more detail below), site-specific cleavage of the target DNA occurs at locations determined by both (i) base-pairing complementarity between the DNA-targeting RNA and the target DNA; and (ii) a short motif (referred to as the protospacer adjacent motif (PAM)) in the target DNA.

[00132] The protein-binding segment of a subject DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (dsRNA duplex).

[00133] In some embodiments, a subject nucleic acid (e.g., a DNA-targeting RNA, a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; a nucleic acid encoding a site-directed polypeptide; etc.) comprises a modification or sequence that provides for an additional desirable feature (e.g., modified or regulated stability; subcellular targeting; tracking, e.g., a fluorescent label; a binding site for a protein or protein complex; etc.). Non-limiting examples include: a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and/or protein complexes); a stability control sequence; a sequence that forms a dsRNA duplex (i.e., a hairpin)); a modification or sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

[00134] In some embodiments, a DNA-targeting RNA comprises an additional segment at either the 5' or 3' end that provides for any of the features described above. For example, a suitable third segment can comprise a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or

regulated accessibility by proteins and protein complexes); a stability control sequence; a sequence that forms a dsRNA duplex (i.e., a hairpin)); a sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

[00135] A subject DNA-targeting RNA and a subject site-directed modifying polypeptide (i.e., site-directed polypeptide) form a complex (i.e., bind via non-covalent interactions). The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-directed modifying polypeptide of the complex provides the site-specific activity. In other words, the site-directed modifying polypeptide is guided to a target DNA sequence (e.g. a target sequence in a chromosomal nucleic acid; a target sequence in an extrachromosomal nucleic acid, e.g. an episomal nucleic acid, a minicircle, etc.; a target sequence in a mitochondrial nucleic acid; a target sequence in a chloroplast nucleic acid; a target sequence in a plasmid; etc.) by virtue of its association with the protein-binding segment of the DNA-targeting RNA.

[00136] In some embodiments, a subject DNA-targeting RNA comprises two separate RNA molecules (RNA polynucleotides: an "activator-RNA" and a "targeter-RNA", see below) and is referred to herein as a "double-molecule DNA-targeting RNA" or a "two-molecule DNA-targeting RNA." In other embodiments, the subject DNA-targeting RNA is a single RNA molecule (single RNA polynucleotide) and is referred to herein as a "single-molecule DNA-targeting RNA," a "single-guide RNA," or an "sgRNA." The term "DNA-targeting RNA" or "gRNA" is inclusive, referring both to double-molecule DNA-targeting RNAs and to single-molecule DNA-targeting RNAs (i.e., sgRNAs).

[00137] An exemplary two-molecule DNA-targeting RNA comprises a crRNA-like ("CRISPR RNA" or "targeter-RNA" or "crRNA" or "crRNA repeat") molecule and a corresponding tracrRNA-like ("trans-acting CRISPR RNA" or "activator-RNA" or "tracrRNA") molecule. A crRNA-like molecule (targeter-RNA) comprises both the DNA-targeting segment (single stranded) of the DNA-targeting RNA and a stretch ("duplex-forming segment") of nucleotides that forms one half of the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. A corresponding tracrRNA-like molecule (activator-RNA) comprises a stretch of nucleotides (duplex-forming segment) that forms the other half of the dsRNA duplex of the protein-binding

segment of the DNA-targeting RNA. In other words, a stretch of nucleotides of a crRNA-like molecule are complementary to and hybridize with a stretch of nucleotides of a tracrRNA-like molecule to form the dsRNA duplex of the protein-binding domain of the DNA-targeting RNA. As such, each crRNA-like molecule can be said to have a corresponding tracrRNA-like molecule. The crRNA-like molecule additionally provides the single stranded DNA-targeting segment. Thus, a crRNA-like and a tracrRNA-like molecule (as a corresponding pair) hybridize to form a DNA-targeting RNA. The exact sequence of a given crRNA or tracrRNA molecule is characteristic of the species in which the RNA molecules are found. Various crRNAs and tracrRNAs are depicted in corresponding complementary pairs in Figures 8. A subject double-molecule DNA-targeting RNA can comprise any corresponding crRNA and tracrRNA pair. A subject double-molecule DNA-targeting RNA can comprise any corresponding crRNA and tracrRNA pair.

[00138] The term "activator-RNA" is used herein to mean a tracrRNA-like molecule of a double-molecule DNA-targeting RNA. The term "targeter-RNA" is used herein to mean a crRNA-like molecule of a double-molecule DNA-targeting RNA. The term "duplex-forming segment" is used herein to mean the stretch of nucleotides of an activator-RNA or a targeter-RNA that contributes to the formation of the dsRNA duplex by hybridizing to a stretch of nucleotides of a corresponding activator-RNA or targeter-RNA molecule. In other words, an activator-RNA comprises a duplex-forming segment that is complementary to the duplex-forming segment of the corresponding targeter-RNA. As such, an activator-RNA comprises a duplex-forming segment while a targeter-RNA comprises both a duplex-forming segment and the DNA-targeting segment of the DNA-targeting RNA. Therefore, a subject double-molecule DNA-targeting RNA can be comprised of any corresponding activator-RNA and targeter-RNA pair.

[00139] A "host cell," as used herein, denotes an in vivo or in vitro eukaryotic cell, a prokaryotic cell (e.g., bacterial or archaeal cell), or a cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells can be, or have been, used as recipients for a nucleic acid, and include the progeny of the original cell which has been transformed by the nucleic acid. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation. A "recombinant host cell" (also referred to as a "genetically modified host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector. For example, a subject bacterial host cell is a genetically modified bacterial host cell by virtue of introduction into a suitable bacterial host cell of an exogenous nucleic acid (e.g., a plasmid or recombinant expression vector) and a subject eukaryotic host cell is a genetically modified eukaryotic host

cell (e.g., a mammalian germ cell), by virtue of introduction into a suitable eukaryotic host cell of an exogenous nucleic acid.

[00140] The term "stem cell" is used herein to refer to a cell (e.g., plant stem cell, vertebrate stem cell) that has the ability both to self-renew and to generate a differentiated cell type (see Morrison et al. (1997) Cell 88:287-298). In the context of cell ontogeny, the adjective "differentiated", or "differentiating" is a relative term. A "differentiated cell" is a cell that has progressed further down the developmental pathway than the cell it is being compared with. Thus, pluripotent stem cells (described below) can differentiate into lineage-restricted progenitor cells (e.g., mesodermal stem cells), which in turn can differentiate into cells that are further restricted (e.g., neuron progenitors), which can differentiate into end-stage cells (i.e., terminally differentiated cells, e.g., neurons, cardiomyocytes, etc.), which play a characteristic role in a certain tissue type, and may or may not retain the capacity to proliferate further. Stem cells may be characterized by both the presence of specific markers (e.g., proteins, RNAs, etc.) and the absence of specific markers. Stem cells may also be identified by functional assays both in vitro and in vivo, particularly assays relating to the ability of stem cells to give rise to multiple differentiated progeny.

[00141] Stem cells of interest include pluripotent stem cells (PSCs). The term "pluripotent stem cell" or "PSC" is used herein to mean a stem cell capable of producing all cell types of the organism. Therefore, a PSC can give rise to cells of all germ layers of the organism (e.g., the endoderm, mesoderm, and ectoderm of a vertebrate). Pluripotent cells are capable of forming teratomas and of contributing to ectoderm, mesoderm, or endoderm tissues in a living organism. Pluripotent stem cells of plants are capable of giving rise to all cell types of the plant (e.g., cells of the root, stem, leaves, etc.).

[00142] PSCs of animals can be derived in a number of different ways. For example, embryonic stem cells (ESCs) are derived from the inner cell mass of an embryo (Thomson et. al, Science. 1998 Nov 6;282(5391):1145-7) whereas induced pluripotent stem cells (iPSCs) are derived from somatic cells (Takahashi et. al, Cell. 2007 Nov 30;131(5):861-72; Takahashi et. al, Nat Protoc. 2007;2(12):3081-9; Yu et. al, Science. 2007 Dec 21;318(5858):1917-20. Epub 2007 Nov 20). Because the term PSC refers to pluripotent stem cells regardless of their derivation, the term PSC encompasses the terms ESC and iPSC, as well as the term embryonic germ stem cells (EGSC), which are another example of a PSC. PSCs may be in the form of an established cell line, they may be obtained directly from primary embryonic tissue, or they may be derived from a somatic cell. PSCs can be target cells of the methods described herein.

[00143] By "embryonic stem cell" (ESC) is meant a PSC that was isolated from an embryo, typically from the inner cell mass of the blastocyst. ESC lines are listed in the NIH Human Embryonic

Stem Cell Registry, e.g. hESBGN-01, hESBGN-02, hESBGN-03, hESBGN-04 (BresaGen, Inc.); HES-1, HES-2, HES-3, HES-4, HES-5, HES-6 (ES Cell International); Miz-hES1 (MizMedi Hospital-Seoul National University); HSF-1, HSF-6 (University of California at San Francisco); and H1, H7, H9, H13, H14 (Wisconsin Alumni Research Foundation (WiCell Research Institute)). Stem cells of interest also include embryonic stem cells from other primates, such as Rhesus stem cells and marmoset stem cells. The stem cells may be obtained from any mammalian species, e.g. human, equine, bovine, porcine, canine, feline, rodent, e.g. mice, rats, hamster, primate, etc. (Thomson et al. (1998) Science 282:1145; Thomson et al. (1995) Proc. Natl. Acad. Sci USA 92:7844; Thomson et al. (1996) Biol. Reprod. 55:254; Shamblott et al., Proc. Natl. Acad. Sci. USA 95:13726, 1998). In culture, ESCs typically grow as flat colonies with large nucleo-cytoplasmic ratios, defined borders and prominent nucleoli. In addition, ESCs express SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, and Alkaline Phosphatase, but not SSEA-1. Examples of methods of generating and characterizing ESCs may be found in, for example, US Patent No. 7,029,913, US Patent No. 5,843,780, and US Patent No. 6,200,806, the disclosures of which are incorporated herein by reference. Methods for proliferating hESCs in the undifferentiated form are described in WO 99/20741, WO 01/51616, and WO 03/020920.

[00144] By "embryonic germ stem cell" (EGSC) or "embryonic germ cell" or "EG cell" is meant a PSC that is derived from germ cells and/or germ cell progenitors, e.g. primordial germ cells, i.e. those that would become sperm and eggs. Embryonic germ cells (EG cells) are thought to have properties similar to embryonic stem cells as described above. Examples of methods of generating and characterizing EG cells may be found in, for example, US Patent No. 7,153,684; Matsui, Y., et al., (1992) Cell 70:841; Shamblott, M., et al. (2001) Proc. Natl. Acad. Sci. USA 98: 113; Shamblott, M., et al. (1998) Proc. Natl. Acad. Sci. USA, 95:13726; and Koshimizu, U., et al. (1996) Development, 122:1235, the disclosures of which are incorporated herein by reference.

[00145] By "induced pluripotent stem cell" or "iPSC" it is meant a PSC that is derived from a cell that is not a PSC (i.e., from a cell this is differentiated relative to a PSC). iPSCs can be derived from multiple different cell types, including terminally differentiated cells. iPSCs have an ES cell-like morphology, growing as flat colonies with large nucleo-cytoplasmic ratios, defined borders and prominent nuclei. In addition, iPSCs express one or more key pluripotency markers known by one of ordinary skill in the art, including but not limited to Alkaline Phosphatase, SSEA3, SSEA4, Sox2, Oct3/4, Nanog, TRA160, TRA181, TDGF 1, Dnmt3b, FoxD3, GDF3, Cyp26a1, TERT, and zfp42. Examples of methods of generating and characterizing iPSCs may be found in, for example, U.S. Patent Publication Nos. US20090047263, US20090068742, US20090191159, US20090227032, US20090246875, and US20090304646, the disclosures of

which are incorporated herein by reference. Generally, to generate iPSCs, somatic cells are provided with reprogramming factors (e.g. Oct4, SOX2, KLF4, MYC, Nanog, Lin28, etc.) known in the art to reprogram the somatic cells to become pluripotent stem cells.

[00146] By "somatic cell" it is meant any cell in an organism that, in the absence of experimental manipulation, does not ordinarily give rise to all types of cells in an organism. In other words, somatic cells are cells that have differentiated sufficiently that they will not naturally generate cells of all three germ layers of the body, i.e. ectoderm, mesoderm and endoderm. For example, somatic cells would include both neurons and neural progenitors, the latter of which may be able to naturally give rise to all or some cell types of the central nervous system but cannot give rise to cells of the mesoderm or endoderm lineages.

[00147] By "mitotic cell" it is meant a cell undergoing mitosis. Mitosis is the process by which a eukaryotic cell separates the chromosomes in its nucleus into two identical sets in two separate nuclei. It is generally followed immediately by cytokinesis, which divides the nuclei, cytoplasm, organelles and cell membrane into two cells containing roughly equal shares of these cellular components.

[00148] By "post-mitotic cell" it is meant a cell that has exited from mitosis, i.e., it is "quiescent", i.e. it is no longer undergoing divisions. This quiescent state may be temporary, i.e. reversible, or it may be permanent.

[00149] By "meiotic cell" it is meant a cell that is undergoing meiosis. Meiosis is the process by which a cell divides its nuclear material for the purpose of producing gametes or spores. Unlike mitosis, in meiosis, the chromosomes undergo a recombination step which shuffles genetic material between chromosomes. Additionally, the outcome of meiosis is four (genetically unique) haploid cells, as compared with the two (genetically identical) diploid cells produced from mitosis.

[00150] By "recombination" it is meant a process of exchange of genetic information between two polynucleotides. As used herein, "homology-directed repair (HDR)" refers to the specialized form DNA repair that takes place, for example, during repair of double-strand breaks in cells. This process requires nucleotide sequence homology, uses a "donor" molecule to template repair of a "target" molecule (i.e., the one that experienced the double-strand break), and leads to the transfer of genetic information from the donor to the target. Homology-directed repair may result in an alteration of the sequence of the target molecule (e.g., insertion, deletion, mutation), if the donor polynucleotide differs from the target molecule and part or all of the sequence of the donor polynucleotide is incorporated into the target DNA. In some embodiments, the donor

polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA.

[00151] By "non-homologous end joining (NHEJ)" it is meant the repair of double-strand breaks in DNA by direct ligation of the break ends to one another without the need for a homologous template (in contrast to homology-directed repair, which requires a homologous sequence to guide repair). NHEJ often results in the loss (deletion) of nucleotide sequence near the site of the double-strand break.

[00152] The terms "treatment", "treating" and the like are used herein to generally mean obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease or symptom in a mammal, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to acquiring the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease or symptom, i.e., arresting its development; or (c) relieving the disease, i.e., causing regression of the disease. The therapeutic agent may be administered before, during or after the onset of disease or injury. The treatment of ongoing disease, where the treatment stabilizes or reduces the undesirable clinical symptoms of the patient, is of particular interest. Such treatment is desirably performed prior to complete loss of function in the affected tissues. The subject therapy will desirably be administered during the symptomatic stage of the disease, and in some cases after the symptomatic stage of the disease.

[00153] The terms "individual," "subject," "host," and "patient," are used interchangeably herein and refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans.

[00154] General methods in molecular and cellular biochemistry can be found in such standard textbooks as Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., HaRBor Laboratory Press 2001); Short Protocols in Molecular Biology, 4th Ed. (Ausubel et al. eds., John Wiley & Sons 1999); Protein Methods (Bollag et al., John Wiley & Sons 1996); Nonviral Vectors for Gene Therapy (Wagner et al. eds., Academic Press 1999); Viral Vectors (Kaplift & Loewy eds., Academic Press 1995); Immunology Methods Manual (I. Lefkovits ed., Academic Press 1997); and Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998), the disclosures of which are incorporated herein by reference.

[00155] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[00156] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[00157] Certain ranges are presented herein with numerical values being preceded by the term "about." The term "about" is used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating unrecited number may be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number.

[00158] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[00159] All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[00160] It is noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "the polypeptide" includes reference to one or more polypeptides and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[00161] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the invention are specifically embraced by the present invention and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present invention and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[00162] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

DETAILED DESCRIPTION – PART I

[00163] The present disclosure provides a DNA-targeting RNA that comprises a targeting sequence and, together with a modifying polypeptide, provides for site-specific modification of a target DNA and/or a polypeptide associated with the target DNA. The present disclosure further provides site-specific modifying polypeptides. The present disclosure further provides methods of site-specific modification of a target DNA and/or a polypeptide associated with the target DNA The present disclosure provides methods of modulating transcription of a target nucleic acid in a target cell, generally involving contacting the target nucleic acid with an enzymatically inactive Cas9 polypeptide and a DNA-targeting RNA. Kits and compositions for carrying out the methods are also provided. The present disclosure provides genetically modified cells that produce Cas9; and Cas9 transgenic non-human multicellular organisms.

NUCLEIC ACIDS

DNA-targeting RNA

[00164] The present disclosure provides a DNA-targeting RNA that directs the activities of an associated polypeptide (e.g., a site-directed modifying polypeptide) to a specific target sequence within a target DNA. A subject DNA-targeting RNA comprises: a first segment (also referred to herein as a "DNA-targeting segment" or a "DNA-targeting sequence") and a second segment (also referred to herein as a "protein-binding segment" or a "protein-binding sequence").

DNA-targeting segment of a DNA-targeting RNA

[00165] The DNA-targeting segment of a subject DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA. In other words, the DNA-targeting segment of a subject DNA-targeting RNA interacts with a target DNA in a sequence-specific manner via hybridization (i.e., base pairing). As such, the nucleotide sequence of the DNA-targeting segment may vary and determines the location within the target DNA that the DNA-targeting RNA and the target DNA will interact. The DNA-targeting segment of a subject DNA-targeting RNA can be modified (e.g., by genetic engineering) to hybridize to any desired sequence within a target DNA.

[00166] The DNA-targeting segment can have a length of from about 12 nucleotides to about 100 nucleotides. For example, the DNA-targeting segment can have a length of from about 12 nucleotides (nt) to about 80 nt, from about 12 nt to about 50nt, from about 12 nt to about 40 nt, from about 12 nt to about 30 nt, from about 12 nt to about 25 nt, from about 12 nt to about 20 nt, or from about 12 nt to about 19 nt. For example, the DNA-targeting segment can have a length of from about 19 nt to about 20 nt, from about 19 nt to about 25 nt, from about 19 nt to about 30 nt, from about 19 nt to about 35 nt, from about 19 nt to about 40 nt, from about 19 nt to about 45 nt, from about 19 nt to about 50 nt, from about 19 nt to about 60 nt, from about 19 nt to about 70 nt, from about 19 nt to about 80 nt, from about 19 nt to about 90 nt, from about 19 nt to about 100 nt, from about 20 nt to about 25 nt, from about 20 nt to about 30 nt, from about 20 nt to about 35 nt, from about 20 nt to about 40 nt, from about 20 nt to about 45 nt, from about 20 nt to about 50 nt, from about 20 nt to about 60 nt, from about 20 nt to about 70 nt, from about 20 nt to about 80 nt, from about 20 nt to about 90 nt, or from about 20 nt to about 100 nt. The nucleotide sequence (the DNA-targeting sequence) of the DNA-targeting segment that is complementary to a nucleotide sequence (target sequence) of the target DNA can have a length at least about 12 nt. For example, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA can have a length at least about 12 nt, at least about 15 nt, at least about 18 nt, at least about 19 nt, at least about 20 nt, at least about 25 nt, at least about 30 nt, at least about 35 nt or at least about 40 nt. For example, the DNA-targeting sequence of the

DNA-targeting segment that is complementary to a target sequence of the target DNA can have a length of from about 12 nucleotides (nt) to about 80 nt, from about 12 nt to about 50nt, from about 12 nt to about 45 nt, from about 12 nt to about 40 nt, from about 12 nt to about 35 nt, from about 12 nt to about 30 nt, from about 12 nt to about 25 nt, from about 12 nt to about 20 nt, from about 12 nt to about 19 nt, from about 19 nt to about 20 nt, from about 19 nt to about 25 nt, from about 19 nt to about 30 nt, from about 19 nt to about 35 nt, from about 19 nt to about 40 nt, from about 19 nt to about 45 nt, from about 19 nt to about 50 nt, from about 19 nt to about 60 nt, from about 20 nt to about 25 nt, from about 20 nt to about 30 nt, from about 20 nt to about 35 nt, from about 20 nt to about 40 nt, from about 20 nt to about 45 nt, from about 20 nt to about 50 nt, or from about 20 nt to about 60 nt. The nucleotide sequence (the DNA-targeting sequence) of the DNA-targeting segment that is complementary to a nucleotide sequence (target sequence) of the target DNA can have a length at least about 12 nt.

[00167] In some cases, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA is 20 nucleotides in length. In some cases, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA is 19 nucleotides in length.

[00168] The percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA can be at least 60% (e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%). In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is 100% over the seven contiguous 5'-most nucleotides of the target sequence of the complementary strand of the target DNA. In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is at least 60% over about 20 contiguous nucleotides. In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is 100% over the fourteen contiguous 5'-most nucleotides of the target sequence of the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 14 nucleotides in length (see Figure 12D-E). In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is 100% over the seven contiguous 5'-most nucleotides of the target sequence of the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 7 nucleotides in length.

Protein-binding segment of a DNA-targeting RNA

[00169] The protein-binding segment of a subject DNA-targeting RNA interacts with a site-directed modifying polypeptide. The subject DNA-targeting RNA guides the bound polypeptide to a specific nucleotide sequence within target DNA via the above mentioned DNA-targeting segment. The protein-binding segment of a subject DNA-targeting RNA comprises two stretches of nucleotides that are complementary to one another. The complementary nucleotides of the protein-binding segment hybridize to form a double stranded RNA duplex (dsRNA) (see Figures 1A and 1B).

[00170] A subject double-molecule DNA-targeting RNA comprises two separate RNA molecules. Each of the two RNA molecules of a subject double-molecule DNA-targeting RNA comprises a stretch of nucleotides that are complementary to one another such that the complementary nucleotides of the two RNA molecules hybridize to form the double stranded RNA duplex of the protein-binding segment (Figure 1A).

[00171] In some embodiments, the duplex-forming segment of the activator-RNA is at least about 60% identical to one of the activator-RNA (tracrRNA) molecules set forth in SEQ ID NOs:431-562, or a complement thereof, over a stretch of at least 8 contiguous nucleotides. For example, the duplex-forming segment of the activator-RNA (or the DNA encoding the duplex-forming segment of the activator-RNA) is at least about 60% identical, at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, or 100 % identical, to one of the tracrRNA sequences set forth in SEQ ID NOs:431-562, or a complement thereof, over a stretch of at least 8 contiguous nucleotides.

[00172] In some embodiments, the duplex-forming segment of the targeter-RNA is at least about 60% identical to one of the targeter-RNA (crRNA) seqeunces set forth in SEQ ID NOs:563-679, or a complement thereof, over a stretch of at least 8 contiguous nucleotides. For example, the duplex-forming segment of the targeter-RNA (or the DNA encoding the duplex-forming segment of the targeter-RNA) is at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical or 100 % identical to one of the crRNA sequences set forth in SEQ ID NOs:563-679, or a complement thereof, over a stretch of at least 8 contiguous nucleotides.

[00173] A two-molecule DNA-targeting RNA can be designed to allow for controlled (i.e., conditional) binding of a targeter-RNA with an activator-RNA. Because a two-molecule DNA-targeting RNA

is not functional unless both the activator-RNA and the targeter-RNA are bound in a functional complex with dCas9, a two-molecule DNA-targeting RNA can be inducible (e.g., drug inducible) by rendering the binding between the activator-RNA and the targeter-RNA to be inducible. As one non-limiting example, RNA aptamers can be used to regulate (i.e., control) the binding of the activator-RNA with the targeter-RNA. Accordingly, the activator-RNA and/or the targeter-RNA can comprise an RNA aptamer sequence.

[00174] RNA aptamers are known in the art and are generally a synthetic version of a riboswitch. The terms "RNA aptamer" and "riboswitch" are used interchangeably herein to encompass both synthetic and natural nucleic acid sequences that provide for inducible regulation of the structure (and therefore the availability of specific sequences) of the RNA molecule of which they are part. RNA aptamers usually comprise a sequence that folds into a particular structure (e.g., a hairpin), which specifically binds a particular drug (e.g., a small molecule). Binding of the drug causes a structural change in the folding of the RNA, which changes a feature of the nucleic acid of which the aptamer is a part. As non-limiting examples: (i) an activator-RNA with an aptamer may not be able to bind to the cognate targeter-RNA unless the aptamer is bound by the appropriate drug; (ii) a targeter-RNA with an aptamer may not be able to bind to the cognate activator-RNA unless the aptamer is bound by the appropriate drug; and (iii) a targeter-RNA and an activator-RNA, each comprising a different aptamer that binds a different drug, may not be able to bind to each other unless both drugs are present. As illustrated by these examples, a two-molecule DNA-targeting RNA can be designed to be inducible.

[00175] Examples of aptamers and riboswitches can be found, for example, in: Nakamura et al., Genes Cells. 2012 May;17(5):344-64; Vavalle et al., Future Cardiol. 2012 May;8(3):371-82; Citartan et al., Biosens Bioelectron. 2012 Apr 15;34(1):1-11; and Liberman et al., Wiley Interdiscip Rev RNA. 2012 May-Jun;3(3):369-84; all of which are herein incorporated by reference in their entirety.

[00176] Non-limiting examples of nucleotide sequences that can be included in a two-molecule DNA-targeting RNA include either of the sequences set forth in SEQ ID NOs:431-562, or complements thereof pairing with any sequences set forth in SEQ ID NOs:563-679, or complements thereof that can hybridize to form a protein binding segment.

[00177] A subject single-molecule DNA-targeting RNA comprises two stretches of nucleotides (a targeter-RNA and an activator-RNA) that are complementary to one another, are covalently linked by intervening nucleotides ("linkers" or "linker nucleotides"), and hybridize to form the double stranded RNA duplex (dsRNA duplex) of the protein-binding segment, thus resulting in a stem-loop structure (Figure 1B). The targeter-RNA and the activator-RNA can be covalently linked via the 3' end of the targeter-RNA and the 5' end of the activator-RNA. Alternatively,

targeter-RNA and the activator-RNA can be covalently linked via the 5' end of the targeter-RNA and the 3' end of the activator-RNA.

[00178] The linker of a single-molecule DNA-targeting RNA can have a length of from about 3 nucleotides to about 100 nucleotides. For example, the linker can have a length of from about 3 nucleotides (nt) to about 90 nt, from about 3 nucleotides (nt) to about 80 nt, from about 3 nucleotides (nt) to about 70 nt, from about 3 nucleotides (nt) to about 60 nt, from about 3 nucleotides (nt) to about 50 nt, from about 3 nucleotides (nt) to about 40 nt, from about 3 nucleotides (nt) to about 30 nt, from about 3 nucleotides (nt) to about 20 nt or from about 3 nucleotides (nt) to about 10 nt. For example, the linker can have a length of from about 3 nt to about 5 nt, from about 5 nt to about 10 nt, from about 10 nt to about 15 nt, from about 15 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. In some embodiments, the linker of a single-molecule DNA-targeting RNA is 4 nt.

[00179] An exemplary single-molecule DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to form a dsRNA duplex. In some embodiments, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 60% identical to one of the activator-RNA (tracrRNA) molecules set forth in SEQ ID NOs:431-562, or a complement thereof, over a stretch of at least 8 contiguous nucleotides. For example, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical or 100 % identical to one of the tracrRNA sequences set forth in SEQ ID NOs:431-562, or a complement thereof, over a stretch of at least 8 contiguous nucleotides.

[00180] In some embodiments, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 60% identical to one of the targeter-RNA (crRNA) sequences set forth in SEQ ID NOs:563-679, or a complement thereof, over a stretch of at least 8 contiguous nucleotides. For example, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99%

identical or 100 % identical to one of the crRNA sequences set forth in SEQ ID NOs:563-679, or a complement thereof, over a stretch of at least 8 contiguous nucleotides.

[00181] Appropriate naturally occurring cognate pairs of crRNAs and tracrRNAs can be routinely determined for SEQ ID NOs:431-679 by taking into account the speices name and base-pairing (for the dsRNA duplex of the protein-binding domain) when determining appropriate cognate pairs (see Figure 8 as a non-limiting example).

[00182] With regard to both a subject single-molecule DNA-targeting RNA and to a subject double-molecule DNA-targeting RNA, **Figure 57** demonstrates that artificial sequences that share very little (roughly 50% identity) with naturally occurring a tracrRNAs and crRNAs can function with Cas9 to cleave target DNA as long as the structure of the protein-binding domain of the DNA-targeting RNA is conserved. Thus, RNA folding structure of a naturally ocuring protein-binding domain of a DNA-trageting RNA can be taken into account in order to design artificial protein-binding domains (either two-molecule or single-molecule versions). As a non-limiting example, the functional artificial DNA-trageting RNA of Figure 57 was designed based on the structure of the protein-binding segment of the naturally occurring DNA-targeting (e.g., including the same number of base pairs along the RNA duplex and including the same "buldge" region as present in the naturally occurring RNA). As structures can readily be produced by one of ordinary skill in the art for any naturally occurring crRNA:tracrRNA pair from any speices (see SEQ ID NOs:431-679 for crRNA and tracrRNA sequences from a wide variety of species), an artificial DNA-targeting-RNA can be designed to mimic the natural structure for a given species when using the Cas9 (or a related Cas9, see Figure 32A) from that species. (see Figure 24D and related details in Example 1). Thus, a suitable DNA-targeting RNA can be an artificially designed RNA (non-naturally occurring) comprising a protein-binding domain that was desgined to mimic the structure of a protein-binding domain of a naturally occurring DNA-targeting RNA. (see SEQ ID NOs:431-679, taking into account the speices name when determining appropriate cognate pairs).

[00183] The protein-binding segment can have a length of from about 10 nucleotides to about 100 nucleotides. For example, the protein-binding segment can have a length of from about 15 nucleotides (nt) to about 80 nt, from about 15 nt to about 50 nt, from about 15 nt to about 40 nt, from about 15 nt to about 30 nt or from about 15 nt to about 25 nt.

[00184] Also with regard to both a subject single-molecule DNA-targeting RNA and to a subject double-molecule DNA-targeting RNA, the dsRNA duplex of the protein-binding segment can have a length from about 6 base pairs (bp) to about 50bp. For example, the dsRNA duplex of the protein-binding segment can have a length from about 6 bp to about 40 bp, from about 6 bp to about 30bp, from about 6 bp to about 25 bp, from about 6 bp to about 20 bp, from about 6 bp to

about 15 bp, from about 8 bp to about 40 bp, from about 8 bp to about 30bp, from about 8 bp to about 25 bp, from about 8 bp to about 20 bp or from about 8 bp to about 15 bp. For example, the dsRNA duplex of the protein-binding segment can have a length from about from about 8 bp to about 10 bp, from about 10 bp to about 15 bp, from about 15 bp to about 18 bp, from about 18 bp to about 20 bp, from about 20 bp to about 25 bp, from about 25 bp to about 30 bp, from about 30 bp to about 35 bp, from about 35 bp to about 40 bp, or from about 40 bp to about 50 bp. In some embodiments, the dsRNA duplex of the protein-binding segment has a length of 36 base pairs. The percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment can be at least about 60%. For example, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment can be at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, or at least about 99% . In some cases, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment is 100%.

Site-directed modifying polypeptide

[00185] A subject DNA-targeting RNA and a subject site-directed modifying polypeptide form a complex. The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA (as noted above). The site-directed modifying polypeptide of the complex provides the site-specific activity. In other words, the site-directed modifying polypeptide is guided to a DNA sequence (e.g. a chromosomal sequence or an extrachromosomal sequence, e.g. an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, etc.) by virtue of its association with at least the protein-binding segment of the DNA-targeting RNA (described above).

[00186] A subject site-directed modifying polypeptide modifies target DNA (e.g., cleavage or methylation of target DNA) and/or a polypeptide associated with target DNA (e.g., methylation or acetylation of a histone tail). A site-directed modifying polypeptide is also referred to herein as a "site-directed polypeptide" or an "RNA binding site-directed modifying polypeptide."

[00187] In some cases, the site-directed modifying polypeptide is a naturally-occurring modifying polypeptide. In other cases, the site-directed modifying polypeptide is not a naturally-occurring polypeptide (e.g., a chimeric polypeptide as discussed below or a naturally-occurring polypeptide that is modified, e.g., mutation, deletion, insertion).

[00188] Exemplary naturally-occurring site-directed modifying polypeptides are set forth in SEQ ID NOs:1-255 as a non-limiting and non-exhaustive list of naturally occurring Cas9/Csn1

endonucleases. These naturally occurring polypeptides, as disclosed herein, bind a DNA-targeting RNA, are thereby directed to a specific sequence within a target DNA, and cleave the target DNA to generate a double strand break. A subject site-directed modifying polypeptide comprises two portions, an RNA-binding portion and an activity portion. In some embodiments, a subject site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that exhibits site-directed enzymatic activity (e.g., activity for DNA methylation, activity for DNA cleavage, activity for histone acetylation, activity for histone methylation, etc.), wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00189] In other embodiments, a subject site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that modulates transcription within the target DNA (e.g., to increase or decrease transcription), wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00190] In some cases, a subject site-directed modifying polypeptide has enzymatic activity that modifies target DNA (e.g., nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity).

[00191] In other cases, a subject site-directed modifying polypeptide has enzymatic activity that modifies a polypeptide (e.g., a histone) associated with target DNA (e.g., methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity).

Exemplary site-directed modifying polypeptides

[00192] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

Nucleic acid modifications

[00193] In some embodiments, a subject nucleic acid (e.g., a DNA-targeting RNA) comprises one or more modifications, e.g., a base modification, a backbone modification, etc, to provide the nucleic acid with a new or enhanced feature (e.g., improved stability). As is known in the art, a nucleoside is a base-sugar combination. The base portion of the nucleoside is normally a heterocyclic base. The two most common classes of such heterocyclic bases are the purines and the pyrimidines. Nucleotides are nucleosides that further include a phosphate group covalently linked to the sugar portion of the nucleoside. For those nucleosides that include a pentofuranosyl sugar, the phosphate group can be linked to the 2', the 3', or the 5' hydroxyl moiety of the sugar. In forming oligonucleotides, the phosphate groups covalently link adjacent nucleosides to one another to form a linear polymeric compound. In turn, the respective ends of this linear polymeric compound can be further joined to form a circular compound, however, linear compounds are generally suitable. In addition, linear compounds may have internal nucleotide base complementarity and may therefore fold in a manner as to produce a fully or partially double-stranded compound. Within oligonucleotides, the phosphate groups are commonly referred to as forming the internucleoside backbone of the oligonucleotide. The normal linkage or backbone of RNA and DNA is a 3' to 5' phosphodiester linkage.

Modified backbones and modified internucleoside linkages

[00194] Examples of suitable nucleic acids containing modifications include nucleic acids containing modified backbones or non-natural internucleoside linkages. Nucleic acids (having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone.

[00195] Suitable modified oligonucleotide backbones containing a phosphorus atom therein include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Suitable oligonucleotides having inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage i.e. a single inverted nucleoside residue which may be a basic (the nucleobase is missing or has a hydroxyl group in place thereof). Various salts (such as, for example, potassium or sodium), mixed salts and free acid forms are also included.

[00196] In some embodiments, a subject nucleic acid comprises one or more phosphorothioate and/or heteroatom internucleoside linkages, in particular $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$ (known as a methylene (methylimino) or MMI backbone), $-CH_2-O-N(CH_3)-CH_2-$, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ and $-O-N(CH_3)-CH_2-CH_2-$ (wherein the native phosphodiester internucleotide linkage is represented as $-O-P(=O)(OH)-O-CH_2-$). MMI type internucleoside linkages are disclosed in the above referenced U.S. Pat. No. 5,489,677. Suitable amide internucleoside linkages are disclosed in t U.S. Pat. No. 5,602,240.

[00197] Also suitable are nucleic acids having morpholino backbone structures as described in, e.g., U.S. Pat. No. 5,034,506. For example, in some embodiments, a subject nucleic acid comprises a 6-membered morpholino ring in place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage replaces a phosphodiester linkage.

[00198] Suitable modified polynucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts.

Mimetics

[00199] A subject nucleic acid can be a nucleic acid mimetic. The term "mimetic" as it is applied to polynucleotides is intended to include polynucleotides wherein only the furanose ring or both the furanose ring and the internucleotide linkage are replaced with non-furanose groups, replacement of only the furanose ring is also referred to in the art as being a sugar surrogate. The heterocyclic base moiety or a modified heterocyclic base moiety is maintained for hybridization with an appropriate target nucleic acid. One such nucleic acid, a polynucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA, the sugar-backbone of a polynucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleotides are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[00200] One polynucleotide mimetic that has been reported to have excellent hybridization properties is a peptide nucleic acid (PNA). The backbone in PNA compounds is two or more linked aminoethylglycine units which gives PNA an amide containing backbone. The heterocyclic base moieties are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Representative U.S. patents that describe the preparation of PNA compounds include, but are not limited to: U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262.

[00201] Another class of polynucleotide mimetic that has been studied is based on linked morpholino units (morpholino nucleic acid) having heterocyclic bases attached to the morpholino ring. A number of linking groups have been reported that link the morpholino monomeric units in a morpholino nucleic acid. One class of linking groups has been selected to give a non-ionic oligomeric compound. The non-ionic morpholino-based oligomeric compounds are less likely to have undesired interactions with cellular proteins. Morpholino-based polynucleotides are non-ionic mimics of oligonucleotides which are less likely to form undesired interactions with cellular proteins (Dwaine A. Braasch and David R. Corey, Biochemistry, 2002, 41(14), 4503-4510). Morpholino-based polynucleotides are disclosed in U.S. Pat. No. 5,034,506. A variety of compounds within the morpholino class of polynucleotides have been prepared, having a variety of different linking groups joining the monomeric subunits.

[00202] A further class of polynucleotide mimetic is referred to as cyclohexenyl nucleic acids (CeNA). The furanose ring normally present in a DNA/RNA molecule is replaced with a cyclohexenyl ring. CeNA DMT protected phosphoramidite monomers have been prepared and used for oligomeric compound synthesis following classical phosphoramidite chemistry. Fully modified CeNA oligomeric compounds and oligonucleotides having specific positions modified with CeNA have been prepared and studied (see Wang et al., J. Am. Chem. Soc., 2000, 122, 8595-8602). In general the incorporation of CeNA monomers into a DNA chain increases its stability of a DNA/RNA hybrid. CeNA oligoadenylates formed complexes with RNA and DNA complements with similar stability to the native complexes. The study of incorporating CeNA structures into natural nucleic acid structures was shown by NMR and circular dichroism to proceed with easy conformational adaptation.

[00203] A further modification includes Locked Nucleic Acids (LNAs) in which the 2'-hydroxyl group is linked to the 4' carbon atom of the sugar ring thereby forming a 2'-C,4'-C-oxymethylene linkage thereby forming a bicyclic sugar moiety. The linkage can be a methylene (-CH_2-), group bridging the 2' oxygen atom and the 4' carbon atom wherein n is 1 or 2 (Singh et al., Chem. Commun., 1998, 4, 455-456). LNA and LNA analogs display very high duplex thermal stabilities with complementary DNA and RNA (Tm=+3 to +10° C), stability towards 3'-exonucleolytic degradation and good solubility properties. Potent and nontoxic antisense

oligonucleotides containing LNAs have been described (Wahlestedt et al., Proc. Natl. Acad. Sci. U.S.A., 2000, 97, 5633-5638).

[00204] The synthesis and preparation of the LNA monomers adenine, cytosine, guanine, 5-methyl-cytosine, thymine and uracil, along with their oligomerization, and nucleic acid recognition properties have been described (Koshkin et al., Tetrahedron, 1998, 54, 3607-3630). LNAs and preparation thereof are also described in WO 98/39352 and WO 99/14226.

Modified sugar moieties

[00205] A subject nucleic acid can also include one or more substituted sugar moieties. Suitable polynucleotides comprise a sugar substituent group selected from: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C.sub.1 to C_{10} alkyl or C_2 to C_{10} alkenyl and alkynyl. Particularly suitable are $O((CH_2)_nO)_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$, and $O(CH_2)_nON((CH_2)_nCH_3)_2$, where n and m are from 1 to about 10. Other suitable polynucleotides comprise a sugar substituent group selected from: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkenyl, alkynyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A suitable modification includes 2'-methoxyethoxy (2'-O-CH_2 CH_2OCH_3, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin et al., Helv. Chim. Acta, 1995, 78, 486-504) i.e., an alkoxyalkoxy group. A further suitable modification includes 2'-dimethylaminooxyethoxy, i.e., a $O(CH_2)_2ON(CH_3)_2$ group, also known as 2'-DMAOE, as described in examples hereinbelow, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethyl-amino-ethoxy-ethyl or 2'-DMAEOE), i.e., 2'-O-CH_2-O-CH_2-$N(CH_3)_2$.

[00206] Other suitable sugar substituent groups include methoxy (-O-CH_3), aminopropoxy (--O CH_2 CH_2 CH_2NH_2), allyl (-CH_2-CH=CH_2), -O-allyl (--O-- CH_2—CH=CH_2) and fluoro (F). 2'-sugar substituent groups may be in the arabino (up) position or ribo (down) position. A suitable 2'-arabino modification is 2'-F. Similar modifications may also be made at other positions on the oligomeric compound, particularly the 3' position of the sugar on the 3' terminal nucleoside or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligomeric compounds may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Base modifications and substitutions

[00207] A subject nucleic acid may also include nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl (-C=C-CH_3) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-amino-adenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further modified nucleobases include tricyclic pyrimidines such as phenoxazine cytidine(1H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g. 9-(2-aminoethoxy)-H-pyrimido(5,4-(b) (1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (H-pyrido(3',2':4,5)pyrrolo(2,3-d)pyrimidin-2-one).

[00208] Heterocyclic base moieties may also include those in which the purine or pyrimidine base is replaced with other heterocycles, for example 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine and 2-pyridone. Further nucleobases include those disclosed in U.S. Pat. No. 3,687,808, those disclosed in The Concise Encyclopedia Of Polymer Science And Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, those disclosed by Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613, and those disclosed by Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993. Certain of these nucleobases are useful for increasing the binding affinity of an oligomeric compound. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2° C. (Sanghvi et al., eds., Antisense Research and Applications, CRC Press, Boca Raton, 1993, pp. 276-278) and are suitable base substitutions, e.g., when combined with 2'-O-methoxyethyl sugar modifications.

Conjugates

[00209] Another possible modification of a subject nucleic acid involves chemically linking to the polynucleotide one or more moieties or conjugates which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. These moieties or conjugates can include conjugate groups covalently bound to functional groups such as primary or secondary hydroxyl groups. Conjugate groups include, but are not limited to, intercalators, reporter molecules, polyamines, polyamides, polyethylene glycols, polyethers, groups that enhance the pharmacodynamic properties of oligomers, and groups that enhance the pharmacokinetic properties of oligomers. Suitable conjugate groups include, but are not limited to, cholesterols, lipids, phospholipids, biotin, phenazine, folate, phenanthridine, anthraquinone, acridine, fluoresceins, rhodamines, coumarins, and dyes. Groups that enhance the pharmacodynamic properties include groups that improve uptake, enhance resistance to degradation, and/or strengthen sequence-specific hybridization with the target nucleic acid. Groups that enhance the pharmacokinetic properties include groups that improve uptake, distribution, metabolism or excretion of a subject nucleic acid.

[00210] Conjugate moieties include but are not limited to lipid moieties such as a cholesterol moiety (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86, 6553-6556), cholic acid (Manoharan et al., Bioorg. Med. Chem. Let., 1994, 4, 1053-1060), a thioether, e.g., hexyl-S-tritylthiol (Manoharan et al., Ann. N.Y. Acad. Sci., 1992, 660, 306-309; Manoharan et al., Bioorg. Med. Chem. Let., 1993, 3, 2765-2770), a thiocholesterol (Oberhauser et al., Nucl. Acids Res., 1992, 20, 533-538), an aliphatic chain, e.g., dodecandiol or undecyl residues (Saison-Behmoaras et al., EMBO J., 1991, 10, 1111-1118; Kabanov et al., FEBS Lett., 1990, 259, 327-330; Svinarchuk et al., Biochimie, 1993, 75, 49-54), a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654; Shea et al., Nucl. Acids Res., 1990, 18, 3777-3783), a polyamine or a polyethylene glycol chain (Manoharan et al., Nucleosides & Nucleotides, 1995, 14, 969-973), or adamantane acetic acid (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654), a palmityl moiety (Mishra et al., Biochim. Biophys. Acta, 1995, 1264, 229-237), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke et al., J. Pharmacol. Exp. Ther., 1996, 277, 923-937.\

[00211] A conjugate may include a "Protein Transduction Domain" or PTD (also known as a CPP – cell penetrating peptide), which may refer to a polypeptide, polynucleotide, carbohydrate, or organic or inorganic compound that facilitates traversing a lipid bilayer, micelle, cell membrane, organelle membrane, or vesicle membrane. A PTD attached to another molecule, which can range from a small polar molecule to a large macromolecule and/or a nanoparticle, facilitates the

molecule traversing a membrane, for example going from extracellular space to intracellular space, or cytosol to within an organelle. In some embodiments, a PTD is covalently linked to the amino terminus of an exogenous polypeptide (e.g., a site-directed modifying polypeptide). In some embodiments, a PTD is covalently linked to the carboxyl terminus of an exogenous polypeptide (e.g., a site-directed modifying polypeptide). In some embodiments, a PTD is covalently linked to a nucleic acid (e.g., a DNA-targeting RNA, a polynucleotide encoding a DNA-targeting RNA, a polynucleotide encoding a site-directed modifying polypeptide, etc.). Exemplary PTDs include but are not limited to a minimal undecapeptide protein transduction domain (corresponding to residues 47-57 of HIV-1 TAT comprising YGRKKRRQRRR; SEQ ID NO:264); a polyarginine sequence comprising a number of arginines sufficient to direct entry into a cell (e.g., 3, 4, 5, 6, 7, 8, 9, 10, or 10-50 arginines); a VP22 domain (Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); an Drosophila Antennapedia protein transduction domain (Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); a truncated human calcitonin peptide (Trehin et al. (2004) *Pharm. Research* 21:1248-1256); polylysine (Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR (SEQ ID NO:265); Transportan GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:266); KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:267); and RQIKIWFQNRRMKWKK (SEQ ID NO:268). Exemplary PTDs include but are not limited to, YGRKKRRQRRR (SEQ ID NO:264), RKKRRQRRR (SEQ ID NO:269); an arginine homopolymer of from 3 arginine residues to 50 arginine residues; Exemplary PTD domain amino acid sequences include, but are not limited to, any of the following: YGRKKRRQRRR (SEQ ID NO:264); RKKRRQRR (SEQ ID NO:270); YARAAARQARA (SEQ ID NO:271); THRLPRRRRRR (SEQ ID NO:272); and GGRRARRRRRR (SEQ ID NO:273). In some embodiments, the PTD is an activatable CPP (ACPP) (Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPPs comprise a polycationic CPP (e.g., Arg9 or "R9") connected via a cleavable linker to a matching polyanion (e.g., Glu9 or "E9"), which reduces the net charge to nearly zero and thereby inhibits adhesion and uptake into cells. Upon cleavage of the linker, the polyanion is released, locally unmasking the polyarginine and its inherent adhesiveness, thus "activating" the ACPP to traverse the membrane.

Exemplary DNA-targeting RNAs

[00212] In some embodiments, a suitable DNA-targeting RNA comprises two separate RNA polynucleotide molecules. The first of the two separate RNA polynucleotide molecules (the activator-RNA) comprises a nucleotide sequence having at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence

identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-562, or complements thereof. The second of the two separate RNA polynucleotide molecules (the targeter-RNA) comprises a nucleotide sequence having at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:563-679, or complements thereof.

[00213] In some embodiments, a suitable DNA-targeting RNA is a single RNA polynucleotide and comprises a first nucleotide sequence having at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs:431-562 and a second nucleotide sequence having at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 8 contiguous nucleotides to any one of the nucleotide sequences set forth in SEQ ID NOs: 463-679.

[00214] In some embodiments, the DNA-targeting RNA is a double-molecule DNA-targeting RNA and the targeter-RNA comprises the sequence 5'GUUUUAGAGCUA-3' (SEQ ID NO:679) linked at its 5' end to a stretch of nucleotides that are complementary to a target DNA. In some embodiments, the DNA-targeting RNA is a double-molecule DNA-targeting RNA and the activator-RNA comprises the sequence 5' UAGCAAGUUAAAAUAAGGCUAGUCCG-3' (SEQ ID NO://).

[00215] In some embodiments, the DNA-targeting RNA is a single-molecule DNA-targeting RNA and comprises the sequence 5'-GUUUUAGAGCUA-linker-UAGCAAGUUAAAAUAAGGCUAGUCCG-3' linked at its 5' end to a stretch of nucleotides that are complementary to a target DNA (where "linker" denotes any a linker nucleotide sequence that can comprise any nucleotide sequence) (SEQ ID NO://). Other exemplary single-molecule DNA-targeting RNAs include those set forth in SEQ ID NOs: 680-682.

Nucleic acids encoding a subject DNA-targeting RNA and/or a subject site-directed modifying polypeptide

[00216] The present disclosure provides a nucleic acid comprising a nucleotide sequence encoding a subject DNA-targeting RNA and/or a subject site-directed modifying polypeptide. In some embodiments, a subject DNA-targeting RNA-encoding nucleic acid is an expression vector, e.g., a recombinant expression vector.

[00217] In some embodiments, a subject method involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide. In some embodiments a cell comprising a target DNA is *in vitro.* In some embodiments a cell comprising a target DNA is *in vivo.* Suitable nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is a "recombinant expression vector."

[00218] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, a recombinant retroviral construct, etc.

[00219] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00220] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00221] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription

enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00222] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide in both prokaryotic and eukaryotic cells.

[00223] Non-limiting examples of suitable eukaryotic promoters (promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also include appropriate sequences for amplifying expression. The expression vector may also include nucleotide sequences encoding protein tags (e.g., 6xHis tag, hemagglutinin tag, green fluorescent protein, etc.) that are fused to the site-directed modifying polypeptide, thus resulting in a chimeric polypeptide.

[00224] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to an inducible promoter. In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to a constitutive promoter.

[00225] Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a cell. Suitable methods include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, e.g., Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like.

CHIMERIC POLYPEPTIDES

[00226] The present disclosure provides a chimeric site-directed modifying polypeptide. A subject chimeric site-directed modifying polypeptide interacts with (e.g., binds to) a subject DNA-targeting RNA (described above). The DNA-targeting RNA guides the chimeric site-directed modifying polypeptide to a target sequence within target DNA (e.g. a chromosomal sequence or an extrachromosomal sequence, e.g. an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, etc.). A subject chimeric site-directed modifying polypeptide modifies target DNA (e.g., cleavage or methylation of target DNA) and/or a polypeptide associated with target DNA (e.g., methylation or acetylation of a histone tail).

[00227] A subject chimeric site-directed modifying polypeptide modifies target DNA (e.g., cleavage or methylation of target DNA) and/or a polypeptide associated with target DNA (e.g., methylation or acetylation of a histone tail). A chimeric site-directed modifying polypeptide is also referred to herein as a "chimeric site-directed polypeptide" or a "chimeric RNA binding site-directed modifying polypeptide."

[00228] A subject chimeric site-directed modifying polypeptide comprises two portions, an RNA-binding portion and an activity portion. A subject chimeric site-directed modifying polypeptide comprises amino acid sequences that are derived from at least two different polypeptides. A subject chimeric site-directed modifying polypeptide can comprise modified and/or naturally-occurring polypeptide sequences (e.g., a first amino acid sequence from a modified or unmodified Cas9/Csn1 protein; and a second amino acid sequence other than the Cas9/Csn1 protein).

RNA-binding portion

[00229] In some cases, the RNA-binding portion of a subject chimeric site-directed modifying polypeptide is a naturally-occurring polypeptide. In other cases, the RNA-binding portion of a subject chimeric site-directed modifying polypeptide is not a naturally-occurring molecule (modified, e.g., mutation, deletion, insertion). Naturally-occurring RNA-binding portions of interest are derived from site-directed modifying polypeptides known in the art. For example, SEQ ID NOs:1-256 and 795-1346 provide a non-limiting and non-exhaustive list of naturally occurring Cas9/Csn1 endonucleases that can be used as site-directed modifying polypeptides. In some cases, the RNA-binding portion of a subject chimeric site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to the RNA-binding portion of a polypeptide having any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346).

[00230] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

Activity portion

[00231] In addition to the RNA-binding portion, the chimeric site-directed modifying polypeptide comprises an "activity portion." In some embodiments, the activity portion of a subject chimeric site-directed modifying polypeptide comprises the naturally-occurring activity portion of a site-directed modifying polypeptide (e.g., Cas9/Csn1 endonuclease). In other embodiments, the activity portion of a subject chimeric site-directed modifying polypeptide comprises a modified amino acid sequence (e.g., substitution, deletion, insertion) of a naturally-occurring activity portion of a site-directed modifying polypeptide. Naturally-occurring activity portions of interest are derived from site-directed modifying polypeptides known in the art. For example, SEQ ID NOs:1-256 and 795-1346 provide a non-limiting and non-exhaustive list of naturally occurring Cas9/Csn1 endonucleases that can be used as site-directed modifying polypeptides. The activity portion of a subject chimeric site-directed modifying polypeptide is variable and may comprise any heterologous polypeptide sequence that may be useful in the methods disclosed herein.

[00232] In some embodiments, a subject chimeric site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that exhibits site-directed enzymatic activity (e.g., activity for DNA methylation, activity for DNA cleavage, activity for histone acetylation, activity for histone methylation, etc.), wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00233] In other embodiments, a subject chimeric site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that modulates transcription within the target DNA (e.g., to increase or decrease transcription), wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00234] In some cases, the activity portion of a subject chimeric site-directed modifying polypeptide has enzymatic activity that modifies target DNA (e.g., nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity,

dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity).

[00235] In other cases, the activity portion of a subject chimeric site-directed modifying polypeptide has enzymatic activity (e.g., methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity) that modifies a polypeptide associated with target DNA (e.g., a histone).

[00236] In some cases, the activity portion of a subject chimeric site-directed modifying polypeptide exhibits enzymatic activity (described above). In other cases, the activity portion of a subject chimeric site-directed modifying polypeptide modulates transcription of the target DNA (described above). The activity portion of a subject chimeric site-directed modifying polypeptide is variable and may comprise any heterologous polypeptide sequence that may be useful in the methods disclosed herein.

Exemplary chimeric site-directed modifying polypeptides

[00237] In some embodiments, the activity portion of the chimeric site-directed modifying polypeptide comprises a modified form of the Cas9/Csn1 protein. In some instances, the modified form of the Cas9/Csn1 protein comprises an amino acid change (e.g., deletion, insertion, or substitution) that reduces the naturally-occurring nuclease activity of the Cas9/Csn1 protein. For example, in some instances, the modified form of the Cas9/Csn1 protein has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9/Csn1 polypeptide. In some cases, the modified form of the Cas9/Csn1 polypeptide has no substantial nuclease activity.

[00238] In some embodiements, the modified form of the Cas9/Csn1 polypeptide is a D10A (aspartate to alanine at amino acid position 10 of SEQ ID NO:8) mutation (or the corresponding mutation of any of the proteins presented in SEQ ID NOs:1-256 and 795-1346) that can cleave the complementary strand of the target DNA but has reduced ability to cleave the non-complementary strand of the target DNA (see Figure 11). In some embodiements, the modified form of the Cas9/Csn1 polypeptide is a H840A (histidine to alanine at amino acid position 840) mutation (or the corresponding mutation of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) that can cleave the non-complementary strand of the target DNA but has reduced ability to cleave the complementary strand of the target DNA (see Figure 11). In some embodiements, the modified form of the Cas9/Csn1 polypeptide harbors both the D10A and the

H840A mutations (or the corresponding mutations of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) such that the polypeptide has a reduced ability to cleave both the complementary and the non-complementary strands of the target DNA. Other residues can be mutated to achieve the above effects (i.e. inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) can be altered (i.e., substituted) (see **Figure 3, Figure 5, Figure 11A**, and **Table 1** for more information regarding the conservation of Cas9 amino acid residues). Also, mutations other than alanine substitutions are suitable. For more information of important

[00239] Table 1. Table 1 lists 4 motifs that are present in Cas9 sequences from various species (see also **Figure 3** and **Figure 5**). The amino acids listed here are from the Cas9 from *S. pyogenes* (SEQ ID NO:8).

Motif #	Motif	Amino acids (residue #s)	Highly conserved
1	RuvC-like I	IGLDIGTNSVGWAVI (7-21) (SEQ ID NO:260)	D10, G12, G17
2	RuvC-like II	IVIEMARE (759-766) (SEQ ID NO:261)	E762
3	HNH-motif	DVDHIVPQSFLKDDSIDNKVLTRSDKN (837-863) (SEQ ID NO:262)	H840, N854, N863
4	RuvC-like II	HHAHDAYL (982-989) (SEQ ID NO:263)	H982, H983, A984, D986, A987

[00240] In some cases, the chimeric site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the chimeric site-directed modifying polypeptide comprises 4 motifs (as listed in Table 4 and depicted in Figure 3A and Figure 5), each with amino acid sequences having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to each of the 4 motifs listed in Table 1(SEQ ID NOs:260-263), or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. In some cases, the chimeric site-directed modifying polypeptide comprises amino acid sequences having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino

acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

[00241] In some embodiments, the activity portion of the site-directed modifying polypeptide comprises a heterologous polypeptide that has DNA-modifying activity and/or transcription factor activity and/or DNA-associated polypeptide-modifying activity. In some cases, a heterologous polypeptide replaces a portion of the Cas9/Csn1 polypeptide that provides nuclease activity. In other embodiments, a subject site-directed modifying polypeptide comprises both a portion of the Cas9/Csn1 polypeptide that normally provides nuclease activity (and that portion can be fully active or can instead be modified to have less than 100% of the corresponding wild-type activity) and a heterologous polypeptide. In other words, in some cases, a subject chimeric site-directed modifying polypeptide is a fusion polypeptide comprising both the portion of the Cas9/Csn1 polypeptide that normally provides nuclease activity and the heterologous polypeptide. In other cases, a subject chimeric site-directed modifying polypeptide is a fusion polypeptide comprising a modified variant of the activity portion of the Cas9/Csn1 polypeptide (e.g., amino acid change, deletion, insertion) and a heterologous polypeptide. In yet other cases, a subject chimeric site-directed modifying polypeptide is a fusion polypeptide comprising a heterologous polypeptide and the RNA-binding portion of a naturally-occurring or a modified site-directed modifying polypeptide.

[00242] For example, in a chimeric Cas9/Csn1 protein, a naturally-occurring (or modified, e.g., mutation, deletion, insertion) bacterial Cas9/Csn1 polypeptide may be fused to a heterologous polypeptide sequence (i.e. a polypeptide sequence from a protein other than Cas9/Csn1 or a polypeptide sequence from another organism). The heterologous polypeptide sequence may exhibit an activity (e.g., enzymatic activity) that will also be exhibited by the chimeric Cas9/Csn1 protein (e.g., methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.). A heterologous nucleic acid sequence may be linked to another nucleic acid sequence (e.g., by genetic engineering) to generate a chimeric nucleotide sequence encoding a chimeric polypeptide. In some embodiments, a chimeric Cas9/Csn1 polypeptide is generated by fusing a Cas9/Csn1 polypeptide (e.g., wild type Cas9 or a Cas9 variant, e.g., a Cas9 with reduced or inactivated nuclease activity) with a heterologous sequence that provides for subcellular localization (e.g., a nuclear localization signal (NLS) for targeting to the nucleus; a mitochondrial localization signal for targeting to the mitochondria; a chloroplast localization signal for targeting to a chloroplast; an ER retention signal; and the like). In some embodiments, the heterologous sequence can provide a tag for ease of tracking or purification (e.g., a fluorescent protein, e.g., green fluorescent protein (GFP), YFP, RFP, CFP, mCherry, tdTomato, and the like; a HIS tag, e.g., a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and

the like). In some embodiments, the heterologous sequence can provide for increased or decreased stability. In some embodiments, the heterologous sequence can provide a binding domain (e.g., to provide the ability of a chimeric Cas9 polypeptide to bind to another protein of interest, e.g., a DNA or histone modifying protein, a transcription factor or transcription repressor, a recruiting protein, etc.).

[00243] Examples of various additional suitable fusion partners (or fragments thereof) for a subject variant Cas9 site-directed polypeptide include, but are not limited to those listed in **Figure 54**.

Nucleic acid encoding a subject chimeric site-directed modifying polypeptide

[00244] The present disclosure provides a nucleic acid comprising a nucleotide sequence encoding a subject chimeric site-directed modifying polypeptide. In some embodiments, the nucleic acid comprising a nucleotide sequence encoding a subject chimeric site-directed modifying polypeptide is an expression vector, e.g., a recombinant expression vector.

[00245] In some embodiments, a subject method involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising a chimeric site-directed modifying polypeptide. Suitable nucleic acids comprising nucleotide sequences encoding a chimeric site-directed modifying polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is a "recombinant expression vector."

[00246] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, etc.

[00247] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia

Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00248] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00249] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00250] In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a chimeric site-directed modifying polypeptide in both prokaryotic and eukaryotic cells.

[00251] Non-limiting examples of suitable eukaryotic promoters (promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also include appropriate sequences for amplifying expression. The expression vector may also include nucleotide sequences encoding protein tags (e.g., 6xHis tag, hemagglutinin (HA) tag, a fluorescent protein (e.g., a green fluorescent protein; a yellow fluorescent protein, etc.), etc.) that are fused to the chimeric site-directed modifying polypeptide.

[00252] In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to an inducible promoter (e.g., heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc.). In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to a spatially restricted and/or temporally

restricted promoter (e.g., a tissue specific promoter, a cell type specific promoter, etc.). In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to a constitutive promoter.

[00253] Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a stem cell or progenitor cell. Suitable methods include, include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, e.g., Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like.

METHODS

[00254] The present disclosure provides methods for modifying a target DNA and/or a target DNA-associated polypeptide. Generally, a subject method involves contacting a target DNA with a complex (a "targeting complex"), which complex comprises a DNA-targeting RNA and a site-directed modifying polypeptide.

[00255] As discussed above, a subject DNA-targeting RNA and a subject site-directed modifying polypeptide form a complex. The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-directed modifying polypeptide of the complex provides the site-specific activity. In some embodiments, a subject complex modifies a target DNA, leading to, for example, DNA cleavage, DNA methylation, DNA damage, DNA repair, etc. In other embodiments, a subject complex modifies a target polypeptide associated with target DNA (e.g., a histone, a DNA-binding protein, etc.), leading to, for example, histone methylation, histone acetylation, histone ubiquitination, and the like. The target DNA may be, for example, naked DNA *in vitro*, chromosomal DNA in cells *in vitro*, chromosomal DNA in cells *in vivo*, etc.

[00256] In some cases, the site-directed modifying polypeptide exhibits nuclease activity that cleaves target DNA at a target DNA sequence defined by the region of complementarity between the DNA-targeting RNA and the target DNA. In some cases, when the site-directed modifying polypeptide is a Cas9 or Cas9 related polypeptide, site-specific cleavage of the target DNA occurs at locations determined by both (i) base-pairing complementarity between the DNA-targeting RNA and the target DNA; and (ii) a short motif [referred to as the protospacer adjacent motif (PAM)] in the target DNA. In some embodiments (e.g., when Cas9 from S. pyogenes, or a

closely related Cas9, is used (see SEQ ID NOs:1-256 and 795-1346)), the PAM sequence of the non-complementary strand is 5'-XGG-3', where X is any DNA nucleotide and X is immediately 3' of the target sequence of the non-complementary strand of the target DNA (see Figure 10). As such, the PAM sequence of the complementary strand is 5'-CCY-3', where Y is any DNA nucleotide and Y is immediately 5' of the target sequence of the complementary strand of the target DNA (see Figure 10 where the PAM of the non-complementary strand is 5'-GGG-3' and the PAM of the complementary strand is 5'-CCC-3'). In some such embodiments, X and Y can be complementary and the X-Y base pair can be any basepair (e.g., X=C and Y=G; X=G and Y=C; X=A and Y=T, X=T and Y=A).

[00257] In some cases, different Cas9 proteins (i.e., Cas9 proteins from various species) may be advantageous to use in the various provided methods in order to capitalize on various enzymatic characteristics of the different Cas9 proteins (e.g., for different PAM sequence preferences; for increased or decreased enzymatic activity; for an increased or decreased level of cellular toxicity; to change the balance between NHEJ, homology-directed repair, single strand breaks, double strand breaks, etc.). Cas9 proteins from various species (see SEQ ID NOs:1-256 and 795-1346) may require different PAM sequences in the target DNA. Thus, for a particular Cas9 protein of choice, the PAM sequence requirement may be different than the 5'-XGG-3' sequence described above.

[00258] Many Cas9 orthologus from a wide variety of species have been identified herein and the protiens share only a few identical amino acids. All identified Cas9 orthologs have the same domain architecture with a central HNH endonuclease domain and a split RuvC/RNaseH domain (See Figures 3A, 3B, Figure 5, and Table 1). Cas9 proteins share 4 key motifs with a conserved architecture. Motifs 1, 2, and 4 are RuvC like motifs while motif 3 is an HNH-motif. In some cases, a suitable site-directed modifying polypeptide comprises an amino acid sequence having 4 motifs, each of motifs 1-4 having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to the motif s 1-4 of the Cas9/Csn1 amino acid sequence depicted in Figure 3A (SEQ ID NOs:260-263, respectively, as depicted in Table 1), or to the corresponding portions in any of the amino acid sequences set forth in SEQ ID NOs:1-256 and 795-1346 (see Figure 5 for an alignment of motifs 1-4 from divergent Cas9 sequences). In some cases, a suitable site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346. Any Cas9 protein as defined above can be used as a

site-directed modifying polypeptide or as part of a chimeric site-directed modifying polypeptide of the subject methods.

[00259] The nuclease activity cleaves target DNA to produce double strand breaks. These breaks are then repaired by the cell in one of two ways: non-homologous end joining, and homology-directed repair (Figure 2). In non-homologous end joining (NHEJ), the double-strand breaks are repaired by direct ligation of the break ends to one another. As such, no new nucleic acid material is inserted into the site, although some nucleic acid material may be lost, resulting in a deletion. In homology-directed repair, a donor polynucleotide with homology to the cleaved target DNA sequence is used as a template for repair of the cleaved target DNA sequence, resulting in the transfer of genetic information from the donor polynucleotide to the target DNA. As such, new nucleic acid material may be inserted/copied into the site. In some cases, a target DNA is contacted with a subject donor polynucleotide. In some cases, a subject donor polynucleotide is introduced into a subject cell. The modifications of the target DNA due to NHEJ and/or homology-directed repair lead to, for example, gene correction, gene replacement, gene tagging, transgene insertion, nucleotide deletion, gene disruption, gene mutation, etc.

[00260] Accordingly, cleavage of DNA by a site-directed modifying polypeptide may be used to delete nucleic acid material from a target DNA sequence (e.g., to disrupt a gene that makes cells susceptible to infection (e.g. the CCR5 or CXCR4 gene, which makes T cells susceptible to HIV infection), to remove disease-causing trinucleotide repeat sequences in neurons, to create gene knockouts and mutations as disease models in research, etc.) by cleaving the target DNA sequence and allowing the cell to repair the sequence in the absence of an exogenously provided donor polynucleotide. Thus, the subject methods can be used to knock out a gene (resulting in complete lack of transcription or altered transcription) or to knock in genetic material into a locus of choice in the target DNA.

[00261] Alternatively, if a DNA-targeting RNA and a site-directed modifying polypeptide are coadministered to cells with a donor polynucleotide sequence that includes at least a segment with homology to the target DNA sequence, the subject methods may be used to add, i.e. insert or replace, nucleic acid material to a target DNA sequence (e.g. to "knock in" a nucleic acid that encodes for a protein, an siRNA, an miRNA, etc.), to add a tag (e.g., 6xHis, a fluorescent protein (e.g., a green fluorescent protein; a yellow fluorescent protein, etc.), hemagglutinin (HA), FLAG, etc.), to add a regulatory sequence to a gene (e.g. promoter, polyadenylation signal, internal ribosome entry sequence (IRES), 2A peptide, start codon, stop codon, splice signal, localization signal, etc.), to modify a nucleic acid sequence (e.g., introduce a mutation), and the like. As such, a complex comprising a DNA-targeting RNA and a site-directed modifying polypeptide is useful in any *in vitro* or *in vivo* application in which it is desirable to modify DNA in a site-

specific, i.e. "targeted", way, for example gene knock-out, gene knock-in, gene editing, gene tagging, etc., as used in, for example, gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, the production of genetically modified organisms in agriculture, the large scale production of proteins by cells for therapeutic, diagnostic, or research purposes, the induction of iPS cells, biological research, the targeting of genes of pathogens for deletion or replacement, etc.

[00262] In some embodiments, the site-directed modifying polypeptide comprises a modified form of the Cas9/Csn1 protein. In some instances, the modified form of the Cas9/Csn1 protein comprises an amino acid change (e.g., deletion, insertion, or substitution) that reduces the naturally-occurring nuclease activity of the Cas9/Csn1 protein. For example, in some instances, the modified form of the Cas9/Csn1 protein has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9/Csn1 polypeptide. In some cases, the modified form of the Cas9/Csn1 polypeptide has no substantial nuclease activity. When a subject site-directed modifying polypeptide is a modified form of the Cas9/Csn1 polypeptide that has no substantial nuclease activity, it can be referred to as "dCas9."

[00263] In some embodiments, the modified form of the Cas9/Csn1 polypeptide is a D10A (aspartate to alanine at amino acid position 10 of SEQ ID NO:8) mutation (or the corresponding mutation of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) that can cleave the complementary strand of the target DNA but has reduced ability to cleave the non-complementary strand of the target DNA (thus resulting in a single strand break (SSB) instead of a DSB; see Figure 11). In some embodiments, the modified form of the Cas9/Csn1 polypeptide is a H840A (histidine to alanine at amino acid position 840 of SEQ ID NO:8) mutation (or the corresponding mutation of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) that can cleave the non-complementary strand of the target DNA but has reduced ability to cleave the complementary strand of the target DNA (thus resulting in a single strand break (SSB) instead of a DSB; see Figure 11). The use of the D10A or H840A variant of Cas9 (or the corresponding mutations in any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) can alter the expected biological outcome because the non-homologous end joining (NHEJ) is much more likely to occur when DSBs are present as opposed to SSBs. Thus, in some cases where one wishes to reduce the likelihood of DSB (and therefore reduce the likelihood of NHEJ), a D10A or H840A variant of Cas9 can be used. Other residues can be mutated to achieve the same effect (i.e. inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the proteins set forth as SEQ ID NOs:1-256 and 795-

1346) can be altered (i.e., substituted) (see **Figure 3, Figure 5, Figure 11A**, and **Table 1** for more information regarding the conservation of Cas9 amino acid residues). Also, mutations other than alanine substitutions are suitable. In some embodiments when a site-directed polypeptide (e.g., site-directred modifying polypeptide) has reduced catalytic activity (e.g., when a Cas9 protein has a D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or a A987 mutation, e.g., D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A), the polypeptide can still bind to target DNA in a site-specific manner (because it is still guided to a target DNA sequence by a DNA-targeting RNA) as long as it retains the ability to interact with the DNA-targeting RNA.

[00264] In some embodiments, the modified form of the Cas9/Csn1 polypeptide harbors both the D10A and the H840A mutations (or the corresponding mutations of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) such that the polypeptide has a reduced ability to cleave both the complementary and the non-complementary strands of the target DNA (i.e., the variant can have no substantial nuclease activity). Other residues can be mutated to achieve the same effect (i.e. inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) can be altered (i.e., substituted) (see **Figure 3, Figure 5, Figure 11A**, and **Table 1** for more information regarding the conservation of Cas9 amino acid residues). Also, mutations other than alanine substitutions are suitable.

[00265] In some embodiments, the site-directed modifying polypeptide comprises a heterologous sequence (e.g., a fusion). In some embodiments, a heterologous sequence can provide for subcellular localization of the site-directed modifying polypeptide (e.g., a nuclear localization signal (NLS) for targeting to the nucleus; a mitochondrial localization signal for targeting to the mitochondria; a chloroplast localization signal for targeting to a chloroplast; a ER retention signal; and the like). In some embodiments, a heterologous sequence can provide a tag for ease of tracking or purification (e.g., a fluorescent protein, e.g., green fluorescent protein (GFP), YFP, RFP, CFP, mCherry, tdTomato, and the like; a his tag, e.g., a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and the like). In some embodiments, the heterologous sequence can provide for increased or decreased stability.

[00266] In some embodiments, a subject site-directed modifying polypeptide can be codon-optimized. This type of optimization is known in the art and entails the mutation of foreign-derived DNA to mimic the codon preferences of the intended host organism or cell while encoding the same protein. Thus, the codons are changed, but the encoded protein remains unchanged. For example, if the intended target cell was a human cell, a human codon-optimized Cas9 (or variant, e.g.,

enzymatically inactive variant) would be a suitable site-directed modifying polypeptide (see SEQ ID NO:256 for an example). Any suitable site-directed modifying polypeptide (e.g., any Cas9 such as any of the sequences set forth in SEQ ID NOs:1-256 and 795-1346) can be codon optimized. As another non-limiting example, if the intended host cell were a mouse cell, than a mouse codon-optimized Cas9 (or variant, e.g., enzymatically inactive variant) would be a suitable site-directed modifying polypeptide. While codon optimization is not required, it is acceptable and may be preferable in certain cases.

[00267] In some embodiments, a subject DNA-targeting RNA and a subject site-directed modifying polypeptide are used as an inducible system for shutting off gene expression in bacterial cells. In some cases, nucleic acids encoding an appropriate DNA-targeting RNA and/or an appropriate site-directed polypeptide are incorporated into the chromosome of a target cell and are under control of an inducible promoter. When the DNA-targeting RNA and/or the site-directed polypeptide are induced, the target DNA is cleaved (or otherwise modified) at the location of interest (e.g., a target gene on a separate plasmid), when both the DNA-targeting RNA and the site-directed modifying polypeptide are present and form a complex. As such, in some cases, bacterial expression strains are engineered to include nucleic acid sequences encoding an appropriate site-directed modifying polypeptide in the bacterial genome and/or an appropriate DNA-targeting RNA on a plasmid (e.g., under control of an inducible promoter), allowing experiments in which the expression of any targeted gene (expressed from a separate plasmid introduced into the strain) could be controlled by inducing expression of the DNA-targeting RNA and the site-directed polypeptide.

[00268] In some cases, the site-directed modifying polypeptide has enzymatic activity that modifies target DNA in ways other than introducing double strand breaks. Enzymatic activity of interest that may be used to modify target DNA (e.g., by fusing a heterologous polypeptide with enzymatic activity to a site-directed modifying polypeptide, thereby generating a chimeric site-directed modifying polypeptide) includes, but is not limited methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity). Methylation and demethylation is recognized in the art as an important mode of epigenetic gene regulation while DNA damage and repair activity is essential for cell survival and for proper genome maintenance in response to environmental stresses.

[00269] As such, the methods herein find use in the epigenetic modification of target DNA and may be employed to control epigenetic modification of target DNA at any location in a target DNA by

genetically engineering the desired complementary nucleic acid sequence into the DNA-targeting segment of a DNA-targeting RNA. The methods herein also find use in the intentional and controlled damage of DNA at any desired location within the target DNA. The methods herein also find use in the sequence-specific and controlled repair of DNA at any desired location within the target DNA. Methods to target DNA-modifying enzymatic activities to specific locations in target DNA find use in both research and clinical applications.

[00270] In some cases, the site-directed modifying polypeptide has activity that modulates the transcription of target DNA (e.g., in the case of a chimeric site-directed modifying polypeptide, etc.). In some cases, a chimeric site-directed modifying polypeptides comprising a heterologous polypeptide that exhibits the ability to increase or decrease transcription (e.g., transcriptional activator or transcription repressor polypeptides) is used to increase or decrease the transcription of target DNA at a specific location in a target DNA, which is guided by the DNA-targeting segment of the DNA-targeting RNA. Examples of source polypeptides for providing a chimeric site-directed modifying polypeptide with transcription modulatory activity include, but are not limited to light-inducible transcription regulators, small molecule/drug-responsive transcription regulators, transcription factors, transcription repressors, etc. In some cases, the subject method is used to control the expression of a targeted coding-RNA (protein-encoding gene) and/or a targeted non-coding RNA (e.g., tRNA, rRNA, snoRNA, siRNA, miRNA, long ncRNA, etc.).

[00271] In some cases, the site-directed modifying polypeptide has enzymatic activity that modifies a polypeptide associated with DNA (e.g. histone). In some embodiments, the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity (i.e., ubiquitination activity), deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, demyristoylation activity glycosylation activity (e.g., from O-GlcNAc transferase) or deglycosylation activity. The enzymatic activities listed herein catalyze covalent modifications to proteins. Such modifications are known in the art to alter the stability or activity of the target protein (e.g., phosphorylation due to kinase activity can stimulate or silence protein activity depending on the target protein). Of particular interest as protein targets are histones. Histone proteins are known in the art to bind DNA and form complexes known as nucleosomes. Histones can be modified (e.g., by methylation, acetylation, ubuitination, phosphorylation) to elicit structural changes in the surrounding DNA, thus controlling the accessibility of potentially large portions of DNA to interacting factors such as transcription factors, polymerases and the like. A single histone can be modified in many different ways and in many different combinations (e.g., trimethylation of lysine 27 of histone 3, H3K27, is associated with DNA regions of repressed

transcription while trimethylation of lysine 4 of histone 3, H3K4, is associated with DNA regions of active transcription). Thus, a site-directed modifying polypeptide with histone-modifying activity finds use in the site specific control of DNA structure and can be used to alter the histone modification pattern in a selected region of target DNA. Such methods find use in both research and clinical applications.

[00272] In some embodiments, multiple DNA-targeting RNAs are used simultaneously to simultaneously modify different locations on the same target DNA or on different target DNAs. In some embodiments, two or more DNA-targeting RNAs target the same gene or transcript or locus. In some embodiments, two or more DNA-targeting RNAs target different unrelated loci. In some embodiments, two or more DNA-targeting RNAs target different, but related loci.

[00273] In some cases, the site-directed modifying polypeptide is provided directly as a protein. As one non-limiting example, fungi (e.g., yeast) can be transformed with exogenous protein and/or nucleic acid using spheroplast transformation (see Kawai et al., Bioeng Bugs. 2010 Nov-Dec;1(6):395-403 : "Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism"; and Tanka et al., Nature. 2004 Mar 18;428(6980):323-8: "Conformational variations in an infectious protein determine prion strain differences"; both of which are herein incorporated by reference in their entirety). Thus, a site-directed modifying polypeptide (e.g., Cas9) can be incorporated into a spheroplast (with or without nucleic acid encoding a DNA-targeting RNA and with or without a donor polynucleotide) and the spheroplast can be used to introduce the content into a yeast cell. A site-directed modifying polypeptide can be introduced into a cell (provided to the cell) by any convenient method; such methods are known to those of ordinary skill in the art. As another non-limiting example, a site-directed modifying polypeptide can be injected directly into a cell (e.g., with or without nucleic acid encoding a DNA-targeting RNA and with or without a donor polynucleotide), e.g., a cell of a zebrafish embryo, the pronucleus of a fertilized mouse oocyte, etc.

Target cells of interest

[00274] In some of the above applications, the subject methods may be employed to induce DNA cleavage, DNA modification, and/or transcriptional modulation in mitotic or post-mitotic cells *in vivo* and/or *ex vivo* and/or *in vitro* (e.g., to produce genetically modified cells that can be reintroduced into an individual). Because the DNA-targeting RNA provide specificity by hybridizing to target DNA, a mitotic and/or post-mitotic cell of interest in the disclosed methods may include a cell from any organism (e.g. a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens C. Agardh*, and the like, a fungal cell (e.g., a yeast cell), an animal cell, a cell from an invertebrate animal

(e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal, a cell from a rodent, a cell from a human, etc.).

[00275] Any type of cell may be of interest (e.g. a stem cell, e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell; a somatic cell, e.g. a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an in vitro or in vivo embryonic cell of an embryo at any stage, e.g., a 1-cell, 2-cell, 4-cell, 8-cell, etc. stage zebrafish embryo; etc.). Cells may be from established cell lines or they may be primary cells, where "primary cells", "primary cell lines", and "primary cultures" are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, i.e. splittings, of the culture. For example, primary cultures are cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Typically, the primary cell lines of the present invention are maintained for fewer than 10 passages in vitro. Target cells are in many embodiments unicellular organisms, or are grown in culture.

[00276] If the cells are primary cells, they may be harvest from an individual by any convenient method. For example, leukocytes may be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. are most conveniently harvested by biopsy. An appropriate solution may be used for dispersion or suspension of the harvested cells. Such solution will generally be a balanced salt solution, e.g. normal saline, phosphate-buffered saline (PBS), Hank's balanced salt solution, etc., conveniently supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration, generally from 5-25 mM. Convenient buffers include HEPES, phosphate buffers, lactate buffers, etc. The cells may be used immediately, or they may be stored, frozen, for long periods of time, being thawed and capable of being reused. In such cases, the cells will usually be frozen in 10% DMSO, 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures, and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

Nucleic acids encoding a subject DNA-targeting RNA and/or a subject site-directed modifying polypeptide

[00277] In some embodiments, a subject method involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a donor polynucleotide. Suitable nucleic acids comprising nucleotide sequences encoding a DNA-

targeting RNA and/or a site-directed modifying polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is a "recombinant expression vector."

[00278] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, etc.

[00279] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00280] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00281] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a mammalian cell, or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide in both prokaryotic and eukaryotic cells.

[00282] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (e.g., U6 promoter, H1 promoter, etc.; see above) (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00283] In some embodiments, a DNA-targeting RNA and/or a site-directed modifying polypeptide can be provided as RNA. In such cases, the DNA-targeting RNA and/or the RNA encoding the site-directed modifying polypeptide can be produced by direct chemical synthesis or may be transcribed *in vitro* from a DNA encoding the DNA-targeting RNA. Methods of synthesizing RNA from a DNA template are well known in the art. In some cases, the DNA-targeting RNA and/or the RNA encoding the site-directed modifying polypeptide will be synthesized *in vitro* using an RNA polymerase enzyme (e.g., T7 polymerase, T3 polymerase, SP6 polymerase, etc.). Once synthesized, the RNA may directly contact a target DNA or may be introduced into a cell by any of the well-known techniques for introducing nucleic acids into cells (e.g., microinjection, electroporation, transfection, etc).

[00284] Nucleotides encoding a DNA-targeting RNA (introduced either as DNA or RNA) and/or a site-directed modifying polypeptide (introduced as DNA or RNA) and/or a donor polynucleotide may be provided to the cells using well-developed transfection techniques; see, e.g. Angel and Yanik (2010) PLoS ONE 5(7): e11756, and the commercially available TransMessenger® reagents from Qiagen, Stemfect™ RNA Transfection Kit from Stemgent, and TransIT®-mRNA Transfection Kit from Mirus Bio LLC. See also Beumer et al. (2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. PNAS 105(50):19821-19826. Alternatively, nucleic acids encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide may be provided on DNA vectors. Many vectors, e.g. plasmids, cosmids, minicircles, phage, viruses, etc., useful for transferring nucleic acids into target cells are available. The vectors comprising the nucleic acid(s) may be maintained episomally, e.g. as plasmids, minicircle DNAs, viruses such cytomegalovirus, adenovirus, etc., or they may be integrated into the target cell genome, through homologous recombination or random integration, e.g. retrovirus-derived vectors such as MMLV, HIV-1, ALV, etc.

[00285] Vectors may be provided directly to the subject cells. In other words, the cells are contacted with vectors comprising the nucleic acid encoding DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide such that the vectors are taken up by the cells. Methods for contacting cells with nucleic acid vectors that are plasmids, including electroporation, calcium chloride transfection,

microinjection, and lipofection are well known in the art. For viral vector delivery, the cells are contacted with viral particles comprising the nucleic acid encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide. Retroviruses, for example, lentiviruses, are particularly suitable to the method of the invention. Commonly used retroviral vectors are "defective", i.e. unable to produce viral proteins required for productive infection. Rather, replication of the vector requires growth in a packaging cell line. To generate viral particles comprising nucleic acids of interest, the retroviral nucleic acids comprising the nucleic acid are packaged into viral capsids by a packaging cell line. Different packaging cell lines provide a different envelope protein (ecotropic, amphotropic or xenotropic) to be incorporated into the capsid, this envelope protein determining the specificity of the viral particle for the cells (ecotropic for murine and rat; amphotropic for most mammalian cell types including human, dog and mouse; and xenotropic for most mammalian cell types except murine cells). The appropriate packaging cell line may be used to ensure that the cells are targeted by the packaged viral particles. Methods of introducing the retroviral vectors comprising the nucleic acid encoding the reprogramming factors into packaging cell lines and of collecting the viral particles that are generated by the packaging lines are well known in the art. Nucleic acids can also introduced by direct micro-injection (e.g., injection of RNA into a zebrafish embryo).

[00286] Vectors used for providing the nucleic acids encoding DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide to the subject cells will typically comprise suitable promoters for driving the expression, that is, transcriptional activation, of the nucleic acid of interest. In other words, the nucleic acid of interest will be operably linked to a promoter. This may include ubiquitously acting promoters, for example, the CMV-β-actin promoter, or inducible promoters, such as promoters that are active in particular cell populations or that respond to the presence of drugs such as tetracycline. By transcriptional activation, it is intended that transcription will be increased above basal levels in the target cell by at least about 10 fold, by at least about 100 fold, more usually by at least about 1000 fold. In addition, vectors used for providing a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide to the subject cells may include nucleic acid sequences that encode for selectable markers in the target cells, so as to identify cells that have taken up the DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide.

[00287] A subject DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide may instead be used to contact DNA or introduced into cells

as RNA. Methods of introducing RNA into cells are known in the art and may include, for example, direct injection, transfection, or any other method used for the introduction of DNA.

[00288] A subject site-directed modifying polypeptide may instead be provided to cells as a polypeptide. Such a polypeptide may optionally be fused to a polypeptide domain that increases solubility of the product. The domain may be linked to the polypeptide through a defined protease cleavage site, e.g. a TEV sequence, which is cleaved by TEV protease. The linker may also include one or more flexible sequences, e.g. from 1 to 10 glycine residues. In some embodiments, the cleavage of the fusion protein is performed in a buffer that maintains solubility of the product, e.g. in the presence of from 0.5 to 2 M urea, in the presence of polypeptides and/or polynucleotides that increase solubility, and the like. Domains of interest include endosomolytic domains, e.g. influenza HA domain; and other polypeptides that aid in production, e.g. IF2 domain, GST domain, GRPE domain, and the like. The polypeptide may be formulated for improved stability. For example, the peptides may be PEGylated, where the polyethyleneoxy group provides for enhanced lifetime in the blood stream.

[00289] Additionally or alternatively, the subject site-directed modifying polypeptide may be fused to a polypeptide permeant domain to promote uptake by the cell. A number of permeant domains are known in the art and may be used in the non-integrating polypeptides of the present invention, including peptides, peptidomimetics, and non-peptide carriers. For example, a permeant peptide may be derived from the third alpha helix of Drosophila melanogaster transcription factor Antennapaedia, referred to as penetratin, which comprises the amino acid sequence RQIKIWFQNRRMKWKK (SEQ ID NO://). As another example, the permeant peptide comprises the HIV-1 tat basic region amino acid sequence, which may include, for example, amino acids 49-57 of naturally-occurring tat protein. Other permeant domains include poly-arginine motifs, for example, the region of amino acids 34-56 of HIV-1 rev protein, nona-arginine, octa-arginine, and the like. (See, for example, Futaki et al. (2003) Curr Protein Pept Sci. 2003 Apr; 4(2): 87-9 and 446; and Wender et al. (2000) Proc. Natl. Acad. Sci. U.S.A 2000 Nov. 21; 97(24):13003-8; published U.S. Patent applications 20030220334; 20030083256; 20030032593; and 20030022831, herein specifically incorporated by reference for the teachings of translocation peptides and peptoids). The nona-arginine (R9) sequence is one of the more efficient PTDs that have been characterized (Wender et al. 2000; Uemura et al. 2002). The site at which the fusion is made may be selected in order to optimize the biological activity, secretion or binding characteristics of the polypeptide. The optimal site will be determined by routine experimentation.

[00290] A subject site-directed modifying polypeptide may be produced *in* vitro or by eukaryotic cells or by prokaryotic cells, and it may be further processed by unfolding, e.g. heat denaturation, DTT reduction, etc. and may be further refolded, using methods known in the art.

[00291] Modifications of interest that do not alter primary sequence include chemical derivatization of polypeptides, e.g., acylation, acetylation, carboxylation, amidation, etc. Also included are modifications of glycosylation, e.g. those made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps; e.g. by exposing the polypeptide to enzymes which affect glycosylation, such as mammalian glycosylating or deglycosylating enzymes. Also embraced are sequences that have phosphorylated amino acid residues, e.g. phosphotyrosine, phosphoserine, or phosphothreonine.

[00292] Also included in the subject invention are DNA-targeting RNAs and site-directed modifying polypeptides that have been modified using ordinary molecular biological techniques and synthetic chemistry so as to improve their resistance to proteolytic degradation, to change the target sequence specificity, to optimize solubility properties, to alter protein activity (e.g., transcription modulatory activity, enzymatic activity, etc) or to render them more suitable as a therapeutic agent. Analogs of such polypeptides include those containing residues other than naturally occurring L-amino acids, e.g. D-amino acids or non-naturally occurring synthetic amino acids. D-amino acids may be substituted for some or all of the amino acid residues.

[00293] The site-directed modifying polypeptides may be prepared by in vitro synthesis, using conventional methods as known in the art. Various commercial synthetic apparatuses are available, for example, automated synthesizers by Applied Biosystems, Inc., Beckman, etc. By using synthesizers, naturally occurring amino acids may be substituted with unnatural amino acids. The particular sequence and the manner of preparation will be determined by convenience, economics, purity required, and the like.

[00294] If desired, various groups may be introduced into the peptide during synthesis or during expression, which allow for linking to other molecules or to a surface. Thus cysteines can be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like.

[00295] The site-directed modifying polypeptides may also be isolated and purified in accordance with conventional methods of recombinant synthesis. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. For the most part, the compositions which are used will comprise at least 20% by weight of the desired product, more usually at least about 75% by weight, preferably at least about 95% by weight, and for therapeutic purposes, usually at

least about 99.5% by weight, in relation to contaminants related to the method of preparation of the product and its purification. Usually, the percentages will be based upon total protein.

[00296] To induce DNA cleavage and recombination, or any desired modification to a target DNA, or any desired modification to a polypeptide associated with target DNA, the DNA-targeting RNA and/or the site-directed modifying polypeptide and/or the donor polynucleotide, whether they be introduced as nucleic acids or polypeptides, are provided to the cells for about 30 minutes to about 24 hours, e.g., 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 12 hours, 16 hours, 18 hours, 20 hours, or any other period from about 30 minutes to about 24 hours, which may be repeated with a frequency of about every day to about every 4 days, e.g., every 1.5 days, every 2 days, every 3 days, or any other frequency from about every day to about every four days. The agent(s) may be provided to the subject cells one or more times, e.g. one time, twice, three times, or more than three times, and the cells allowed to incubate with the agent(s) for some amount of time following each contacting event e.g. 16-24 hours, after which time the media is replaced with fresh media and the cells are cultured further.

[00297] In cases in which two or more different targeting complexes are provided to the cell (e.g., two different DNA-targeting RNAs that are complementary to different sequences within the same or different target DNA), the complexes may be provided simultaneously (e.g. as two polypeptides and/or nucleic acids), or delivered simultaneously. Alternatively, they may be provided consecutively, e.g. the targeting complex being provided first, followed by the second targeting complex, etc. or vice versa.

[00298] Typically, an effective amount of the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide is provided to the target DNA or cells to induce cleavage. An effective amount of the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide is the amount to induce a 2--fold increase or more in the amount of target modification observed between two homologous sequences relative to a negative control, e.g. a cell contacted with an empty vector or irrelevant polypeptide. That is to say, an effective amount or dose of the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide will induce a 2-fold increase, a 3-fold increase, a 4-fold increase or more in the amount of target modification observed at a target DNA region, in some instances a 5-fold increase, a 6-fold increase or more, sometimes a 7-fold or 8-fold increase or more in the amount of recombination observed, e.g. an increase of 10-fold, 50-fold, or 100-fold or more, in some instances, an increase of 200-fold, 500-fold, 700-fold, or 1000-fold or more, e.g. a 5000-fold, or 10,000-fold increase in the amount of recombination observed. The amount of target modification may be measured by any convenient method. For example, a silent reporter construct comprising complementary sequence to the targeting segment (targeting

sequence) of the DNA-targeting RNA flanked by repeat sequences that, when recombined, will reconstitute a nucleic acid encoding an active reporter may be cotransfected into the cells, and the amount of reporter protein assessed after contact with the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide, e.g. 2 hours, 4 hours, 8 hours, 12 hours, 24 hours, 36 hours, 48 hours, 72 hours or more after contact with the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide. As another, more sensitivity assay, for example, the extent of recombination at a genomic DNA region of interest comprising target DNA sequences may be assessed by PCR or Southern hybridization of the region after contact with a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide, e.g. 2 hours, 4 hours, 8 hours, 12 hours, 24 hours, 36 hours, 48 hours, 72 hours or more after contact with the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide.

[00299] Contacting the cells with a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may occur in any culture media and under any culture conditions that promote the survival of the cells. For example, cells may be suspended in any appropriate nutrient medium that is convenient, such as Iscove's modified DMEM or RPMI 1640, supplemented with fetal calf serum or heat inactivated goat serum (about 5-10%), L-glutamine, a thiol, particularly 2-mercaptoethanol, and antibiotics, e.g. penicillin and streptomycin. The culture may contain growth factors to which the cells are responsive. Growth factors, as defined herein, are molecules capable of promoting survival, growth and/or differentiation of cells, either in culture or in the intact tissue, through specific effects on a transmembrane receptor. Growth factors include polypeptides and non-polypeptide factors. Conditions that promote the survival of cells are typically permissive of nonhomologous end joining and homology-directed repair.

[00300] In applications in which it is desirable to insert a polynucleotide sequence into a target DNA sequence, a polynucleotide comprising a donor sequence to be inserted is also provided to the cell. By a "donor sequence" or "donor polynucleotide" it is meant a nucleic acid sequence to be inserted at the cleavage site induced by a site-directed modifying polypeptide. The donor polynucleotide will contain sufficient homology to a genomic sequence at the cleavage site, e.g. 70%, 80%, 85%, 90%, 95%, or 100% homology with the nucleotide sequences flanking the cleavage site, e.g. within about 50 bases or less of the cleavage site, e.g. within about 30 bases, within about 15 bases, within about 10 bases, within about 5 bases, or immediately flanking the cleavage site, to support homology-directed repair between it and the genomic sequence to which it bears homology. Approximately 25, 50, 100, or 200 nucleotides, or more than 200 nucleotides, of sequence homology between a donor and a genomic sequence (or any integral value between 10 and 200 nucleotides, or more) will support homology-directed repair. Donor

sequences can be of any length, e.g. 10 nucleotides or more, 50 nucleotides or more, 100 nucleotides or more, 250 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, etc.

[00301] The donor sequence is typically not identical to the genomic sequence that it replaces. Rather, the donor sequence may contain at least one or more single base changes, insertions, deletions, inversions or rearrangements with respect to the genomic sequence, so long as sufficient homology is present to support homology-directed repair. In some embodiments, the donor sequence comprises a non-homologous sequence flanked by two regions of homology, such that homology-directed repair between the target DNA region and the two flanking sequences results in insertion of the non-homologous sequence at the target region. Donor sequences may also comprise a vector backbone containing sequences that are not homologous to the DNA region of interest and that are not intended for insertion into the DNA region of interest. Generally, the homologous region(s) of a donor sequence will have at least 50% sequence identity to a genomic sequence with which recombination is desired. In certain embodiments, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or 99.9% sequence identity is present. Any value between 1% and 100% sequence identity can be present, depending upon the length of the donor polynucleotide.

[00302] The donor sequence may comprise certain sequence differences as compared to the genomic sequence, e.g. restriction sites, nucleotide polymorphisms, selectable markers (e.g., drug resistance genes, fluorescent proteins, enzymes etc.), etc., which may be used to assess for successful insertion of the donor sequence at the cleavage site or in some cases may be used for other purposes (e.g., to signify expression at the targeted genomic locus). In some cases, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (i.e., changes which do not affect the structure or function of the protein). Alternatively, these sequences differences may include flanking recombination sequences such as FLPs, loxP sequences, or the like, that can be activated at a later time for removal of the marker sequence.

[00303] The donor sequence may be provided to the cell as single-stranded DNA, single-stranded RNA, double-stranded DNA, or double-stranded RNA. It may be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence may be protected (e.g., from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. See, for example, Chang et al. (1987) Proc. Natl. Acad Sci USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of

modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues. As an alternative to protecting the termini of a linear donor sequence, additional lengths of sequence may be included outside of the regions of homology that can be degraded without impacting recombination. A donor sequence can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor sequences can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (e.g., adenovirus, AAV), as described above for nucleic acids encoding a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide.

[00304] Following the methods described above, a DNA region of interest may be cleaved and modified, i.e. "genetically modified", ex vivo. In some embodiments, as when a selectable marker has been inserted into the DNA region of interest, the population of cells may be enriched for those comprising the genetic modification by separating the genetically modified cells from the remaining population. Prior to enriching, the "genetically modified" cells may make up only about 1% or more (e.g., 2% or more, 3% or more, 4% or more, 5% or more, 6% or more, 7% or more, 8% or more, 9% or more, 10% or more, 15% or more, or 20% or more) of the cellular population. Separation of "genetically modified" cells may be achieved by any convenient separation technique appropriate for the selectable marker used. For example, if a fluorescent marker has been inserted, cells may be separated by fluorescence activated cell sorting, whereas if a cell surface marker has been inserted, cells may be separated from the heterogeneous population by affinity separation techniques, e.g. magnetic separation, affinity chromatography, "panning" with an affinity reagent attached to a solid matrix, or other convenient technique. Techniques providing accurate separation include fluorescence activated cell sorters, which can have varying degrees of sophistication, such as multiple color channels, low angle and obtuse light scattering detecting channels, impedance channels, etc. The cells may be selected against dead cells by employing dyes associated with dead cells (e.g. propidium iodide). Any technique may be employed which is not unduly detrimental to the viability of the genetically modified cells. Cell compositions that are highly enriched for cells comprising modified DNA are achieved in this manner. By "highly enriched", it is meant that the genetically modified cells will be 70% or more, 75% or more, 80% or more, 85% or more, 90% or more of the cell composition, for example, about 95% or more, or 98% or more of the cell composition. In other words, the composition may be a substantially pure composition of genetically modified cells.

[00305] Genetically modified cells produced by the methods described herein may be used immediately. Alternatively, the cells may be frozen at liquid nitrogen temperatures and stored for long periods of time, being thawed and capable of being reused. In such cases, the cells will usually be frozen in 10% dimethylsulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures, and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

[00306] The genetically modified cells may be cultured in vitro under various culture conditions. The cells may be expanded in culture, i.e. grown under conditions that promote their proliferation. Culture medium may be liquid or semi-solid, e.g. containing agar, methylcellulose, etc. The cell population may be suspended in an appropriate nutrient medium, such as Iscove's modified DMEM or RPMI 1640, normally supplemented with fetal calf serum (about 5-10%), L-glutamine, a thiol, particularly 2-mercaptoethanol, and antibiotics, e.g. penicillin and streptomycin. The culture may contain growth factors to which the regulatory T cells are responsive. Growth factors, as defined herein, are molecules capable of promoting survival, growth and/or differentiation of cells, either in culture or in the intact tissue, through specific effects on a transmembrane receptor. Growth factors include polypeptides and non-polypeptide factors.

[00307] Cells that have been genetically modified in this way may be transplanted to a subject for purposes such as gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, for the production of genetically modified organisms in agriculture, or for biological research. The subject may be a neonate, a juvenile, or an adult. Of particular interest are mammalian subjects. Mammalian species that may be treated with the present methods include canines and felines; equines; bovines; ovines; etc. and primates, particularly humans. Animal models, particularly small mammals (e.g. mouse, rat, guinea pig, hamster, lagomorpha (e.g., rabbit), etc.) may be used for experimental investigations.

[00308] Cells may be provided to the subject alone or with a suitable substrate or matrix, e.g. to support their growth and/or organization in the tissue to which they are being transplanted. Usually, at least 1x103 cells will be administered, for example 5x103 cells, 1x104 cells, 5x104 cells, 1x105 cells, 1 x 106 cells or more. The cells may be introduced to the subject via any of the following routes: parenteral, subcutaneous, intravenous, intracranial, intraspinal, intraocular, or into spinal fluid. The cells may be introduced by injection, catheter, or the like. Examples of methods for local delivery, that is, delivery to the site of injury, include, e.g. through an Ommaya reservoir, e.g. for intrathecal delivery (see e.g. US Patent Nos. 5,222,982 and 5385582, incorporated herein by reference); by bolus injection, e.g. by a syringe, e.g. into a joint; by continuous infusion, e.g. by cannulation, e.g. with convection (see e.g. US Application No. 20070254842, incorporated

here by reference); or by implanting a device upon which the cells have been reversably affixed (see e.g. US Application Nos. 20080081064 and 20090196903, incorporated herein by reference). Cells may also be introduced into an embryo (e.g., a blastocyst) for the purpose of generating a transgenic animal (e.g., a transgenic mouse).

[00309] The number of administrations of treatment to a subject may vary. Introducing the genetically modified cells into the subject may be a one-time event; but in certain situations, such treatment may elicit improvement for a limited period of time and require an on-going series of repeated treatments. In other situations, multiple administrations of the genetically modified cells may be required before an effect is observed. The exact protocols depend upon the disease or condition, the stage of the disease and parameters of the individual subject being treated.

[00310] In other aspects of the invention, the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide are employed to modify cellular DNA in vivo, again for purposes such as gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, for the production of genetically modified organisms in agriculture, or for biological research. In these in vivo embodiments, a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide are administered directly to the individual. A DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may be administered by any of a number of well-known methods in the art for the administration of peptides, small molecules and nucleic acids to a subject. A DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide can be incorporated into a variety of formulations. More particularly, a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide of the present invention can be formulated into pharmaceutical compositions by combination with appropriate pharmaceutically acceptable carriers or diluents.

[00311] Pharmaceutical preparations are compositions that include one or more a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide present in a pharmaceutically acceptable vehicle. "Pharmaceutically acceptable vehicles" may be vehicles approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, such as humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is formulated for administration to a mammal. Such pharmaceutical vehicles can be lipids, e.g. liposomes, e.g. liposome dendrimers; liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like, saline; gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents

may be used. Pharmaceutical compositions may be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants, gels, microspheres, and aerosols. As such, administration of the a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide can be achieved in various ways, including oral, buccal, rectal, parenteral, intraperitoneal, intradermal, transdermal, intratracheal, intraocular, etc., administration. The active agent may be systemic after administration or may be localized by the use of regional administration, intramural administration, or use of an implant that acts to retain the active dose at the site of implantation. The active agent may be formulated for immediate activity or it may be formulated for sustained release.

[00312] For some conditions, particularly central nervous system conditions, it may be necessary to formulate agents to cross the blood-brain barrier (BBB). One strategy for drug delivery through the blood-brain barrier (BBB) entails disruption of the BBB, either by osmotic means such as mannitol or leukotrienes, or biochemically by the use of vasoactive substances such as bradykinin. The potential for using BBB opening to target specific agents to brain tumors is also an option. A BBB disrupting agent can be co-administered with the therapeutic compositions of the invention when the compositions are administered by intravascular injection. Other strategies to go through the BBB may entail the use of endogenous transport systems, including Caveolin-1 mediated transcytosis, carrier-mediated transporters such as glucose and amino acid carriers, receptor-mediated transcytosis for insulin or transferrin, and active efflux transporters such as p-glycoprotein. Active transport moieties may also be conjugated to the therapeutic compounds for use in the invention to facilitate transport across the endothelial wall of the blood vessel. Alternatively, drug delivery of therapeutics agents behind the BBB may be by local delivery, for example by intrathecal delivery, e.g. through an Ommaya reservoir (see e.g. US Patent Nos. 5,222,982 and 5385582, incorporated herein by reference); by bolus injection, e.g. by a syringe, e.g. intravitreally or intracranially; by continuous infusion, e.g. by cannulation, e.g. with convection (see e.g. US Application No. 20070254842, incorporated here by reference); or by implanting a device upon which the agent has been reversably affixed (see e.g. US Application Nos. 20080081064 and 20090196903, incorporated herein by reference).

[00313] Typically, an effective amount of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide are provided. As discussed above with regard to ex vivo methods, an effective amount or effective dose of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide in vivo is the amount to induce a 2 fold increase or more in the amount of recombination observed between two homologous sequences relative to a negative control, e.g. a cell contacted with an empty vector or irrelevant

polypeptide. The amount of recombination may be measured by any convenient method, e.g. as described above and known in the art. The calculation of the effective amount or effective dose of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide to be administered is within the skill of one of ordinary skill in the art, and will be routine to those persons skilled in the art. The final amount to be administered will be dependent upon the route of administration and upon the nature of the disorder or condition that is to be treated.

[00314] The effective amount given to a particular patient will depend on a variety of factors, several of which will differ from patient to patient. A competent clinician will be able to determine an effective amount of a therapeutic agent to administer to a patient to halt or reverse the progression the disease condition as required. Utilizing LD50 animal data, and other information available for the agent, a clinician can determine the maximum safe dose for an individual, depending on the route of administration. For instance, an intravenously administered dose may be more than an intrathecally administered dose, given the greater body of fluid into which the therapeutic composition is being administered. Similarly, compositions which are rapidly cleared from the body may be administered at higher doses, or in repeated doses, in order to maintain a therapeutic concentration. Utilizing ordinary skill, the competent clinician will be able to optimize the dosage of a particular therapeutic in the course of routine clinical trials.

[00315] For inclusion in a medicament, a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may be obtained from a suitable commercial source. As a general proposition, the total pharmaceutically effective amount of the a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide administered parenterally per dose will be in a range that can be measured by a dose response curve.

[00316] Therapies based on a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotides, i.e. preparations of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide to be used for therapeutic administration, must be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 μm membranes). Therapeutic compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle. The therapies based on a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-mL vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous solution of compound, and the resulting

mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized compound using bacteriostatic Water-for-Injection.

[00317] Pharmaceutical compositions can include, depending on the formulation desired, pharmaceutically-acceptable, non-toxic carriers of diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, buffered water, physiological saline, PBS, Ringer's solution, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation can include other carriers, adjuvants, or non-toxic, nontherapeutic, nonimmunogenic stabilizers, excipients and the like. The compositions can also include additional substances to approximate physiological conditions, such as pH adjusting and buffering agents, toxicity adjusting agents, wetting agents and detergents.

[00318] The composition can also include any of a variety of stabilizing agents, such as an antioxidant for example. When the pharmaceutical composition includes a polypeptide, the polypeptide can be complexed with various well-known compounds that enhance the in vivo stability of the polypeptide, or otherwise enhance its pharmacological properties (e.g., increase the half-life of the polypeptide, reduce its toxicity, enhance solubility or uptake). Examples of such modifications or complexing agents include sulfate, gluconate, citrate and phosphate. The nucleic acids or polypeptides of a composition can also be complexed with molecules that enhance their in vivo attributes. Such molecules include, for example, carbohydrates, polyamines, amino acids, other peptides, ions (e.g., sodium, potassium, calcium, magnesium, manganese), and lipids.

[00319] Further guidance regarding formulations that are suitable for various types of administration can be found in Remington's Pharmaceutical Sciences, Mace Publishing Company, Philadelphia, Pa., 17th ed. (1985). For a brief review of methods for drug delivery, see, Langer, Science 249:1527-1533 (1990).

[00320] The pharmaceutical compositions can be administered for prophylactic and/or therapeutic treatments. Toxicity and therapeutic efficacy of the active ingredient can be determined according to standard pharmaceutical procedures in cell cultures and/or experimental animals, including, for example, determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Therapies that exhibit large therapeutic indices are preferred.

[00321] The data obtained from cell culture and/or animal studies can be used in formulating a range of dosages for humans. The dosage of the active ingredient typically lines within a range of circulating concentrations that include the ED50 with low toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized.

[00322] The components used to formulate the pharmaceutical compositions are preferably of high purity and are substantially free of potentially harmful contaminants (e.g., at least National Food (NF) grade, generally at least analytical grade, and more typically at least pharmaceutical grade). Moreover, compositions intended for in vivo use are usually sterile. To the extent that a given compound must be synthesized prior to use, the resulting product is typically substantially free of any potentially toxic agents, particularly any endotoxins, which may be present during the synthesis or purification process. Compositions for parental administration are also sterile, substantially isotonic and made under GMP conditions.

[00323] The effective amount of a therapeutic composition to be given to a particular patient will depend on a variety of factors, several of which will differ from patient to patient. A competent clinician will be able to determine an effective amount of a therapeutic agent to administer to a patient to halt or reverse the progression the disease condition as required. Utilizing LD50 animal data, and other information available for the agent, a clinician can determine the maximum safe dose for an individual, depending on the route of administration. For instance, an intravenously administered dose may be more than an intrathecally administered dose, given the greater body of fluid into which the therapeutic composition is being administered. Similarly, compositions which are rapidly cleared from the body may be administered at higher doses, or in repeated doses, in order to maintain a therapeutic concentration. Utilizing ordinary skill, the competent clinician will be able to optimize the dosage of a particular therapeutic in the course of routine clinical trials.

GENETICALLY MODIFIED HOST CELLS

[00324] The present disclosure provides genetically modified host cells, including isolated genetically modified host cells, where a subject genetically modified host cell comprises (has been genetically modified with: 1) an exogenous DNA-targeting RNA; 2) an exogenous nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; 3) an exogenous site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.); 4) an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide; or 5) any combination of the above. A subject genetically modified cell is generated by genetically modifying a host cell with, for example: 1) an exogenous DNA-targeting RNA; 2) an exogenous nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; 3) an exogenous site-directed modifying

polypeptide; 4) an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide; or 5) any combination of the above.).

[00325] All cells suitable to be a target cell are also suitable to be a genetically modified host cell. For example, a genetically modified host cells of interest can be a cell from any organism (e.g. a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens C. Agardh*, and the like, a fungal cell (e.g., a yeast cell), an animal cell, a cell from an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., a pig, a cow, a goat, a sheep, a rodent, a rat, a mouse, a non-human primate, a human, etc.), etc.

[00326] In some embodiments, a genetically modified host cell has been genetically modified with an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.). The DNA of a genetically modified host cell can be targeted for modification by introducing into the cell a DNA-targeting RNA (or a DNA encoding a DNA-targeting RNA, which determines the genomic location/sequence to be modified) and optionally a donor nucleic acid. In some embodiments, the nucleotide sequence encoding a site-directed modifying polypeptide is operably linked to an inducible promoter (e.g., heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc.). In some embodiments, the nucleotide sequence encoding a site-directed modifying polypeptide is operably linked to a spatially restricted and/or temporally restricted promoter (e.g., a tissue specific promoter, a cell type specific promoter, etc.). In some embodiments, the nucleotide sequence encoding a site-directed modifying polypeptide is operably linked to a constitutive promoter.

[00327] In some embodiments, a subject genetically modified host cell is *in vitro*. In some embodiments, a subject genetically modified host cell is *in vivo*. In some embodiments, a subject genetically modified host cell is a prokaryotic cell or is derived from a prokaryotic cell. In some embodiments, a subject genetically modified host cell is a bacterial cell or is derived from a bacterial cell. In some embodiments, a subject genetically modified host cell is an archaeal cell or is derived from an archaeal cell. In some embodiments, a subject genetically modified host cell is a eukaryotic cell or is derived from a eukaryotic cell. In some embodiments, a subject genetically modified host cell is a plant cell or is derived from a plant cell. In some embodiments, a subject genetically modified host cell is an animal cell or is derived from an animal cell. In some embodiments, a subject genetically modified host cell is an invertebrate cell

or is derived from an invertebrate cell. In some embodiments, a subject genetically modified host cell is a vertebrate cell or is derived from a vertebrate cell. In some embodiments, a subject genetically modified host cell is a mammalian cell or is derived from a mammalian cell. In some embodiments, a subject genetically modified host cell is a rodent cell or is derived from a rodent cell. In some embodiments, a subject genetically modified host cell is a human cell or is derived from a human cell.

[00328] The present disclosure further provides progeny of a subject genetically modified cell, where the progeny can comprise the same exogenous nucleic acid or polypeptide as the subject genetically modified cell from which it was derived. The present disclosure further provides a composition comprising a subject genetically modified host cell.

Genetically modified stem cells and genetically modified progenitor cells

[00329] In some embodiments, a subject genetically modified host cell is a genetically modified stem cell or progenitor cell. Suitable host cells include, e.g., stem cells (adult stem cells, embryonic stem cells, iPS cells, etc.) and progenitor cells (e.g., cardiac progenitor cells, neural progenitor cells, etc.). Suitable host cells include mammalian stem cells and progenitor cells, including, e.g., rodent stem cells, rodent progenitor cells, human stem cells, human progenitor cells, etc. Suitable host cells include *in vitro* host cells, e.g., isolated host cells.

[00330] In some embodiments, a subject genetically modified host cell comprises an exogenous DNA-targeting RNA nucleic acid. In some embodiments, a subject genetically modified host cell comprises an exogenous nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA. In some embodiments, a subject genetically modified host cell comprises an exogenous site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.). In some embodiments, a subject genetically modified host cell comprises an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide. In some embodiments, a subject genetically modified host cell comprises exogenous nucleic acid comprising a nucleotide sequence encoding 1) a DNA-targeting RNA and 2) a site-directed modifying polypeptide.

[00331] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

COMPOSITIONS

[00332] The present invention provides a composition comprising a subject DNA-targeting RNA and/or a site-directed modifying polypeptide. In some cases, the site-directed modifying polypeptide is a subject chimeric polypeptide. A subject composition is useful for carrying out a method of the present disclosure, e.g., a method for site-specific modification of a target DNA; a method for site-specific modification of a polypeptide associated with a target DNA; etc.

Compositions comprising a DNA-targeting RNA

[00333] The present invention provides a composition comprising a subject DNA-targeting RNA. The composition can comprise, in addition to the DNA-targeting RNA, one or more of: a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), 2-(N-Morpholino)ethanesulfonic acid (MES), MES sodium salt, 3-(N-Morpholino)propanesulfonic acid (MOPS), N-tris[Hydroxymethyl]methy1-3-aminopropanesulfonic acid (TAPS), etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a nuclease inhibitor; and the like. For example, in some cases, a subject composition comprises a subject DNA-targeting RNA and a buffer for stabilizing nucleic acids.

[00334] In some embodiments, a DNA-targeting RNA present in a subject composition is pure, e.g., at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or more than 99% pure, where "% purity" means that DNA-targeting RNA is the recited percent free from other macromolecules, or contaminants that may be present during the production of the DNA-targeting RNA.

Compositions comprising a subject chimeric polypeptide

[00335] The present invention provides a composition a subject chimeric polypeptide. The composition can comprise, in addition to the DNA-targeting RNA, one or more of: a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, HEPES, MES, MES sodium salt, MOPS, TAPS, etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a protease inhibitor; a reducing agent (e.g., dithiothreitol); and the like.

[00336] In some embodiments, a subject chimeric polypeptide present in a subject composition is pure, e.g., at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or more than 99% pure, where "% purity" means that the site-directed modifying polypeptide is the recited percent free from other proteins, other macromolecules, or contaminants that may be present during the production of the chimeric polypeptide.

Compositions comprising a DNA-targeting RNA and a site-directed modifying polypeptide

[00337] The present invention provides a composition comprising: (i) a DNA-targeting RNA or a DNA polynucleotide encoding the same; and ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same. In some cases, the site-directed modifying polypeptide is a subject chimeric site-directed modifying polypeptide. In other cases, the site-directed modifying polypeptide is a naturally-occurring site-directed modifying polypeptide. In some instances, the site-directed modifying polypeptide exhibits enzymatic activity that modifies a target DNA. In other cases, the site-directed modifying polypeptide exhibits enzymatic activity that modifies a polypeptide that is associated with a target DNA. In still other cases, the site-directed modifying polypeptide modulates transcription of the target DNA.

[00338] The present invention provides a composition comprising: (i) a DNA-targeting RNA, as described above, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00339] In some instances, a subject composition comprises: a composition comprising: (i) a subject DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00340] In other embodiments, a subject composition comprises: (i) a polynucleotide encoding a subject DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a polynucleotide encoding the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00341] In some embodiments, a subject composition includes both RNA molecules of a double-molecule DNA-targeting RNA. As such, in some embodiments, a subject composition includes an activator-RNA that comprises a duplex-forming segment that is complementary to the duplex-forming segment of a targeter-RNA (see Figure 1A). The duplex-forming segments of the activator-RNA and the targeter-RNA hybridize to form the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. The targeter-RNA further provides the DNA-targeting segment (single stranded) of the DNA-targeting RNA and therefore targets the DNA-targeting RNA to a specific sequence within the target DNA. As one non-limiting example, the duplex-forming segment of the activator-RNA comprises a nucleotide sequence that has at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 98%, or 100% identity with the sequence 5'-UAGCAAGUUAAAAU-3' (SEQ ID NO:562). As another non-limiting example, the duplex-forming segment of the targeter-RNA comprises a nucleotide sequence that has at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 98%, or 100% identity with the sequence 5'-GUUUUAGAGCUA-3' (SEQ ID NO:679).

[00342] The present disclosure provides a composition comprising: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00343] For example, in some cases, a subject composition comprises: (i) a DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00344] As another example, in some cases, a subject composition comprises: (i) a DNA polynucleotide encoding a DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b)

a second segment that interacts with a site-directed modifying polypeptide; and (ii) a polynucleotide encoding the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00345] A subject composition can comprise, in addition to i) a subject DNA-targeting RNA, or a DNA polynucleotide encoding the same; and ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, one or more of: a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, HEPES, MES, MES sodium salt, MOPS, TAPS, etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a protease inhibitor; a reducing agent (e.g., dithiothreitol); and the like.

[00346] In some cases, the components of the composition are individually pure, e.g., each of the components is at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or at least 99%, pure. In some cases, the individual components of a subject composition are pure before being added to the composition.

[00347] For example, in some embodiments, a site-directed modifying polypeptide present in a subject composition is pure, e.g., at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or more than 99% pure, where "% purity" means that the site-directed modifying polypeptide is the recited percent free from other proteins (e.g., proteins other than the site-directed modifying polypeptide), other macromolecules, or contaminants that may be present during the production of the site-directed modifying polypeptide.

KITS

[00348] The present disclosure provides kits for carrying out a subject method. A subject kit can include one or more of: a site-directed modifying polypeptide; a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide; a DNA-targeting RNA; a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; an activator-RNA; a nucleic acid comprising a nucleotide sequence encoding an activator-RNA; a targeter-RNA; and a nucleic acid comprising a nucleotide sequence encoding a targeter-RNA. A site-directed modifying polypeptide; a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide; a DNA-targeting RNA; a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; an activator-RNA; a nucleic acid comprising a nucleotide sequence encoding an activator-RNA; a targeter-RNA; and a nucleic acid comprising a nucleotide sequence encoding a targeter-RNA, are described in detail above. A kit may comprise a complex that comprises two or more of: a site-directed modifying polypeptide; a nucleic acid

comprising a nucleotide encoding a site-directed modifying polypeptide; a DNA-targeting RNA; a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; an activator-RNA; a nucleic acid comprising a nucleotide sequence encoding an activator-RNA; a targeter-RNA; and a nucleic acid comprising a nucleotide sequence encoding a targeter-RNA.

[00349] In some embodiments, a subject kit comprises a site-directed modifying polypeptide, or a polynucleotide encoding the same. In some embodiments, the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA. In some cases, the activity portion of the site-directed modifying polypeptide exhibits reduced or inactivated nuclease activity. In some cases, the site-directed modifying polypeptide is a chimeric site-directed modifying polypeptide.

[00350] In some embodiments, a subject kit comprises: a site-directed modifying polypeptide, or a polynucleotide encoding the same, and a reagent for reconstituting and/or diluting the site-directed modifying polypeptide. In other embodiments, a subject kit comprises a nucleic acid (e.g., DNA, RNA) comprising a nucleotide encoding a site-directed modifying polypeptide. In some embodiments, a subject kit comprises: a nucleic acid (e.g., DNA, RNA) comprising a nucleotide encoding a site-directed modifying polypeptide; and a reagent for reconstituting and/or diluting the site-directed modifying polypeptide.

[00351] A subject kit comprising a site-directed modifying polypeptide, or a polynucleotide encoding the same, can further include one or more additional reagents, where such additional reagents can be selected from: a buffer for introducing the site-directed modifying polypeptide into a cell; a wash buffer; a control reagent; a control expression vector or RNA polynucleotide; a reagent for *in vitro* production of the site-directed modifying polypeptide from DNA, and the like. In some cases, the site-directed modifying polypeptide included in a subject kit is a chimeric site-directed modifying polypeptide, as described above.

[00352] In some embodiments, a subject kit comprises a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide. In some embodiments, the DNA-targeting RNA further comprises a third segment (as described above). In some embodiments, a subject kit comprises: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a site-directed

modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some embodiments, the activity portion of the site-directed modifying polypeptide does not exhibit enzymatic activity (comprises an inactivated nuclease, e.g., via mutation). In some cases, the kit comprises a DNA-targeting RNA and a site-directed modifying polypeptide. In other cases, the kit comprises: (i) a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; and (ii) a nucleic acid comprising a nucleotide sequence encoding site-directed modifying polypeptide.

[00353] As another example, a subject kit can include: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA In some cases, the kit comprises: (i) a DNA-targeting RNA; and a site-directed modifying polypeptide. In other cases, the kit comprises: (i) a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; and (ii) a nucleic acid comprising a nucleotide sequence encoding site-directed modifying polypeptide.

[00354] The present disclosure provides a kit comprising: (1) a recombinant expression vector comprising (i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a nucleotide sequence encoding the site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.; and (2) a reagent for reconstitution and/or dilution of the expression vector.

[00355] The present disclosure provides a kit comprising: (1) a recombinant expression vector comprising: (i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a nucleotide sequence encoding the site-directed

modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA; and (2) a reagent for reconstitution and/or dilution of the recombinant expression vector.

[00356] The present disclosure provides a kit comprising: (1) a recombinant expression vector comprising a nucleic acid comprising a nucleotide sequence that encodes a DNA targeting RNA comprising: (i) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) a second segment that interacts with a site-directed modifying polypeptide; and (2) a reagent for reconstitution and/or dilution of the recombinant expression vector. In some embodiments of this kit, the kit comprises: a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In other embodiments of this kit, the kit comprises: a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00357] In some embodiments of any of the above kits, the kit comprises an activator-RNA or a targeter-RNA. In some embodiments of any of the above kits, the kit comprises a single-molecule DNA-targeting RNA. In some embodiments of any of the above kits, the kit comprises two or more double-molecule or single-molecule DNA-targeting RNAs. In some embodiments of any of the above kits, a DNA-targeting RNA (e.g., including two or more DNA-targeting RNAs) can be provided as an array (e.g., an array of RNA molecules, an array of DNA molecules encoding the DNA-targeting RNA(s), etc.). Such kits can be useful, for example, for use in conjunction with the above described genetically modified host cells that comprise a subject site-directed modifying polypeptide. In some embodiments of any of the above kits, the kit further comprises a donor polynucleotide to effect the desired genetic modification. Components of a subject kit can be in separate containers; or can be combined in a single container.

[00358] Any of the above-described kits can further include one or more additional reagents, where such additional reagents can be selected from: a dilution buffer; a reconstitution solution; a wash

buffer; a control reagent; a control expression vector or RNA polynucleotide; a reagent for *in vitro* production of the site-directed modifying polypeptide from DNA, and the like.

[00359] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, flash drive, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

NON-HUMAN GENETICALLY MODIFIED ORGANISMS

[00360] In some embodiments, a genetically modified host cell has been genetically modified with an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.). If such a cell is a eukaryotic single-cell organism, then the modified cell can be considered a genetically modified organism. In some embodiments, subject non-human genetically modified organism is a Cas9 transgenic multicellular organism.

[00361] In some embodiments, a subject genetically modified non-human host cell (e.g., a cell that has been genetically modified with an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) can generate a subject genetically modified non-human organism (e.g., a mouse, a fish, a frog, a fly, a worm, etc.). For example, if the genetically modified host cell is a pluripotent stem cell (i.e., PSC) or a germ cell (e.g., sperm, oocyte, etc.), an entire genetically modified organism can be derived from the genetically modified host cell. In some embodiments, the genetically modified host cell is a pluripotent stem cell (e.g., ESC, iPSC, pluripotent plant stem cell, etc.) or a germ cell (e.g., sperm cell, oocyte, etc.), either in vivo or in vitro, that can give rise to a genetically modified organism. In some embodiments the genetically modified host cell is a vertebrate PSC (e.g., ESC, iPSC, etc.) and is used to generate a genetically modified organism (e.g. by injecting a PSC into a blastocyst to produce a chimeric/mosaic animal, which could then be mated to generate non-chimeric/non-mosaic

genetically modified organisms; grafting in the case of plants; etc.). Any convenient method/protocol for producing a genetically modified organism, including the methods described herein, is suitable for producing a genetically modified host cell comprising an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.). Methods of producing genetically modified organisms are known in the art. For example, see Cho et al., Curr Protoc Cell Biol. 2009 Mar;Chapter 19:Unit 19.11: Generation of transgenic mice; Gama et al., Brain Struct Funct. 2010 Mar;214(2-3):91-109. Epub 2009 Nov 25: Animal transgenesis: an overview; Husaini et al., GM Crops. 2011 Jun-Dec;2(3):150-62. Epub 2011 Jun 1: Approaches for gene targeting and targeted gene expression in plants.

[00362] In some embodiments, a genetically modified organism comprises a target cell for methods of the invention, and thus can be considered a source for target cells. For example, if a genetically modified cell comprising an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) is used to generate a genetically modified organism, then the cells of the genetically modified organism comprise the exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.). In some such embodiments, the DNA of a cell or cells of the genetically modified organism can be targeted for modification by introducing into the cell or cells a DNA-targeting RNA (or a DNA encoding a DNA-targeting RNA) and optionally a donor nucleic acid. For example, the introduction of a DNA-targeting RNA (or a DNA encoding a DNA-targeting RNA) into a subset of cells (e.g., brain cells, intestinal cells, kidney cells, lung cells, blood cells, etc.) of the genetically modified organism can target the DNA of such cells for modification, the genomic location of which will depend on the DNA-targeting sequence of the introduced DNA-targeting RNA.

[00363] In some embodiments, a genetically modified organism is a source of target cells for methods of the invention. For example, a genetically modified organism comprising cells that are genetically modified with an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) can provide a source of genetically modified cells, for example PSCs (e.g., ESCs, iPSCs, sperm, oocytes, etc.), neurons, progenitor cells, cardiomyocytes, etc.

[00364] In some embodiments, a genetically modified cell is a PSC comprising an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a

naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.). As such, the PSC can be a target cell such that the DNA of the PSC can be targeted for modification by introducing into the PSC a DNA-targeting RNA (or a DNA encoding a DNA-targeting RNA) and optionally a donor nucleic acid, and the genomic location of the modification will depend on the DNA-targeting sequence of the introduced DNA-targeting RNA. Thus, in some embodiments, the methods described herein can be used to modify the DNA (e.g., delete and/or replace any desired genomic location) of PSCs derived from a subject genetically modified organism. Such modified PSCs can then be used to generate organisms having both (i) an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) and (ii) a DNA modification that was introduced into the PSC.

[00365] An exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) can be under the control of (i.e., operably linked to) an unknown promoter (e.g., when the nucleic acid randomly integrates into a host cell genome) or can be under the control of (i.e., operably linked to) a known promoter. Suitable known promoters can be any known promoter and include constitutively active promoters (e.g., CMV promoter), inducible promoters (e.g., heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc.), spatially restricted and/or temporally restricted promoters (e.g., a tissue specific promoter, a cell type specific promoter, etc.), etc.

[00366] A subject genetically modified organism (e.g. an organism whose cells comprise a nucleotide sequence encoding a site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) can be any organism including for example, a plant; algae; an invertebrate (e.g., a cnidarian, an echinoderm, a worm, a fly, etc.); a vertebrate (e.g., a fish (e.g., zebrafish, puffer fish, gold fish, etc.), an amphibian (e.g., salamander, frog, etc.), a reptile, a bird, a mammal, etc.); an ungulate (e.g., a goat, a pig, a sheep, a cow, etc.); a rodent (e.g., a mouse, a rat, a hamster, a guinea pig); a lagomorpha (e.g., a rabbit); etc.

[00367] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

Transgenic non-human animals

[00368] As described above, in some embodiments, a subject nucleic acid (e.g., a nucleotide sequence encoding a site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) or a subject recombinant expression vector is used as a transgene to generate a transgenic animal that produces a site-directed modifying polypeptide. Thus, the present invention further provides a transgenic non-human animal, which animal comprises a transgene comprising a subject nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc., as described above. In some embodiments, the genome of the transgenic non-human animal comprises a subject nucleotide sequence encoding a site-directed modifying polypeptide. In some embodiments, the transgenic non-human animal is homozygous for the genetic modification. In some embodiments, the transgenic non-human animal is heterozygous for the genetic modification. In some embodiments, the transgenic non-human animal is a vertebrate, for example, a fish (e.g., zebra fish, gold fish, puffer fish, cave fish, etc.), an amphibian (frog, salamander, etc.), a bird (e.g., chicken, turkey, etc.), a reptile (e.g., snake, lizard, etc.), a mammal (e.g., an ungulate, e.g., a pig, a cow, a goat, a sheep, etc.; a lagomorph (e.g., a rabbit); a rodent (e.g., a rat, a mouse); a non-human primate; etc.), etc.

[00369] An exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) can be under the control of (i.e., operably linked to) an unknown promoter (e.g., when the nucleic acid randomly integrates into a host cell genome) or can be under the control of (i.e., operably linked to) a known promoter. Suitable known promoters can be any known promoter and include constitutively active promoters (e.g., CMV promoter), inducible promoters (e.g., heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc.), spatially restricted and/or temporally restricted promoters (e.g., a tissue specific promoter, a cell type specific promoter, etc.), etc.

[00370] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

Transgenic plants

[00371] As described above, in some embodiments, a subject nucleic acid (e.g., a nucleotide sequence encoding a site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) or a subject recombinant expression vector is used as a transgene to generate a transgenic plant that produces a site-directed modifying polypeptide. Thus, the present invention further provides a transgenic plant, which plant comprises a transgene comprising a subject nucleic acid comprising a nucleotide sequence encoding site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc., as described above. In some embodiments, the genome of the transgenic plant comprises a subject nucleic acid. In some embodiments, the transgenic plant is homozygous for the genetic modification. In some embodiments, the transgenic plant is heterozygous for the genetic modification.

[00372] Methods of introducing exogenous nucleic acids into plant cells are well known in the art. Such plant cells are considered "transformed," as defined above. Suitable methods include viral infection (such as double stranded DNA viruses), transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, silicon carbide whiskers technology, *Agrobacterium*-mediated transformation and the like. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (i.e. *in vitro*, *ex vivo*, or *in vivo*).

[00373] Transformation methods based upon the soil bacterium *Agrobacterium tumefaciens* are particularly useful for introducing an exogenous nucleic acid molecule into a vascular plant. The wild type form of *Agrobacterium* contains a Ti (tumor-inducing) plasmid that directs production of tumorigenic crown gall growth on host plants. Transfer of the tumor-inducing T-DNA region of the Ti plasmid to a plant genome requires the Ti plasmid-encoded virulence genes as well as T-DNA borders, which are a set of direct DNA repeats that delineate the region to be transferred. An *Agrobacterium*-based vector is a modified form of a Ti plasmid, in which the tumor inducing functions are replaced by the nucleic acid sequence of interest to be introduced into the plant host.

[00374] *Agrobacterium*-mediated transformation generally employs cointegrate vectors or binary vector systems, in which the components of the Ti plasmid are divided between a helper vector, which resides permanently in the *Agrobacterium* host and carries the virulence genes, and a shuttle vector, which contains the gene of interest bounded by T-DNA sequences. A variety of binary vectors are well known in the art and are commercially available, for example, from Clontech (Palo Alto, Calif.). Methods of coculturing *Agrobacterium* with cultured plant cells or wounded tissue such as leaf tissue, root explants, hypocotyledons, stem pieces or tubers, for example, also

are well known in the art. See., e.g., Glick and Thompson, (eds.), *Methods in Plant Molecular Biology and Biotechnology*, Boca Raton, Fla.: CRC Press (1993).

[00375] Microprojectile-mediated transformation also can be used to produce a subject transgenic plant. This method, first described by Klein et al. (*Nature* 327:70-73 (1987)), relies on microprojectiles such as gold or tungsten that are coated with the desired nucleic acid molecule by precipitation with calcium chloride, spermidine or polyethylene glycol. The microprojectile particles are accelerated at high speed into an angiosperm tissue using a device such as the BIOLISTIC PD-1000 (Biorad; Hercules Calif.).

[00376] A subject nucleic acid may be introduced into a plant in a manner such that the nucleic acid is able to enter a plant cell(s), e.g., via an *in vivo* or *ex vivo* protocol. By "*in vivo*," it is meant in the nucleic acid is administered to a living body of a plant *e.g.* infiltration. By "*ex vivo*" it is meant that cells or explants are modified outside of the plant, and then such cells or organs are regenerated to a plant. A number of vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described, including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642. Alternatively, non-Ti vectors can be used to transfer the DNA into plants and cells by using free DNA delivery techniques. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9:957-9 and 4462) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48 and for Agrobacterium-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750). Exemplary methods for introduction of DNA into chloroplasts are biolistic bombardment, polyethylene glycol transformation of protoplasts, and microinjection (Danieli et al Nat. Biotechnol 16:345-348, 1998; Staub et al Nat. Biotechnol 18: 333-338, 2000; O'Neill et al Plant J. 3:729-738, 1993; Knoblauch et al Nat. Biotechnol 17: 906-909; U.S. Pat. Nos. 5,451,513, 5,545,817, 5,545,818, and 5,576,198; in Intl. Application No. WO 95/16783; and in Boynton et al., Methods in Enzymology 217: 510-536 (1993), Svab et al., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993), and McBride et al., Proc. Nati. Acad. Sci. USA 91: 7301-7305 (1994)). Any vector suitable for the methods of biolistic bombardment, polyethylene glycol transformation of protoplasts and microinjection will be suitable as a targeting vector for chloroplast

transformation. Any double stranded DNA vector may be used as a transformation vector, especially when the method of introduction does not utilize *Agrobacterium*.

[00377] Plants which can be genetically modified include grains, forage crops, fruits, vegetables, oil seed crops, palms, forestry, and vines. Specific examples of plants which can be modified follow: maize, banana, peanut, field peas, sunflower, tomato, canola, tobacco, wheat, barley, oats, potato, soybeans, cotton, carnations, sorghum, lupin and rice.

[00378] Also provided by the subject invention are transformed plant cells, tissues, plants and products that contain the transformed plant cells. A feature of the subject transformed cells, and tissues and products that include the same is the presence of a subject nucleic acid integrated into the genome, and production by plant cells of a site-directed modifying polypeptide, e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc. Recombinant plant cells of the present invention are useful as populations of recombinant cells, or as a tissue, seed, whole plant, stem, fruit, leaf, root, flower, stem, tuber, grain, animal feed, a field of plants, and the like.

[00379] A nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide (e.g., a naturally occurring Cas9; a modified, i.e., mutated or variant, Cas9; a chimeric Cas9; etc.) can be under the control of (i.e., operably linked to) an unknown promoter (e.g., when the nucleic acid randomly integrates into a host cell genome) or can be under the control of (i.e., operably linked to) a known promoter. Suitable known promoters can be any known promoter and include constitutively active promoters, inducible promoters, spatially restricted and/or temporally restricted promoters, etc.

[00380] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3, or to the corresponding portions in any of the amino acid sequences set forth as SEQ ID NOs:1-256 and 795-1346.

[00381] Also provided by the subject invention is reproductive material of a subject transgenic plant, where reproductive material includes seeds, progeny plants and clonal material.

DEFINITIONS – PART II

[00382] The term "naturally-occurring" or "unmodified" as used herein as applied to a nucleic acid, a polypeptide, a cell, or an organism, refers to a nucleic acid, polypeptide, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in an

organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by a human in the laboratory is naturally occurring.

[00383] "Heterologous," as used herein, means a nucleotide or polypeptide sequence that is not found in the native nucleic acid or protein, respectively. For example, in a fusion variant Cas9 site-directed polypeptide, a variant Cas9 site-directed polypeptide may be fused to a heterologous polypeptide (i.e. a polypeptide other than Cas9). The heterologous polypeptide may exhibit an activity (e.g., enzymatic activity) that will also be exhibited by the fusion variant Cas9 site-directed polypeptide. A heterologous nucleic acid sequence may be linked to a variant Cas9 site-directed polypeptide (e.g., by genetic engineering) to generate a nucleotide sequence encoding a fusion variant Cas9 site-directed polypeptide.

[00384] The term "chimeric polypeptide" refers to a polypeptide which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of amino sequence through human intervention. Thus, a chimeric polypeptide is also the result of human intervention. Thus, a polypeptide that comprises a chimeric amino acid sequence is a chimeric polypeptide.

[00385] By "site-directed polypeptide" or "RNA-binding site-directed polypeptide" or "RNA-binding site-directed polypeptide" it is meant a polypeptide that binds RNA and is targeted to a specific DNA sequence. A site-directed polypeptide as described herein is targeted to a specific DNA sequence by the RNA molecule to which it is bound. The RNA molecule comprises a sequence that is complementary to a target sequence within the target DNA, thus targeting the bound polypeptide to a specific location within the target DNA (the target sequence).

[00386] In some embodiments, a subject nucleic acid (e.g., a DNA-targeting RNA, a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; a nucleic acid encoding a site-directed polypeptide; etc.) comprises a modification or sequence that provides for an additional desirable feature (e.g., modified or regulated stability; subcellular targeting; tracking, e.g., a fluorescent label; a binding site for a protein or protein complex; etc.). Non-limiting examples include: a 5' cap (e.g., a 7-methylguanylate cap (m^7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and/or protein complexes); a modification or sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional

repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

[00387] In some embodiments, a DNA-targeting RNA comprises an additional segment at either the 5' or 3' end that provides for any of the features described above. For example, a suitable third segment can comprise a 5' cap (e.g., a 7-methylguanylate cap (m^7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and protein complexes); a sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

[00388] A subject DNA-targeting RNA and a subject site-directed polypeptide form a complex (i.e., bind via non-covalent interactions). The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-directed polypeptide of the complex provides the site-specific activity. In other words, the site-directed polypeptide is guided to a target DNA sequence (e.g. a target sequence in a chromosomal nucleic acid; a target sequence in an extrachromosomal nucleic acid, e.g. an episomal nucleic acid, a minicircle, etc.; a target sequence in a mitochondrial nucleic acid; a target sequence in a chloroplast nucleic acid; a target sequence in a plasmid; etc.) by virtue of its association with the protein-binding segment of the DNA-targeting RNA.

[00389] In some embodiments, a subject DNA-targeting RNA comprises two separate RNA molecules (RNA polynucleotides) and is referred to herein as a "double-molecule DNA-targeting RNA" or a "two-molecule DNA-targeting RNA." In other embodiments, a subject DNA-targeting RNA is a single RNA molecule (single RNA polynucleotide) and is referred to herein as a "single-molecule DNA-targeting RNA.". If not otherwise specified, the term "DNA-targeting RNA" is inclusive, referring to both single-molecule DNA-targeting RNAs and double-molecule DNA-targeting RNAs.

[00390] A subject two-molecule DNA-targeting RNA comprises two separate RNA molecules (a "targeter-RNA" and an "activator-RNA"). Each of the two RNA molecules of a subject two-molecule DNA-targeting RNA comprises a stretch of nucleotides that are complementary to one another such that the complementary nucleotides of the two RNA molecules hybridize to form the double stranded RNA duplex of the protein-binding segment.

[00391] A subject single-molecule DNA-targeting RNA comprises two stretches of nucleotides (a targeter-RNA and an activator-RNA) that are complementary to one another, are covalently linked by intervening nucleotides ("linkers" or "linker nucleotides"), and hybridize to form the double stranded RNA duplex (dsRNA duplex) of the protein-binding segment, thus resulting in a stem-loop structure. The targeter-RNA and the activator-RNA can be covalently linked via the 3' end of the targeter-RNA and the 5' end of the activator-RNA. Alternatively, targeter-RNA and the activator-RNA can be covalently linked via the 5' end of the targeter-RNA and the 3' end of the activator-RNA.

[00392] An exemplary two-molecule DNA-targeting RNA comprises a crRNA-like ("CRISPR RNA" or "targeter-RNA" or "crRNA" or "crRNA repeat") molecule and a corresponding tracrRNA-like ("trans-acting CRISPR RNA" or "activator-RNA" or "tracrRNA") molecule. A crRNA-like molecule (targeter-RNA) comprises both the DNA-targeting segment (single stranded) of the DNA-targeting RNA and a stretch ("duplex-forming segment") of nucleotides that forms one half of the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. A corresponding tracrRNA-like molecule (activator-RNA) comprises a stretch of nucleotides (duplex-forming segment) that forms the other half of the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. In other words, a stretch of nucleotides of a crRNA-like molecule are complementary to and hybridize with a stretch of nucleotides of a tracrRNA-like molecule to form the dsRNA duplex of the protein-binding domain of the DNA-targeting RNA. As such, each crRNA-like molecule can be said to have a corresponding tracrRNA-like molecule. The crRNA-like molecule additionally provides the single stranded DNA-targeting segment. Thus, a crRNA-like and a tracrRNA-like molecule (as a corresponding pair) hybridize to form a DNA-targeting RNA. The exact sequence of a given crRNA or tracrRNA molecule is characteristic of the species in which the RNA molecules are found.

[00393] The term "activator-RNA" is used herein to mean a tracrRNA-like molecule of a double-molecule DNA-targeting RNA. The term "targeter-RNA" is used herein to mean a crRNA-like molecule of a double-molecule DNA-targeting RNA. The term "duplex-forming segment" is used herein to mean the stretch of nucleotides of an activator-RNA or a targeter-RNA that contributes to the formation of the dsRNA duplex by hybridizing to a stretch of nucleotides of a corresponding activator-RNA or targeter-RNA molecule. In other words, an activator-RNA comprises a duplex-forming segment that is complementary to the duplex-forming segment of the corresponding targeter-RNA. As such, an activator-RNA comprises a duplex-forming segment while a targeter-RNA comprises both a duplex-forming segment and the DNA-targeting segment of the DNA-targeting RNA. Therefore, a subject double-molecule DNA-targeting RNA can be comprised of any corresponding activator-RNA and targeter-RNA pair.

[00394] A two-molecule DNA-targeting RNA can be designed to allow for controlled (i.e., conditional) binding of a targeter-RNA with an activator-RNA. Because a two-molecule DNA-targeting RNA is not functional unless both the activator-RNA and the targeter-RNA are bound in a functional complex with dCas9, a two-molecule DNA-targeting RNA can be inducible (e.g., drug inducible) by rendering the binding between the activator-RNA and the targeter-RNA to be inducible. As one non-limiting example, RNA aptamers can be used to regulate (i.e., control) the binding of the activator-RNA with the targeter-RNA. Accordingly, the activator-RNA and/or the targeter-RNA can comprise an RNA aptamer sequence.

[00395] RNA aptamers are known in the art and are generally a synthetic version of a riboswitch. The terms "RNA aptamer" and "riboswitch" are used interchangeably herein to encompass both synthetic and natural nucleic acid sequences that provide for inducible regulation of the structure (and therefore the availability of specific sequences) of the RNA molecule of which they are part. RNA aptamers usually comprise a sequence that folds into a particular structure (e.g., a hairpin), which specifically binds a particular drug (e.g., a small molecule). Binding of the drug causes a structural change in the folding of the RNA, which changes a feature of the nucleic acid of which the aptamer is a part. As non-limiting examples: (i) an activator-RNA with an aptamer may not be able to bind to the cognate targeter-RNA unless the aptamer is bound by the appropriate drug; (ii) a targeter-RNA with an aptamer may not be able to bind to the cognate activator-RNA unless the aptamer is bound by the appropriate drug; and (iii) a targeter-RNA and an activator-RNA, each comprising a different aptamer that binds a different drug, may not be able to bind to each other unless both drugs are present. As illustrated by these examples, a two-molecule DNA-targeting RNA can be designed to be inducible.

[00396] Examples of aptamers and riboswitches can be found, for example, in: Nakamura et al., Genes Cells. 2012 May;17(5):344-64; Vavalle et al., Future Cardiol. 2012 May;8(3):371-82; Citartan et al., Biosens Bioelectron. 2012 Apr 15;34(1):1-11; and Liberman et al., Wiley Interdiscip Rev **RNA**. 2012 May-Jun;3(3):369-84; all of which are herein incorporated by reference in their entirety.

[00397] Non-limiting examples of nucleotide sequences that can be included in a two-molecule DNA-targeting RNA include targeter RNAs (e.g., SEQ ID NOs:566-567) that can pair with the duplex forming seqment of any one of the activator RNAs set forth in SEQ ID NOs:671-678.

[00398] An exemplary single-molecule DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to form a dsRNA duplex. In some embodiments, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 60% identical to one of the activator-RNA

(tracrRNA) sequences set forth in SEQ ID NOs:431-562 over a stretch of at least 8 contiguous nucleotides. For example, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical or 100 % identical to one of the tracrRNA sequences set forth in SEQ ID NOs:431-562over a stretch of at least 8 contiguous nucleotides.

[00399] In some embodiments, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 60% identical to one of the targeter-RNA (crRNA) sequences set forth in SEQ ID NOs:563-679 over a stretch of at least 8 contiguous nucleotides. For example, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical or 100 % identical to one of the crRNA sequences set forth in SEQ ID NOs:563-679 over a stretch of at least 8 contiguous nucleotides.

[00400] As above, a "host cell," as used herein, denotes an *in vivo* or *in vitro* eukaryotic cell, a prokaryotic cell (e.g., bacterial or archaeal cell), or a cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells can be, or have been, used as recipients for a nucleic acid, and include the progeny of the original cell which has been transformed by the nucleic acid. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation. A "recombinant host cell" (also referred to as a "genetically modified host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector. For example, a subject bacterial host cell is a genetically modified bacterial host cell by virtue of introduction into a suitable bacterial host cell of an exogenous nucleic acid (e.g., a plasmid or recombinant expression vector) and a subject eukaryotic host cell is a genetically modified eukaryotic host cell (e.g., a mammalian germ cell), by virtue of introduction into a suitable eukaryotic host cell of an exogenous nucleic acid.

[00401] Definitions provided in “Definitions – Part I” are also applicable to the instant section; see “Definitions – Part I” for additional clarification of terms.

[00402] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[00403] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[00404] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[00405] It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an enzymatically inactive Cas9 polypeptide" includes a plurality of such polypeptides and reference to "the target nucleic acid" includes reference to one or more target nucleic acids and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[00406] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the invention are specifically embraced by the present invention and are disclosed herein just as if each and every combination

was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present invention and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[00407] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

DETAILED DESCRIPTION – PART II

[00408] The present disclosure provides methods of modulating transcription of a target nucleic acid in a host cell. The methods generally involve contacting the target nucleic acid with an enzymatically inactive Cas9 polypeptide and a single-guide RNA. The methods are useful in a variety of applications, which are also provided.

[00409] A transcriptional modulation method of the present disclosure overcomes some of the drawbacks of methods involving RNAi. A transcriptional modulation method of the present disclosure finds use in a wide variety of applications, including research applications, drug discovery (e.g., high throughput screening), target validation, industrial applications (e.g., crop engineering; microbial engineering, etc.), diagnostic applications, therapeutic applications, and imaging techniques.

METHODS OF MODULATING TRANSCRIPTION

[00410] The present disclosure provides a method of selectively modulating transcription of a target DNA in a host cell. The method generally involves: a) introducing into the host cell: i) a DNA-targeting RNA, or a nucleic acid comprising a nucleotide sequence encoding the DNA-targeting RNA; and ii) a variant Cas9 site-directed polypeptide ("variant Cas9 polypeptide"), or a nucleic acid comprising a nucleotide sequence encoding the variant Cas9 polypeptide, where the variant Cas9 polypeptide exhibits reduced endodeoxyribonuclease activity.

[00411] The DNA-targeting RNA (also referred to herein as "crRNA"; or "guide RNA"; or "gRNA") comprises: i) a first segment comprising a nucleotide sequence that is complementary to a target sequence in a target DNA; ii) a second segment that interacts with a site-directed polypeptide; and iii) a transcriptional terminator. The first segment, comprising a nucleotide sequence that is complementary to a target sequence in a target DNA, is referred to herein as a "targeting segment". The second segment, which interacts with a site-directed polypeptide, is

also referred to herein as a "protein-binding sequence" or "dCas9-binding hairpin," or "dCas9 handle." By "segment" it is meant a segment/section/region of a molecule, e.g., a contiguous stretch of nucleotides in an RNA. The definition of "segment," unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, and may include regions of RNA molecules that are of any total length and may or may not include regions with complementarity to other molecules. A DNA-targeting RNA according to the present disclosure can be a single RNA molecule (single RNA polynucleotide), which can be referred to herein as a "single-molecule DNA-targeting RNA," a "single-guide RNA," or an "sgRNA." A DNA-targeting RNA according to the present disclosure can comprise two RNA molecules. The term "DNA-targeting RNA" or "gRNA" is inclusive, referring both to two-molecule DNA-targeting RNAs and to single-molecule DNA-targeting RNAs (i.e., sgRNAs).

[00412] The variant Cas9 site-directed polypeptide comprises: i) an RNA-binding portion that interacts with the DNA-targeting RNA; and ii) an activity portion that exhibits reduced endodeoxyribonuclease activity.

[00413] The DNA-targeting RNA and the variant Cas9 polypeptide form a complex in the host cell; the complex selectively modulates transcription of a target DNA in the host cell.

[00414] In some cases, a transcription modulation method of the present disclosure provides for selective modulation (e.g., reduction or increase) of a target nucleic acid in a host cell. For example, "selective" reduction of transcription of a target nucleic acid reduces transcription of the target nucleic acid by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or greater than 90%, compared to the level of transcription of the target nucleic acid in the absence of a DNA-targeting RNA/variant Cas9 polypeptide complex. Selective reduction of transcription of a target nucleic acid reduces transcription of the target nucleic acid, but does not substantially reduce transcription of a non-target nucleic acid, e.g., transcription of a non-target nucleic acid is reduced, if at all, by less than 10% compared to the level of transcription of the non-target nucleic acid in the absence of the DNA-targeting RNA/variant Cas9 polypeptide complex.

Increased transcription

[00415] "Selective" increased transcription of a target DNA can increase transcription of the target DNA by at least about 1.1 fold (e.g., at least about 1.2 fold, at least about 1.3 fold, at least about 1.4 fold, at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about 1.9 fold, at least about 2 fold, at least about 2.5 fold, at least about 3 fold, at least about 3.5 fold, at least about 4 fold, at least about 4.5 fold, at least about 5 fold, at least about 6 fold, at least about 7 fold, at least about 8 fold, at least about 9 fold, at least about 10

fold, at least about 12 fold, at least about 15 fold, or at least about 20-fold) compared to the level of transcription of the target DNA in the absence of a DNA-targeting RNA/variant Cas9 polypeptide complex. Selective increase of transcription of a target DNA increases transcription of the target DNA, but does not substantially increase transcription of a non-target DNA, e.g., transcription of a non-target DNA is increased, if at all, by less than about 5-fold (e.g., less than about 4-fold, less than about 3-fold, less than about 2-fold, less than about 1.8-fold, less than about 1.6-fold, less than about 1.4-fold, less than about 1.2-fold, or less than about 1.1-fold) compared to the level of transcription of the non-targeted DNA in the absence of the DNA-targeting RNA/variant Cas9 polypeptide complex.

[00416] As a non-limiting example, increased can be achieved by fusing dCas9 to a heterologous sequence. Suitable fusion partners include, but are not limited to, a polypeptide that provides an activity that indirectly increases transcription by acting directly on the target DNA or on a polypeptide (e.g., a histone or other DNA-binding protein) associated with the target DNA. Suitable fusion partners include, but are not limited to, a polypeptide that provides for methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, or demyristoylation activity.

[00417] Additional suitable fusion partners include, but are not limited to, a polypeptide that directly provides for increased transcription of the target nucleic acid (e.g., a transcription activator or a fragment thereof, a protein or fragment thereof that recruits a transcription activator, a small molecule/drug-responsive transcription regulator, etc.).

[00418] A non-limiting example of a subject method using a dCas9 fusion protein to increase transcription in a prokaryote includes a modification of the bacterial one-hybrid (B1H) or two-hybrid (B2H) system. In the B1H system, a DNA binding domain (BD) is fused to a bacterial transcription activation domain (AD, e.g., the alpha subunit of the *Escherichia coli* RNA polymerase (RNAPα)). Thus, a subject dCas9 can be fused to a heterologous sequence comprising an AD. When the subject dCas9 fusion protein arrives at the upstream region of a promoter (targeted there by the DNA-targeting RNA) the AD (e.g., RNAPα) of the dCas9 fusion protein recruits the RNAP holoenzyme, leading to transcription activation. In the B2H system, the BD is not directly fused to the AD; instead, their interaction is mediated by a protein-protein interaction (e.g., GAL11P - GAL4 interaction). To modify such a system for use in the subject methods, dCas9 can be fused to a first protein sequence that provides for protein-protein interaction (e.g., the yeast GAL11P and/or GAL4 protein) and RNAα can be fused to a second protein sequence that completes the protein-protein interaction (e.g., GAL4 if GAL11P is fused

to dCas9, GAL11P if GAL4 is fused to dCas9, etc.). The binding affinity between GAL11P and GAL4 increases the efficiency of binding and transcription firing rate.

[00419] A non-limiting example of a subject method using a dCas9 fusion protein to increase transcription in a eukaryotes includes fusion of dCas9 to an activation domain (AD) (e.g., GAL4, herpesvirus activation protein VP16 or VP64, human nuclear factor NF-κB p65 subunit, etc.). To render the system inducible, expression of the dCas9 fusion protein can be controlled by an inducible promoter (e.g., Tet-ON, Tet-OFF, etc.). The DNA-targeting RNA can be design to target known transcription response elements (e.g., promoters, enhancers, etc.), known upstream activating sequences (UAS), sequences of unknown or known function that are suspected of being able to control expression of the target DNA, etc.

Additional fusion partners

[00420] Non-limiting examples of fusion partners to accomplish increased or decreased transcription are listed in **Figure 54** and include transcription activator and transcription repressor domains (e.g., the Krüppel associated box (KRAB or SKD); the Mad mSIN3 interaction domain (SID); the ERF repressor domain (ERD), etc). In some such cases, the dCas9 fusion protein is targeted by the DNA-targeting RNA to a specific location (i.e., sequence) in the target DNA and exerts locus-specific regulation such as blocking RNA polymerase binding to a promoter (which selectively inhibits transcription activator function), and/or modifying the local chromatin status (e.g., when a fusion sequence is used that modifies the target DNA or modifies a polypeptide associated with the target DNA). In some cases, the changes are transient (e.g., transcription repression or activation). In some cases, the changes are inheritable (e.g., when epigenetic modifications are made to the target DNA or to proteins associated with the target DNA, e.g., nucleosomal histones).

[00421] In some embodiments, the heterologous sequence can be fused to the C-terminus of the dCas9 polypeptide. In some embodiments, the heterologous sequence can be fused to the N-terminus of the dCas9 polypeptide. In some embodiments, the heterologous sequence can be fused to an internal portion (i.e., a portion other than the N- or C- terminus) of the dCas9 polypeptide.

[00422] The biological effects of a method using a subject dCas9 fusion protein can be detected by any convenient method (e.g., gene expression assays; chromatin-based assays, e.g., Chromatin immunoPrecipitation (ChiP), Chromatin in vivo Assay (CiA), etc.; and the like).

[00423] In some cases, a subject method involves use of two or more different DNA-targeting RNAs. For example, two different DNA-targeting RNAs can be used in a single host cell, where the two different DNA-targeting RNAs target two different target sequences in the same target nucleic acid.

[00424] Thus, for example, a subject transcriptional modulation method can further comprise introducing into the host cell a second DNA-targeting RNA, or a nucleic acid comprising a nucleotide sequence encoding the second DNA-targeting RNA, where the second DNA-targeting RNA comprises: i) a first segment comprising a nucleotide sequence that is complementary to a second target sequence in the target DNA; ii) a second segment that interacts with the site-directed polypeptide; and iii) a transcriptional terminator. In some cases, use of two different DNA-targeting RNAs targeting two different targeting sequences in the same target nucleic acid provides for increased modulation (e.g., reduction or increase) in transcription of the target nucleic acid.

[00425] As another example, two different DNA-targeting RNAs can be used in a single host cell, where the two different DNA-targeting RNAs target two different target nucleic acids. Thus, for example, a subject transcriptional modulation method can further comprise introducing into the host cell a second DNA-targeting RNA, or a nucleic acid comprising a nucleotide sequence encoding the second DNA-targeting RNA, where the second DNA-targeting RNA comprises: i) a first segment comprising a nucleotide sequence that is complementary to a target sequence in at least a second target DNA; ii) a second segment that interacts with the site-directed polypeptide; and iii) a transcriptional terminator.

[00426] In some embodiments, a subject nucleic acid (e.g., a DNA-targeting RNA, e.g., a single-molecule DNA-targeting RNA, an activator-RNA, a targeter-RNA, etc.; a donor polynucleotide; a nucleic acid encoding a site-directed modifying polypeptide; etc.) comprises a modification or sequence that provides for an additional desirable feature (e.g., modified or regulated stability; subcellular targeting; tracking, e.g., a fluorescent label; a binding site for a protein or protein complex; etc.). Non-limiting examples include: a 5' cap (e.g., a 7-methylguanylate cap (m^7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence or an aptamer sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and/or protein complexes); a terminator sequence; a sequence that forms a dsRNA duplex (i.e., a hairpin)); a modification or sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

DNA-targeting segment

[00427] The DNA-targeting segment (or "DNA-targeting sequence") of a DNA-targeting RNA ("crRNA") comprises a nucleotide sequence that is complementary to a specific sequence within a target DNA (the complementary strand of the target DNA).

[00428] In other words, the DNA-targeting segment of a subject DNA-targeting RNA interacts with a target DNA in a sequence-specific manner via hybridization (i.e., base pairing). As such, the nucleotide sequence of the DNA-targeting segment may vary and determines the location within the target DNA that the DNA-targeting RNA and the target DNA will interact. The DNA-targeting segment of a subject DNA-targeting RNA can be modified (e.g., by genetic engineering) to hybridize to any desired sequence within a target DNA.

[00429] The DNA-targeting segment can have a length of from about 12 nucleotides to about 100 nucleotides. For example, the DNA-targeting segment can have a length of from about 12 nucleotides (nt) to about 80 nt, from about 12 nt to about 50nt, from about 12 nt to about 40 nt, from about 12 nt to about 30 nt, from about 12 nt to about 25 nt, from about 12 nt to about 20 nt, or from about 12 nt to about 19 nt. For example, the DNA-targeting segment can have a length of from about 19 nt to about 20 nt, from about 19 nt to about 25 nt, from about 19 nt to about 30 nt, from about 19 nt to about 35 nt, from about 19 nt to about 40 nt, from about 19 nt to about 45 nt, from about 19 nt to about 50 nt, from about 19 nt to about 60 nt, from about 19 nt to about 70 nt, from about 19 nt to about 80 nt, from about 19 nt to about 90 nt, from about 19 nt to about 100 nt, from about 20 nt to about 25 nt, from about 20 nt to about 30 nt, from about 20 nt to about 35 nt, from about 20 nt to about 40 nt, from about 20 nt to about 45 nt, from about 20 nt to about 50 nt, from about 20 nt to about 60 nt, from about 20 nt to about 70 nt, from about 20 nt to about 80 nt, from about 20 nt to about 90 nt, or from about 20 nt to about 100 nt.

[00430] The nucleotide sequence (the DNA-targeting sequence) of the DNA-targeting segment that is complementary to a nucleotide sequence (target sequence) of the target DNA can have a length at least about 12 nt. For example, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA can have a length at least about 12 nt, at least about 15 nt, at least about 18 nt, at least about 19 nt, at least about 20 nt, at least about 25 nt, at least about 30 nt, at least about 35 nt or at least about 40 nt. For example, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA can have a length of from about 12 nucleotides (nt) to about 80 nt, from about 12 nt to about 50nt, from about 12 nt to about 45 nt, from about 12 nt to about 40 nt, from about 12 nt to about 35 nt, from about 12 nt to about 30 nt, from about 12 nt to about 25 nt, from about 12 nt to about 20 nt, from about 12 nt to about 19 nt, from about 19 nt to about 20 nt, from about 19 nt to about 25 nt, from about 19 nt to about 30 nt, from about 19 nt to about 35 nt, from about 19

nt to about 40 nt, from about 19 nt to about 45 nt, from about 19 nt to about 50 nt, from about 19 nt to about 60 nt, from about 20 nt to about 25 nt, from about 20 nt to about 30 nt, from about 20 nt to about 35 nt, from about 20 nt to about 40 nt, from about 20 nt to about 45 nt, from about 20 nt to about 50 nt, or from about 20 nt to about 60 nt. The nucleotide sequence (the DNA-targeting sequence) of the DNA-targeting segment that is complementary to a nucleotide sequence (target sequence) of the target DNA can have a length at least about 12 nt.

[00431] In some cases, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA is 20 nucleotides in length. In some cases, the DNA-targeting sequence of the DNA-targeting segment that is complementary to a target sequence of the target DNA is 19 nucleotides in length.

[00432] The percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA can be at least 60% (e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%). In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is 100% over the seven contiguous 5'-most nucleotides of the target sequence of the complementary strand of the target DNA. In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is at least 60% over about 20 contiguous nucleotides. In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is 100% over the fourteen contiguous 5'-most nucleotides of the target sequence of the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 14 nucleotides in length. In some cases, the percent complementarity between the DNA-targeting sequence of the DNA-targeting segment and the target sequence of the target DNA is 100% over the seven contiguous 5'-most nucleotides of the target sequence of the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 7 nucleotides in length.

Protein-binding segment

[00433] The protein-binding segment (i.e., "protein-binding sequence") of a DNA-targeting RNA interacts with a variant site-directed polypeptide. When the variant Cas9 site-directed polypeptide, together with the DNA-targeting RNA, binds to a target DNA, transcription of the target DNA is reduced.

[00434] The protein-binding segment of a DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (dsRNA duplex).

[00435] The protein-binding segment of a DNA-targeting RNA of the present disclosure comprises two stretches of nucleotides (a targeter-RNA and an activator-RNA) that are complementary to one another, are covalently linked by intervening nucleotides (e.g., in the case of a single-molecule DNA-targeting RNA)("linkers" or "linker nucleotides"), and hybridize to form the double stranded RNA duplex (dsRNA duplex, or "dCas9-binding hairpin") of the protein-binding segment, thus resulting in a stem-loop structure. This stem-loop structure is shown schematically in Figrue 39A. The targeter-RNA and the activator-RNA can be covalently linked via the 3' end of the targeter-RNA and the 5' end of the activator-RNA. Alternatively, targeter-RNA and the activator-RNA can be covalently linked via the 5' end of the targeter-RNA and the 3' end of the activator-RNA.

[00436] The protein-binding segment can have a length of from about 10 nucleotides to about 100 nucleotides, e.g., from about 10 nucleotides (nt) to about 20 nt, from about 20 nt to about 30 nt, from about 30 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. For example, the protein-binding segment can have a length of from about 15 nucleotides (nt) to about 80 nt, from about 15 nt to about 50 nt, from about 15 nt to about 40 nt, from about 15 nt to about 30 nt or from about 15 nt to about 25 nt.

[00437] The dsRNA duplex of the protein-binding segment can have a length from about 6 base pairs (bp) to about 50bp. For example, the dsRNA duplex of the protein-binding segment can have a length from about 6 bp to about 40 bp, from about 6 bp to about 30bp, from about 6 bp to about 25 bp, from about 6 bp to about 20 bp, from about 6 bp to about 15 bp, from about 8 bp to about 40 bp, from about 8 bp to about 30bp, from about 8 bp to about 25 bp, from about 8 bp to about 20 bp or from about 8 bp to about 15 bp. For example, the dsRNA duplex of the protein-binding segment can have a length from about from about 8 bp to about 10 bp, from about 10 bp to about 15 bp, from about 15 bp to about 18 bp, from about 18 bp to about 20 bp, from about 20 bp to about 25 bp, from about 25 bp to about 30 bp, from about 30 bp to about 35 bp, from about 35 bp to about 40 bp, or from about 40 bp to about 50 bp. In some embodiments, the dsRNA duplex of the protein-binding segment has a length of 36 base pairs. The percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment can be at least about 60%. For example, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment can be at least about 65%, at least about 70%, at least about 75%, at least about 80%, at

least about 85%, at least about 90%, at least about 95%, at least about 98%, or at least about 99%. In some cases, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment is 100%.

[00438] The linker can have a length of from about 3 nucleotides to about 100 nucleotides. For example, the linker can have a length of from about 3 nucleotides (nt) to about 90 nt, from about 3 nucleotides (nt) to about 80 nt, from about 3 nucleotides (nt) to about 70 nt, from about 3 nucleotides (nt) to about 60 nt, from about 3 nucleotides (nt) to about 50 nt, from about 3 nucleotides (nt) to about 40 nt, from about 3 nucleotides (nt) to about 30 nt, from about 3 nucleotides (nt) to about 20 nt or from about 3 nucleotides (nt) to about 10 nt. For example, the linker can have a length of from about 3 nt to about 5 nt, from about 5 nt to about 10 nt, from about 10 nt to about 15 nt, from about 15 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. In some embodiments, the linker of a DNA-targeting RNA is 4 nt.

[00439] Non-limiting examples of nucleotide sequences that can be included in a suitable protein-binding segment (i.e., dCas9 handle) are set forth in SEQ ID NOs:563-682 (For examples, see Figure 8 and Figure 9).

[00440] In some cases, a suitable protein-binding segment comprises a nucleotide sequence that differs by 1, 2, 3, 4, or 5 nucleotides from any one of the above-listed sequences.

Stability control sequence (e.g., transcriptional terminator segment)

[00441] A stability control sequence influences the stability of an RNA (e.g., a DNA-targeting RNA, a targeter-RNA, an activator-RNA, etc.). One example of a suitable stability control sequence is a transcriptional terminator segment (i.e., a transcription termination sequence). A transcriptional terminator segment of a subject DNA-targeting RNA can have a total length of from about 10 nucleotides to about 100 nucleotides, e.g., from about 10 nucleotides (nt) to about 20 nt, from about 20 nt to about 30 nt, from about 30 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. For example, the transcriptional terminator segment can have a length of from about 15 nucleotides (nt) to about 80 nt, from about 15 nt to about 50 nt, from about 15 nt to about 40 nt, from about 15 nt to about 30 nt or from about 15 nt to about 25 nt.

[00442] In some cases, the transcription termination sequence is one that is functional in a eukaryotic cell. In some cases, the transcription termination sequence is one that is functional in a prokaryotic cell.

[00443] Non-limiting examples of nucleotide sequences that can be included in a stability control sequence (e.g., transcriptional termination segment, or in any segment of the DNA-targeting RNA to provide for increased stability) include sequences set forth in SEQ ID NO:683-696 and, for example,

5'-UAAUCCCACAGCCGCCAGUUCCGCUGGCGGCAUUUU-5' (SEQ ID NO:795) (a Rho-independent *trp* termination site).

Additional sequences

[00444] In some embodiments, a DNA-targeting RNA comprises at least one additional segment at either the 5' or 3' end. For example, a suitable additional segment can comprise a 5' cap (e.g., a 7-methylguanylate cap (m^7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and protein complexes); a sequence that forms a dsRNA duplex (i.e., a hairpin)); a sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like) a modification or sequence that provides for increased, decreased, and/or controllable stability; and combinations thereof.

Multiple simultaneous DNA-targeting RNAs

[00445] In some embodiments, multiple DNA-targeting RNAs are used simultaneously in the same cell to simultaneously modulate transcription at different locations on the same target DNA or on different target DNAs. In some embodiments, two or more DNA-targeting RNAs target the same gene or transcript or locus. In some embodiments, two or more DNA-targeting RNAs target different unrelated loci. In some embodiments, two or more DNA-targeting RNAs target different, but related loci.

[00446] Because the DNA-targeting RNAs are small and robust they can be simultaneously present on the same expression vector and can even be under the same transcriptional control if so desired. In some embodiments, two or more (e.g., 3 or more, 4 or more, 5 or more, 10 or more, 15 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, or 50 or more)

DNA-targeting RNAs are simultaneously expressed in a target cell (from the same or different vectors). The expressed DNA-targeting RNAs can be differently recognized by dCas9 proteins from different bacteria, such as *S. pyogenes, S. thermophilus, L. innocua, and N. meningitidis*.

[00447] To express multiple DNA-targeting RNAs, an artificial RNA processing system mediated by the Csy4 endoribonuclease can be used. Multiple DNA-targeting RNAs can be concatenated into a tandem array on a precursor transcript (e.g., expressed from a U6 promoter), and separated by Csy4-specific RNA sequence. Co-expressed Csy4 protein cleaves the precursor transcript into multiple DNA-targeting RNAs. Advantages for using an RNA processing system include: first, there is no need to use multiple promoters; second, since all DNA-targeting RNAs are processed from a precursor transcript, their concentrations are normalized for similar dCas9-binding.

[00448] Csy4 is a small endoribonuclease (RNase) protein derived from bacteria *Pseudomonas aeruginosa*. Csy4 specifically recognizes a minimal 17-bp RNA hairpin, and exhibits rapid (<1 min) and highly efficient (>99.9%) RNA cleavage. Unlike most RNases, the cleaved RNA fragment remains stable and functionally active. The Csy4-based RNA cleavage can be repurposed into an artificial RNA processing system. In this system, the 17-bp RNA hairpins are inserted between multiple RNA fragments that are transcribed as a precursor transcript from a single promoter. Co-expression of Csy4 is effective in generating individual RNA fragments.

Site-directed polypeptide

[00449] As noted above, a subject DNA-targeting RNA and a variant Cas9 site-directed polypeptide form a complex. The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA.

[00450] The variant Cas9 site-directed polypeptide has reduced endodeoxyribonuclease activity. For example, a variant Cas9 site-directed polypeptide suitable for use in a transcription modulation method of the present disclosure exhibits less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 1%, or less than about 0.1%, of the endodeoxyribonuclease activity of a wild-type Cas9 polypeptide, e.g., a wild-type Cas9 polypeptide comprising an amino acid sequence as depicted in Figure 3 (SEQ ID NO:8). In some embodiments, the variant Cas9 site-directed polypeptide has substantially no detectable endodeoxyribonuclease activity. In some embodiments when a site-directed polypeptide has reduced catalytic activity (e.g., when a Cas9 protein has a D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or a A987 mutation, e.g., D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A), the polypeptide can still bind to target DNA in a site-specific manner (because it is still guided to a target DNA sequence by a DNA-targeting RNA) as long as it retains the ability to interact with the DNA-targeting RNA.

[00451] In some cases, a suitable variant Cas9 site-directed polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 3 (SEQ ID NO:8), or to the corresponding portions in any one of the amino acid sequences SEQ ID NOs:1-256 and 795-1346.

[00452] In some cases, the variant Cas9 site-directed polypeptide can cleave the complementary strand of the target DNA but has reduced ability to cleave the non-complementary strand of the target DNA. For example, the variant Cas9 site-directed polypeptide can have a mutation (amino acid substitution) that reduces the function of the RuvC domain (e.g., "domain 1" of Figure 3). As a non-limiting example, in some cases, the variant Cas9 site-directed polypeptide is a D10A (aspartate to alanine) mutation of the amino acid sequence depicted in Figure 3 (or the corresponding mutation of any of the amino acid sequences set forth in SEQ ID NOs:1-256 and 795-1346).

[00453] In some cases, the variant Cas9 site-directed polypeptide can cleave the non-complementary strand of the target DNA but has reduced ability to cleave the complementary strand of the target DNA. For example, the variant Cas9 site-directed polypeptide can have a mutation (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs, "domain 2" of Figure 3). As a non-limiting example, in some cases, the variant Cas9 site-directed polypeptide is a H840A (histidine to alanine at amino acid position 840 of SEQ ID NO:8) or the corresponding mutation of any of the amino acid sequences set forth in SEQ ID NOs:1-256 and 795-1346).

[00454] In some cases, the variant Cas9 site-directed polypeptide has a reduced ability to cleave both the complementary and the non-complementary strands of the target DNA. As a non-limiting example, in some cases, the variant Cas9 site-directed polypeptide harbors both D10A and H840A mutations of the amino acid sequence depicted in Figure 3 (or the corresponding mutations of any of the amino acid sequences set forth in SEQ ID NOs:1-256 and 795-1346).

[00455] Other residues can be mutated to achieve the same effect (i.e. inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the proteins set forth as SEQ ID NOs:1-256 and 795-1346) can be altered (i.e., substituted) (see **Figure 3, Figure 5, Figure 11A**, and **Table 1** for more information regarding the conservation of Cas9 amino acid residues). Also, mutations other than alanine substitutions are suitable.

[00456] In some cases, the variant Cas9 site-directed polypeptide is a fusion polypeptide (a "variant Cas9 fusion polypeptide"), i.e., a fusion polypeptide comprising: i) a variant Cas9 site-directed polypeptide; and b) a covalently linked heterologous polypeptide (also referred to as a "fusion partner").

[00457] The heterologous polypeptide may exhibit an activity (e.g., enzymatic activity) that will also be exhibited by the variant Cas9 fusion polypeptide (e.g., methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.). A heterologous nucleic acid sequence may be linked to another nucleic acid sequence (e.g., by genetic engineering) to generate a chimeric nucleotide sequence encoding a chimeric polypeptide. In some embodiments, a variant Cas9 fusion polypeptide is generated by fusing a variant Cas9 polypeptide with a heterologous sequence that provides for subcellular localization (i.e., the heterologous sequence is a subcellular localization sequence, e.g., a nuclear localization signal (NLS) for targeting to the nucleus; a mitochondrial localization signal for targeting to the mitochondria; a chloroplast localization signal for targeting to a chloroplast; an ER retention signal; and the like). In some embodiments, the heterologous sequence can provide a tag (i.e., the heterologous sequence is a detectable label) for ease of tracking and/or purification (e.g., a fluorescent protein, e.g., green fluorescent protein (GFP), YFP, RFP, CFP, mCherry, tdTomato, and the like; a histidine tag, e.g., a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and the like). In some embodiments, the heterologous sequence can provide for increased or decreased stability (i.e., the heterologous sequence is a stability control peptide, e.g., a degron, which in some cases is controllable (e.g., a temperature sensitive or drug controllable degron sequence, see below). In some embodiments, the heterologous sequence can provide for increased or decreased transcription from the target DNA (i.e., the heterologous sequence is a transcription modulation sequence, e.g., a transcription factor/activator or a fragment thereof, a protein or fragment thereof that recruits a transcription factor/activator, a transcription repressor or a fragment thereof, a protein or fragment thereof that recruits a transcription repressor, a small molecule/drug-responsive transcription regulator, etc.). In some embodiments, the heterologous sequence can provide a binding domain (i.e., the heterologous sequence is a protein binding sequence, e.g., to provide the ability of a chimeric dCas9 polypeptide to bind to another protein of interest, e.g., a DNA or histone modifying protein, a transcription factor or transcription repressor, a recruiting protein, etc.).

[00458] Suitable fusion partners that provide for increased or decreased stability include, but are not limited to degron sequences. Degrons are readily understood by one of ordinary skill in the art to be amino acid sequences that control the stability of the protein of which they are part. For example, the stability of a protein comprising a degron sequence is controlled at least in part by

the degron sequence. In some cases, a suitable degron is constitutive such that the degron exerts its influence on protein stability independent of experimental control (i.e., the degron is not drug inducible, temperature inducible, etc.) In some cases, the degron provides the variant Cas9 polypeptide with controllable stability such that the variant Cas9 polypeptide can be turned "on" (i.e., stable) or "off" (i.e., unstable, degraded) depending on the desired conditions. For example, if the degron is a temperature sensitive degron, the variant Cas9 polypeptide may be functional (i.e., "on", stable) below a threshold temperature (e.g., 42°C, 41°C, 40°C, 39°C, 38°C, 37°C, 36°C, 35°C, 34°C, 33°C, 32°C, 31°C, 30°C, etc.) but non-functional (i.e., "off", degraded) above the threshold temperature. As another example, if the degron is a drug inducible degron, the presence or absence of drug can switch the protein from an "off" (i.e., unstable) state to an "on" (i.e., stable) state or vice versa. An exemplary drug inducible degron is derived from the FKBP12 protein. The stability of the degron is controlled by the presence or absence of a small molecule that binds to the degron.

[00459] Examples of suitable degrons include, but are not limited to those degrons controlled by Shield-1, DHFR, auxins, and/or temperature. Non-limiting examples of suitable degrons are known in the art (e.g., Dohmen et al., Science, 1994. 263(5151): p. 1273-1276: Heat-inducible degron: a method for constructing temperature-sensitive mutants; Schoeber et al., Am J Physiol Renal Physiol. 2009 Jan;296(1):F204-11 : Conditional fast expression and function of multimeric TRPV5 channels using Shield-1 ; Chu et al., Bioorg Med Chem Lett. 2008 Nov 15;18(22):5941-4: Recent progress with FKBP-derived destabilizing domains ; Kanemaki, Pflugers Arch. 2012 Dec 28: Frontiers of protein expression control with conditional degrons; Yang et al., Mol Cell. 2012 Nov 30;48(4):487-8: Titivated for destruction: the methyl degron; Barbour et al., Biosci Rep. 2013 Jan 18;33(1).: Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase; and Greussing et al., J Vis Exp. 2012 Nov 10;(69): Monitoring of ubiquitin-proteasome activity in living cells using a Degron (dgn)-destabilized green fluorescent protein (GFP)-based reporter protein; all of which are hereby incorporated in their entirety by reference).

[00460] Exemplary degron sequences have been well-characterized and tested in both cells and animals. Thus, fusing dCas9 to a degron sequence produces a "tunable" and "inducible" dCas9 polypeptide. Any of the fusion partners described herein can be used in any desirable combination. As one non-limiting example to illustrate this point, a dCas9 fusion protein can comprise a YFP sequence for detection, a degron sequence for stability, and transcription activator sequence to increase transcription of the target DNA. Furthermore, the number of fusion partners that can be used in a dCas9 fusion protein is unlimited. In some cases, a dCas9

fusion protein comprises one or more (e.g. two or more, three or more, four or more, or five or more) heterologous sequences.

[00461] Suitable fusion partners include, but are not limited to, a polypeptide that provides for methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, or demyristoylation activity, any of which can be directed at modifying the DNA directly (e.g., methylation of DNA) or at modifying a DNA-associated polypeptide (e.g., a histone or DNA binding protein). Further suitable fusion partners include, but are not limited to boundary elements (e.g., CTCF), proteins and fragments thereof that provide periphery recruitment (e.g., Lamin A, Lamin B, etc.), and protein docking elements (e.g., FKBP/FRB, Pil1/Aby1, etc.).

[00462] Examples of various additional suitable fusion partners (or fragments thereof) for a subject variant Cas9 site-directed polypeptide include, but are not limited to those listed in **Figure 54**.

[00463] In some embodiments, a subject site-directed modifying polypeptide can be codon-optimized. This type of optimization is known in the art and entails the mutation of foreign-derived DNA to mimic the codon preferences of the intended host organism or cell while encoding the same protein. Thus, the codons are changed, but the encoded protein remains unchanged. For example, if the intended target cell was a human cell, a human codon-optimized dCas9 (or dCas9 variant) would be a suitable site-directed modifying polypeptide. As another non-limiting example, if the intended host cell were a mouse cell, than a mouse codon-optimized Cas9 (or variant, e.g., enzymatically inactive variant) would be a suitable Cas9 site-directed polypeptide. While codon optimization is not required, it is acceptable and may be preferable in certain cases.

Host cells

[00464] A method of the present disclosure to modulate transcription may be employed to induce transcriptional modulation in mitotic or post-mitotic cells *in vivo* and/or *ex vivo* and/or *in vitro*. Because the DNA-targeting RNA provides specificity by hybridizing to target DNA, a mitotic and/or post-mitotic cell can be any of a variety of host cell, where suitable host cells include, but are not limited to, a bacterial cell; an archaeal cell; a single-celled eukaryotic organism; a plant cell; an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like; a fungal cell; an animal cell; a cell from an invertebrate animal (e.g., an insect, a cnidarian, an echinoderm, a nematode, etc.); a eukaryotic parasite (e.g., a malarial parasite, e.g., *Plasmodium falciparum*; a helminth; etc.); a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal); a

mammalian cell, e.g., a rodent cell, a human cell, a non-human primate cell, etc. Suitable host cells include naturally-occurring cells; genetically modified cells (e.g., cells genetically modified in a laboratory, e.g., by the "hand of man"); and cells manipulated *in vitro* in any way. In some cases, a host cell is isolated.

[00465] Any type of cell may be of interest (e.g. a stem cell, e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell; a somatic cell, e.g. a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an in vitro or in vivo embryonic cell of an embryo at any stage, e.g., a 1-cell, 2-cell, 4-cell, 8-cell, etc. stage zebrafish embryo; etc.). Cells may be from established cell lines or they may be primary cells, where "primary cells", "primary cell lines", and "primary cultures" are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, i.e. splittings, of the culture. For example, primary cultures include cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Primary cell lines can be are maintained for fewer than 10 passages *in vitro*. Target cells are in many embodiments unicellular organisms, or are grown in culture.

[00466] If the cells are primary cells, such cells may be harvest from an individual by any convenient method. For example, leukocytes may be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. are most conveniently harvested by biopsy. An appropriate solution may be used for dispersion or suspension of the harvested cells. Such solution will generally be a balanced salt solution, e.g. normal saline, phosphate-buffered saline (PBS), Hank's balanced salt solution, etc., conveniently supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration, e.g., from 5-25 mM. Convenient buffers include HEPES, phosphate buffers, lactate buffers, etc. The cells may be used immediately, or they may be stored, frozen, for long periods of time, being thawed and capable of being reused. In such cases, the cells will usually be frozen in 10% dimethyl sulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures, and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

Introducing nucleic acid into a host cell

[00467] A DNA-targeting RNA, or a nucleic acid comprising a nucleotide sequence encoding same, can be introduced into a host cell by any of a variety of well-known methods. Similarly, where a subject method involves introducing into a host cell a nucleic acid comprising a nucleotide

sequence encoding a variant Cas9 site-directed polypeptide, such a nucleic acid can be introduced into a host cell by any of a variety of well-known methods.

[00468] Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a stem cell or progenitor cell. Suitable methods include, include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, e.g., Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like.

NUCLEIC ACIDS

[00469] The present disclosure provides an isolated nucleic acid comprising a nucleotide sequence encoding a subject DNA-targeting RNA. In some cases, a subject nucleic acid also comprises a nucleotide sequence encoding a variant Cas9 site-directed polypeptide.

[00470] In some embodiments, a subject method involves introducing into a host cell (or a population of host cells) one or more nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a variant Cas9 site-directed polypeptide. In some embodiments a cell comprising a target DNA is *in vitro.* In some embodiments a cell comprising a target DNA is *in vivo.* Suitable nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed polypeptide is a "recombinant expression vector."

[00471] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, a recombinant retroviral construct, etc.

[00472] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683

690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00473] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00474] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00475] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a variant Cas9 site-directed polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a variant Cas9 site-directed polypeptide is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a DNA-targeting RNA and/or a variant Cas9 site-directed polypeptide in both prokaryotic and eukaryotic cells.

[00476] A promoter can be a constitutively active promoter (i.e., a promoter that is constitutively in an active/"ON" state), it may be an inducible promoter (i.e., a promoter whose state, active/"ON" or inactive/"OFF", is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein.), it may be a spatially restricted promoter (i.e., transcriptional control element, enhancer, etc.)(e.g., tissue specific promoter, cell type specific promoter, etc.), and it may be a temporally restricted promoter (i.e., the promoter is in the "ON" state or "OFF" state during specific stages of embryonic development or during specific stages of a biological process, e.g., hair follicle cycle in mice).

[00477] Suitable promoters can be derived from viruses and can therefore be referred to as viral promoters, or they can be derived from any organism, including prokaryotic or eukaryotic

organisms. Suitable promoters can be used to drive expression by any RNA polymerase (e.g., pol I, pol II, pol III). Exemplary promoters include, but are not limited to the SV40 early promoter, mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), a rous sarcoma virus (RSV) promoter, a human U6 small nuclear promoter (U6) (Miyagishi et al. , Nature Biotechnology 20, 497 - 500 (2002)), an enhanced U6 promoter (e.g., Xia et al., Nucleic Acids Res. 2003 Sep 1;31(17)), a human H1 promoter (H1), and the like.

[00478] Examples of inducible promoters include, but are not limited toT7 RNA polymerase promoter, T3 RNA polymerase promoter, Isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, lactose induced promoter, heat shock promoter, Tetracycline-regulated promoter (e.g., Tet-ON, Tet-OFF, etc.), Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc. Inducible promoters can therefore be regulated by molecules including, but not limited to, doxycycline; RNA polymerase, e.g., T7 RNA polymerase; an estrogen receptor; an estrogen receptor fusion; etc.

[00479] In some embodiments, the promoter is a spatially restricted promoter (i.e., cell type specific promoter, tissue specific promoter, etc.) such that in a multi-cellular organism, the promoter is active (i.e., "ON") in a subset of specific cells. Spatially restricted promoters may also be referred to as enhancers, transcriptional control elements, control sequences, etc. Any convenient spatially restricted promoter may be used and the choice of suitable promoter (e.g., a brain specific promoter, a promoter that drives expression in a subset of neurons, a promoter that drives expression in the germline, a promoter that drives expression in the lungs, a promoter that drives expression in muscles, a promoter that drives expression in islet cells of the pancreas, etc.) will depend on the organism. For example, various spatially restricted promoters are known for plants, flies, worms, mammals, mice, etc. Thus, a spatially restricted promoter can be used to regulate the expression of a nucleic acid encoding a subject site-directed polypeptide in a wide variety of different tissues and cell types, depending on the organism. Some spatially restricted promoters are also temporally restricted such that the promoter is in the "ON" state or "OFF" state during specific stages of embryonic development or during specific stages of a biological process (e.g., hair follicle cycle in mice).

[00480] For illustration purposes, examples of spatially restricted promoters include, but are not limited to, neuron-specific promoters, adipocyte-specific promoters, cardiomyocyte-specific promoters, smooth muscle-specific promoters, photoreceptor-specific promoters, etc. Neuron-specific spatially restricted promoters include, but are not limited to, a neuron-specific enolase (NSE) promoter (see, e.g., EMBL HSENO2, X51956); an aromatic amino acid decarboxylase (AADC)

promoter; a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al. (1987) *Cell* 51:7-19; and Llewellyn, et al. (2010) *Nat. Med.* 16(10):1161-1166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., Oh et al. (2009) *Gene Ther* 16:437; Sasaoka et al. (1992) *Mol. Brain Res.* 16:274; Boundy et al. (1998) *J. Neurosci.* 18:9989; and Kaneda et al. (1991) *Neuron* 6:583-594); a GnRH promoter (see, e.g., Radovick et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:3402-3406); an L7 promoter (see, e.g., Oberdick et al. (1990) *Science* 248:223-226); a DNMT promoter (see, e.g., Bartge et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al. (1988) *EMBO J.* 17:3793-3805); a myelin basic protein (MBP) promoter; a Ca^{2+}-calmodulin-dependent protein kinase II-alpha (CamKIIα) promoter (see, e.g., Mayford et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:13250; and Casanova et al. (2001) *Genesis* 31:37); a CMV enhancer/platelet-derived growth factor-β promoter (see, e.g., Liu et al. (2004) *Gene Therapy* 11:52-60); and the like.

[00481] Adipocyte-specific spatially restricted promoters include, but are not limited to aP2 gene promoter/enhancer, e.g., a region from -5.4 kb to +21 bp of a human aP2 gene (see, e.g., Tozzo et al. (1997) *Endocrinol.* 138:1604; Ross et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:9590; and Pavjani et al. (2005) *Nat. Med.* 11:797); a glucose transporter-4 (GLUT4) promoter (see, e.g., Knight et al. (2003) *Proc. Natl. Acad. Sci. USA* 100:14725); a fatty acid translocase (FAT/CD36) promoter (see, e.g., Kuriki et al. (2002) *Biol. Pharm. Bull.* 25:1476; and Sato et al. (2002) *J. Biol. Chem.* 277:15703); a stearoyl-CoA desaturase-1 (SCD1) promoter (Tabor et al. (1999) *J. Biol. Chem.* 274:20603); a leptin promoter (see, e.g., Mason et al. (1998) *Endocrinol.* 139:1013; and Chen et al. (1999) *Biochem. Biophys. Res. Comm.* 262:187); an adiponectin promoter (see, e.g., Kita et al. (2005) *Biochem. Biophys. Res. Comm.* 331:484; and Chakrabarti (2010) *Endocrinol.* 151:2408); an adipsin promoter (see, e.g., Platt et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:7490); a resistin promoter (see, e.g., Seo et al. (2003) *Molec. Endocrinol.* 17:1522); and the like.

[00482] Cardiomyocyte-specific spatially restricted promoters include, but are not limited to control sequences derived from the following genes: myosin light chain-2, α-myosin heavy chain, AE3, cardiac troponin C, cardiac actin, and the like. Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617; and Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

[00483] Smooth muscle-specific spatially restricted promoters include, but are not limited to an SM22α promoter (see, e.g., Akyürek et al. (2000) *Mol. Med.* 6:983; and U.S. Patent No. 7,169,874); a

smoothelin promoter (see, e.g., WO 2001/018048); an α-smooth muscle actin promoter; and the like. For example, a 0.4 kb region of the SM22α promoter, within which lie two CArG elements, has been shown to mediate vascular smooth muscle cell-specific expression (see, e.g., Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; and Moessler, et al. (1996) Development 122, 2415-2425).

[00484] Photoreceptor-specific spatially restricted promoters include, but are not limited to, a rhodopsin promoter; a rhodopsin kinase promoter (Young et al. (2003) *Ophthalmol. Vis. Sci.* 44:4076); a beta phosphodiesterase gene promoter (Nicoud et al. (2007) *J. Gene Med.* 9:1015); a retinitis pigmentosa gene promoter (Nicoud et al. (2007) *supra*); an interphotoreceptor retinoid-binding protein (IRBP) gene enhancer (Nicoud et al. (2007) *supra*); an IRBP gene promoter (Yokoyama et al. (1992) *Exp Eye Res.* 55:225); and the like.

LIBRARIES

[00485] The present disclosure provides a library of DNA-targeting RNAs. The present disclosure provides a library of nucleic acids comprising nucleotides encoding DNA-targeting RNAs. A subject library of nucleic acids comprising nucleotides encoding DNA-targeting RNAs can comprises a library of recombinant expression vectors comprising nucleotides encoding the DNA-targeting RNAs.

[00486] A subject library can comprise from about 10 individual members to about 10^{12} individual members; e.g., a subject library can comprise from about 10 individual members to about 10^{2} individual members, from about 10^{2} individual members to about 10^{3} individual members, from about 10^{3} individual members to about 10^{5} individual members, from about 10^{5} individual members to about 10^{7} individual members, from about 10^{7} individual members to about 10^{9} individual members, or from about 10^{9} individual members to about 10^{12} individual members.

[00487] An "individual member" of a subject library differs from other members of the library in the nucleotide sequence of the DNA targeting segment of the DNA-targeting RNA. Thus, e.g., each individual member of a subject library can comprise the same or substantially the same nucleotide sequence of the protein-binding segment as all other members of the library; and can comprise the same or substantially the same nucleotide sequence of the transcriptional termination segment as all other members of the library; but differs from other members of the library in the nucleotide sequence of the DNA targeting segment of the DNA-targeting RNA. In this way, the library can comprise members that bind to different target nucleic acids.

UTILITY

[00488] A method for modulating transcription according to the present disclosure finds use in a variety of applications, which are also provided. Applications include research applications; diagnostic applications; industrial applications; and treatment applications.

[00489] Research applications include, e.g., determining the effect of reducing or increasing transcription of a target nucleic acid on, e.g., development, metabolism, expression of a downstream gene, and the like.

[00490] High through-put genomic analysis can be carried out using a subject transcription modulation method, in which only the DNA-targeting segment of the DNA-targeting RNA needs to be varied, while the protein-binding segment and the transcription termination segment can (in some cases) be held constant. A library (e.g., a subject library) comprising a plurality of nucleic acids used in the genomic analysis would include: a promoter operably linked to a DNA-targeting RNA-encoding nucleotide sequence, where each nucleic acid would include a different DNA-targeting segment, a common protein-binding segment, and a common transcription termination segment. A chip could contain over 5×10^4 unique DNA-targeting RNAs. Applications would include large-scale phenotyping, gene-to-function mapping, and meta-genomic analysis.

[00491] The subject methods disclosed herein find use in the field of metabolic engineering. Because transcription levels can be efficiently and predictably controlled by designing an appropriate DNA-targeting RNA, as disclosed herein, the activity of metabolic pathways (e.g., biosynthetic pathways) can be precisely controlled and tuned by controlling the level of specific enzymes (e.g., via increased or decreased transcription) within a metabolic pathway of interest. Metabolic pathways of interest include those used for chemical (fine chemicals, fuel, antibiotics, toxins, agonists, antagonists, etc.) and/or drug production.

[00492] Biosynthetic pathways of interest include but are not limited to (1) the mevalonate pathway (e.g., HMG-CoA reductase pathway) (converts acetyl-CoA to dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP), which are used for the biosynthesis of a wide variety of biomolecules including terpenoids/isoprenoids), (2) the non-mevalonate pathway (i.e., the "2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway" or "MEP/DOXP pathway" or "DXP pathway")(also produces DMAPP and IPP, instead by converting pyruvate and glyceraldehyde 3-phosphate into DMAPP and IPP via an alternative pathway to the mevalonate pathway), (3) the polyketide synthesis pathway (produces a variety of polyketides via a variety of polyketide synthase enzymes. Polyketides include naturally occurring small molecules used for chemotherapy (e. g., tetracyclin, and macrolides) and

industrially important polyketides include rapamycin (immunosuppressant), erythromycin (antibiotic), lovastatin (anticholesterol drug), and epothilone B (anticancer drug)), (4) fatty acid synthesis pathways, (5) the DAHP (3-deoxy-D-arabino-heptulosonate 7-phosphate) synthesis pathway, (6) pathways that produce potential biofuels (such as short-chain alcohols and alkane, fatty acid methyl esters and fatty alcohols, isoprenoids, etc.), etc.

Networks and Cascades

[00493] The methods disclosed herein can be used to design integrated networks (i.e., a cascade or cascades) of control. For example, a subject DNA-targeting RNA / variant Cas9 site-directed polypeptide may be used to control (i.e., modulate, e.g., increase, decrease) the expression of another DNA-tageting RNA or another subject variant Cas9 site-directed polypeptide. For example, a first DNA-targeting RNA may be designed to target the modulation of transcription of a second chimeric dCas9 polypeptide with a function that is different than the first variant Cas9 site-directed polypeptide (e.g., methyltransferase activity, demethylase activity, acetyltansferase activity, deacetylase activity, etc.). In addition, because different dCas9 proteins (e.g., derived from different species) may require a different Cas9 handle (i.e., protein binding segment), the second chimeric dCas9 polypeptide can be derived from a different species than the first dCas9 polypeptide above. Thus, in some cases, the second chimeric dCas9 polypeptide can be selected such that it may not interact with the first DNA-targeting RNA. In other cases, the second chimeric dCas9 polypeptide can be selected such that it does interact with the first DNA-targeting RNA. In some such cases, the activities of the two (or more) dCas9 proteins may compete (e.g., if the polypeptides have opposing activities) or may synergize (e.g., if the polypeptides have similar or synergistic activities). Likewise, as noted above, any of the complexes (i.e., DNA-targeting RNA / dCas9 polypeptide) in the network can be designed to control other DNA-targeting RNAs or dCas9 polypeptides. Because a subject DNA-targeting RNA and subject variant Cas9 site-directed polypeptide can be targeted to any desired DNA sequence, the methods described herein can be used to control and regulate the expression of any desired target. The integrated networks (i.e., cascades of interactions) that can be designed range from very simple to very complex, and are without limit.

[00494] In a network wherein two or more components (e.g., DNA-targeting RNAs, activator-RNAs, targeter-RNAs, or dCas9 polypeptides) are each under regulatory control of another DNA-targeting RNA/dCas9 polypeptide complex, the level of expression of one component of the network may affect the level of expression (e.g., may increase or decrease the expression) of another component of the network. Through this mechanism, the expression of one component may affect the expression of a different component in the same network, and the network may include a mix of components that increase the expression of other components, as well as

components that decrease the expression of other components. As would be readily understood by one of skill in the art, the above examples whereby the level of expression of one component may affect the level of expression of one or more different component(s) are for illustrative purposes, and are not limiting. An additional layer of complexity may be optionally introduced into a network when one or more components are modified (as described above) to be manipulable (i.e., under experimental control, e.g., temperature control; drug control, i.e., drug inducible control; light control; etc.).

[00495] As one non-limiting example, a first DNA-targeting RNA can bind to the promoter of a second DNA-targeting RNA, which controls the expression of a target therapeutic/metabolic gene. In such a case, conditional expression of the first DNA-targeting RNA indirectly activates the therapeutic/metabolic gene. RNA cascades of this type are useful, for example, for easily converting a repressor into an activator, and can be used to control the logics or dynamics of expression of a target gene.

[00496] A subject transcription modulation method can also be used for drug discovery and target validation.

KITS

[00497] The present disclosure provides a kit for carrying out a subject method. A subject kit comprises: a) a DNA-targeting RNA of the present disclosure, or a nucleic acid comprising a nucleotide sequence encoding the DNA-targeting RNA, wherein the DNA-targeting RNA comprises: i)) a first segment comprising a nucleotide sequence that is complementary to a target sequence in the target DNA; ii)) a second segment that interacts with a site-directed polypeptide; and iii) a transcriptional terminator; and b) a buffer. In some cases, the nucleic acid comprising a nucleotide sequence encoding the DNA-targeting RNA further comprises a nucleotide sequence encoding a variant Cas9 site-directed polypeptide that exhibits reduced endodeoxyribonuclease activity relative to wild-type Cas9.

[00498] In some cases, a subject kit further comprises a variant Cas9 site-directed polypeptide that exhibits reduced endodeoxyribonuclease activity relative to wild-type Cas9.

[00499] In some cases, a subject kit further comprises a nucleic acid comprising a nucleotide sequence encoding a variant Cas9 site-directed polypeptide that exhibits reduced endodeoxyribonuclease activity relative to wild-type Cas9.

[00500] A subject can further include one or more additional reagents, where such additional reagents can be selected from: a buffer; a wash buffer; a control reagent; a control expression vector or RNA polynucleotide; a reagent for *in vitro* production of the variant Cas9 site-directed

polypeptide from DNA; and the like. In some cases, the variant Cas9 site-directed polypeptide included in a subject kit is a fusion variant Cas9 site-directed polypeptide, as described above.

[00501] Components of a subject kit can be in separate containers; or can be combined in a single container.

[00502] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, flash drive, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

EXAMPLES

[00503] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); pl, picoliter(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); i.m., intramuscular(ly); i.p., intraperitoneal(ly); s.c., subcutaneous(ly); and the like.

Example 1: Use of Cas9 to generate modifications in target DNA.

MATERIALS AND METHODS

Bacterial strains and culture conditions

[00504] *Streptococcus pyogenes*, cultured in THY medium (Todd Hewitt Broth (THB, Bacto, Becton Dickinson) supplemented with 0.2% yeast extract (Oxoid)) or on TSA (trypticase soy agar, BBL, Becton Dickinson) supplemented with 3% sheep blood, was incubated at 37°C in an atmosphere supplemented with 5% CO_2 without shaking. Escherichia coli, cultured in Luria-Bertani (LB) medium and agar, was incubated at 37°C with shaking. When required, suitable antibiotics were added to the medium at the following final concentrations: ampicillin, 100 μg/ml for *E. coli*; chloramphenicol, 33 μg/ml for *Escherichia coli*; kanamycin, 25 μg/ml for *E. coli* and 300 μg/ml for *S. pyogenes*. Bacterial cell growth was monitored periodically by measuring the optical density of culture aliquots at 620 nm using a microplate reader (SLT Spectra Reader).

Transformation of bacterial cells

[00505] Plasmid DNA transformation into *E. coli* cells was performed according to a standard heat shock protocol. Transformation of *S. pyogenes* was performed as previously described with some modifications. The transformation assay performed to monitor in vivo CRISPR/Cas activity on plasmid maintenance was essentially carried out as described previously. Briefly, electrocompetent cells of *S. pyogenes* were equalized to the same cell density and electroporated with 500 ng of plasmid DNA. Every transformation was plated two to three times and the experiment was performed three times independently with different batches of competent cells for statistical analysis. Transformation efficiencies were calculated as CFU (colony forming units) per μg of DNA. Control transformations were performed with sterile water and backbone vector pEC85.

DNA manipulations

[00506] DNA manipulations including DNA preparation, amplification, digestion, ligation, purification, agarose gel electrophoresis were performed according to standard techniques with minor modifications. Protospacer plasmids for the in vitro cleavage and *S. pyogenes* transformation assays were constructed as described previously (4). Additional pUC19-based protospacer plasmids for in vitro cleavage assays were generated by ligating annealed oligonucleotides between digested EcoRI and BamHI sites in pUC19. The GFP gene-containing plasmid has been described previously (41). Kits (Qiagen) were used for DNA purification and plasmid preparation. Plasmid mutagenesis was performed using QuikChange® II XL kit (Stratagene) or QuikChange site-directed mutagenesis kit (Agilent). VBC-Biotech Services, Sigma-Aldrich and Integrated DNA Technologies supplied the synthetic oligonucleotides and RNAs.

Oligonucleotides for *in vitro* transcription templates

[00507] Templates for *in vitro* transcribed CRISPR Type II-A tracrRNA and crRNAs of *S. pyogenes* (for tracrRNA – PCR on chr. DNA SF370; for crRNA – annealing of two oligonucleotides)

T7-tracrRNA (75nt)

OLEC1521 (F 5' tracrRNA): SEQ ID NO:340

OLEC1522 (R 3' tracrRNA): SEQ ID NO:341

T7-crRNA (template)

OLEC2176 (F crRNA-sp1): SEQ ID NO:342

OLEC2178 (R crRNA-sp1): SEQ ID NO:343

OLEC2177 (F crRNA-sp2): SEQ ID NO:344

OLEC2179 (R crRNA-sp2): SEQ ID NO:345

[00508] Templates for in vitro transcribed N. meningitidis tracrRNA and engineered crRNA-sp2 (for tracrRNA – PCR on chr. DNA Z2491; for crRNA – annealing of two oligonucleotides)

T7-tracrRNA

OLEC2205 (F predicted 5'): SEQ ID NO:346

OLEC2206 (R predicted 3'): SEQ ID NO:347

T7-crRNA (template)

OLEC2209 (F sp2(*speM*) + *N.m.* repeat): SEQ ID NO:348

OLEC2214 (R sp2(*speM*) + *N.m.* repeat): SEQ ID NO:349

[00509] Templates for in vitro transcribed L. innocua tracrRNA and engineered crRNA-sp2 (for tracrRNA – PCR on chr. DNA Clip11262; for crRNA – annealing of two oligonucleotides)

T7-tracrRNA

OLEC2203 (F predicted 5'): SEQ ID NO:350

OLEC2204 (R predicted 3'): SEQ ID NO:351

T7-crRNA (template)

OLEC2207 (F sp2(*speM*) + *L.in.* repeat): SEQ ID NO:352

OLEC2212 (R sp2(*speM*) + *L.in.* repeat): SEQ ID NO:353

Oligonucleotides for constructing plasmids with protospacer for in vitro and in vivo studies

[00510] Plasmids for *speM* (spacer 2 (CRISPR Type II-A, SF370; protospacer prophage ø8232.3 from MGAS8232) analysis *in vitro* and in *S. pyogenes* (template: chr. DNA MGAS8232 or plasmids containing *speM* fragments)

pEC287

OLEC1555 (F *speM*): SEQ ID NO:354

OLEC1556 (R *speM*): SEQ ID NO:355

pEC488

OLEC2145 (F *speM*): SEQ ID NO:356

OLEC2146 (R *speM*): SEQ ID NO:357

pEC370

OLEC1593 (F pEC488 protospacer 2 A22G): SEQ ID NO:358

OLEC1594 (R pEC488 protospacer 2 A22G): SEQ ID NO:359

pEC371

OLEC1595 (F pEC488 protospacer 2 T10C): SEQ ID NO:360

OLEC1596 (R pEC488 protospacer 2 T10C): SEQ ID NO:361

pEC372

OLEC2185 (F pEC488 protospacer 2 T7A): SEQ ID NO:362

OLEC2186 (R pEC488 protospacer 2 T7A): SEQ ID NO:363

pEC373

OLEC2187 (F pEC488 protospacer 2 A6T): SEQ ID NO:364

OLEC2188 (R pEC488 protospacer 2 A6T): SEQ ID NO:365

pEC374

OLEC2235 (F pEC488 protospacer 2 A5T): SEQ ID NO:366

OLEC2236 (R pEC488 protospacer 2 A5T): SEQ ID NO:367

pEC375

OLEC2233 (F pEC488 protospacer 2 A4T): SEQ ID NO:368

OLEC2234 (R pEC488 protospacer 2 A4T): SEQ ID NO:369

pEC376

OLEC2189 (F pEC488 protospacer 2 A3T): SEQ ID NO:370

OLEC2190 (R pEC488 protospacer 2 A3T): SEQ ID NO:371

pEC377

OLEC2191 (F pEC488 protospacer 2 PAM G1C): SEQ ID NO:372

OLEC2192 (R pEC488 protospacer 2 PAM G1C): SEQ ID NO:373

pEC378

OLEC2237 (F pEC488 protospacer 2 PAM GG1, 2CC): SEQ ID NO:374

OLEC2238 (R pEC488 protospacer 2 PAM GG1, 2CC): SEQ ID NO:375

[00511] Plasmids for SPy_0700 (spacer 1 (CRISPR Type II-A, SF370; protospacer prophage ø370.1 from SF370) analysis *in vitro* and in *S. pyogenes* (template: chr. DNA SF370 or plasmids containing SPy_0700 fragments)

pEC489

OLEC2106 (F Spy_0700): SEQ ID NO:376

OLEC2107 (R Spy_0700): SEQ ID NO:377

pEC573

OLEC2941 (F PAM TG1, 2GG): SEQ ID NO:378

OLEC2942 (R PAM TG1, 2GG): SEQ ID NO:379

[00512] Oligonucleotides for verification of plasmid constructs and cutting sites by sequencing analysis

ColE1 (pEC85)

oliRN228 (R sequencing): SEQ ID NO:380

speM (pEC287)

OLEC1557 (F sequencing): SEQ ID NO:381

OLEC1556 (R sequencing): SEQ ID NO:382

repDEG-pAMbeta1 (pEC85)

OLEC787 (F sequencing): SEQ ID NO:383

Oligonucleotides for *in vitro* cleavage assays

[00513] crRNA

Spacer 1 crRNA (1-42): SEQ ID NO:384

Spacer 2 crRNA (1-42): SEQ ID NO:385

Spacer 4 crRNA (1-42): SEQ ID NO:386

Spacer 2 crRNA (1-36): SEQ ID NO:387

Spacer 2 crRNA (1-32): SEQ ID NO:388

Spacer 2 crRNA (11-42): SEQ ID NO:389

[00514] tracrRNA

(4-89): SEQ ID NO:390

(15-89): SEQ ID NO:391

(23-89): SEQ ID NO:392

(15-53): SEQ ID NO:393

(15-44): SEQ ID NO:394

(15-36): SEQ ID NO:395

(23-53): SEQ ID NO:396

(23-48): SEQ ID NO:397

(23-44): SEQ ID NO:398

(1-26): SEQ ID NO:399

[00515] chimeric RNAs

Spacer 1 – chimera A: SEQ ID NO:400

Spacer 1 – chimera B: SEQ ID NO:401

Spacer 2 – chimera A: SEQ ID NO:402

Spacer 2 – chimera B: SEQ ID NO:403

Spacer 4 – chimera A: SEQ ID NO:404

Spacer 4 – chimera B: SEQ ID NO:405

GFP1: SEQ ID NO:406

GFP2: SEQ ID NO:407

GFP3: SEQ ID NO:408

GFP4: SEQ ID NO:409

GFP5: SEQ ID NO:410

[00516] DNA oligonucleotides as substrates for cleavage assays (protospacer in bold, PAM underlined)

protospacer 1 - complementary – WT: SEQ ID NO:411

protospacer 1 - noncomplementary – WT: SEQ ID NO:412

protospacer 2 - complementary – WT: SEQ ID NO:413

protospacer 2 - noncomplementary – WT: SEQ ID NO:414

protospacer 4 - complementary – WT: SEQ ID NO:415

protospacer 4 - noncomplementary – WT: SEQ ID NO:416

protospacer 2 - complementary – PAM1: SEQ ID NO:417

protospacer 2 - noncomplementary – PAM1: SEQ ID NO:418

protospacer 2 - complementary – PAM2: SEQ ID NO:419

protospacer 2 - noncomplementary – PAM2: SEQ ID NO:420

protospacer 4 - complementary – PAM1: SEQ ID NO:421

protospacer 4 - noncomplementary – PAM1: SEQ ID NO:422

protospacer 4 - complementary – PAM2: SEQ ID NO:423

protospacer 4 - noncomplementary – PAM2: SEQ ID NO:424

In vitro transcription and purification of RNA

[00517] RNA was in vitro transcribed using T7 Flash in vitro Transcription Kit (Epicentre, Illumina company) and PCR-generated DNA templates carrying a T7 promoter sequence. RNA was gel-purified and quality-checked prior to use. The primers used for the preparation of RNA templates from S. pyogenes SF370, Listeria innocua Clip 11262 and Neisseria meningitidis A Z2491 are described above.

Protein purification

[00518] The sequence encoding Cas9 (residues 1-1368) was PCRamplified from the genomic DNA of S. pyogenes SF370 and inserted into a custom pET-based expression vector using ligation-independent cloning (LIC). The resulting fusion construct contained an N-terminal hexahistidine-maltose binding protein (His6- MBP) tag, followed by a peptide sequence containing a tobacco etch virus (TEV) protease cleavage site. The protein was expressed in E. coli strain BL21 Rosetta 2 (DE3) (EMD Biosciences), grown in 2xTY medium at 18°C for 16 h following induction with 0.2 mM IPTG. The protein was purified by a combination of affinity, ion exchange and size exclusion chromatographic steps. Briefly, cells were lysed in 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM TCEP (supplemented with protease inhibitor cocktail (Roche)) in a homogenizer (Avestin). Clarified lysate was bound in batch to Ni-NTA agarose (Qiagen). The resin was washed extensively with 20 mM Tris pH 8.0, 500 mM NaCl and the bound protein was eluted in 20 mM Tris pH 8.0, 250 mM NaCl, 10% glycerol. The His6-MBP affinity tag was removed by cleavage with TEV protease, while the protein was dialyzed overnight against 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM TCEP, 10% glycerol. The cleaved Cas9 protein was separated from the fusion tag by purification on a 5 ml SP Sepharose HiTrap column (GE Life Sciences), eluting with a linear gradient of 100 mM – 1 M KCl. The protein was further purified by size exclusion chromatography on a Superdex 200 16/60 column in 20 mM HEPES pH 7.5, 150 mM KCl and 1 mM TCEP. Eluted protein was concentrated to ~8 mg/ml, flash-frozen in liquid nitrogen and stored at -80°C. Cas9 D10A, H840A and D10A/H840A point mutants were generated using the QuikChange site-directed mutagenesis kit (Agilent) and confirmed by DNA sequencing. The proteins were purified following the same procedure as for the wildtype Cas9 protein.

[00519] Cas9 orthologs from Streptococcus thermophilus (LMD-9,YP_820832.1), L. innocua (Clip11262, NP_472073.1), Campylobacter jejuni (subsp. jejuni NCTC 11168, YP_002344900.1) and N. meningitidis (Z2491, YP_002342100.1) were expressed in BL21

Rosetta (DE3) pLysS cells (Novagen) as His6-MBP (N. meningitidis and C. jejuni), His6-Thioredoxin (L. innocua) and His6-GST (S. thermophilus) fusion proteins, and purified essentially as for S. pyogenes Cas9 with the following modifications. Due to large amounts of co-purifying nucleic acids, all four Cas9 proteins were purified by an additional heparin sepharose step prior to gel filtration, eluting the bound protein with a linear gradient of 100 mM – 2 M KCl. This successfully removed nucleic acid contamination from the C. jejuni, N. meningitidis and L. innocua proteins, but failed to remove co-purifying nucleic acids from the S. thermophilus Cas9 preparation. All proteins were concentrated to 1-8 mg/ml in 20 mM HEPES pH 7.5, 150 mM KCl and 1 mM TCEP, flash-frozen in liquid N2 and stored at -80°C.

Plasmid DNA cleavage assay

[00520] Synthetic or in vitro-transcribed tracrRNA and crRNA were pre-annealed prior to the reaction by heating to 95°C and slowly cooling down to room temperature. Native or restriction digest-linearized plasmid DNA (300 ng (~8 nM)) was incubated for 60 min at 37°C with purified Cas9 protein (50-500 nM) and tracrRNA:crRNA duplex (50-500 nM, 1:1) in a Cas9 plasmid cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) with or without 10 mM $MgCl_2$. The reactions were stopped with 5X DNA loading buffer containing 250 mM EDTA, resolved by 0.8 or 1% agarose gel electrophoresis and visualized by ethidium bromide staining. For the Cas9 mutant cleavage assays, the reactions were stopped with 5X SDS loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA) prior to loading on the agarose gel.

Metal-dependent cleavage assay

[00521] Protospacer 2 plasmid DNA (5 nM) was incubated for 1 h at 37°C with Cas9 (50 nM) pre-incubated with 50 nM tracrRNA:crRNA-sp2 in cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) supplemented with 1, 5 or 10 mM $MgCl_2$, 1 or 10 mM of $MnCl_2$, $CaCl_2$, $ZnCl_2$, $CoCl_2$, $NiSO_4$ or $CuSO_4$. The reaction was stopped by adding 5X SDS loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA), resolved by 1% agarose gel electrophoresis and visualized by ethidium bromide staining.

Single-turnover assay

[00522] Cas9 (25 nM) was pre-incubated 15 min at 37°C in cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA) with duplexed tracrRNA:crRNA-sp2 (25 nM, 1:1) or both RNAs (25 nM) not preannealed and the reaction was started by adding protospacer 2 plasmid DNA (5 nM). The reaction mix was incubated at 37°C. At defined time intervals, samples were withdrawn from the reaction, 5X SDS loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA) was added to stop the reaction and the cleavage was monitored by 1% agarose gel electrophoresis and ethidium bromide staining. The same was done for the single turnover kinetics without pre-incubation of Cas9 and RNA, where protospacer 2 plasmid DNA

(5 nM) was mixed in cleavage buffer with duplex tracrRNA:crRNA-sp2 (25 nM) or both RNAs (25 nM) not pre-annealed, and the reaction was started by addition of Cas9 (25 nM). Percentage of cleavage was analyzed by densitometry and the average of three independent experiments was plotted against time. The data were fit by nonlinear regression analysis and the cleavage rates (k_{obs} [min^{-1}]) were calculated.

Multiple-turnover assay

[00523] Cas9 (1 nM) was pre-incubated for 15 min at 37°C in cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA) with pre-annealed tracrRNA:crRNA-sp2 (1 nM, 1:1). The reaction was started by addition of protospacer 2 plasmid DNA (5 nM). At defined time intervals, samples were withdrawn and the reaction was stopped by adding 5X SDS loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA). The cleavage reaction was resolved by 1% agarose gel electrophoresis, stained with ethidium bromide and the percentage of cleavage was analyzed by densitometry. The results of four independent experiments were plotted against time (min).

Oligonucleotide DNA cleavage assay

[00524] DNA oligonucleotides (10 pmol) were radiolabeled by incubating with 5 units T4 polynucleotide kinase (New England Biolabs) and ~3–6 pmol (~20–40 mCi) [γ-32P]-ATP (Promega) in 1X T4 polynucleotide kinase reaction buffer at 37°C for 30 min, in a 50 μL reaction. After heat inactivation (65°C for 20 min), reactions were purified through an Illustra MicroSpin G-25 column (GE Healthcare) to remove unincorporated label. Duplex substrates (100 nM) were generated by annealing labeled oligonucleotides with equimolar amounts of unlabeled complementary oligonucleotide at 95°C for 3 min, followed by slow cooling to room temperature. For cleavage assays, tracrRNA and crRNA were annealed by heating to 95°C for 30 s, followed by slow cooling to room temperature. Cas9 (500 nM final concentration) was pre-incubated with the annealed tracrRNA:crRNA duplex (500 nM) in cleavage assay buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol) in a total volume of 9 μl. Reactions were initiated by the addition of 1 μl target DNA (10 nM) and incubated for 1 h at 37°C. Reactions were quenched by the addition of 20 μl of loading dye (5 mM EDTA, 0.025% SDS, 5% glycerol in formamide) and heated to 95°C for 5 min. Cleavage products were resolved on 12% denaturing polyacrylamide gels containing 7 M urea and visualized by phosphorimaging (Storm, GE Life Sciences). Cleavage assays testing PAM requirements (Figure 13B) were carried out using DNA duplex substrates that had been pre-annealed and purified on an 8% native acrylamide gel, and subsequently radiolabeled at both 5' ends. The reactions were set-up and analyzed as above.

Electrophoretic mobility shift assays

[00525] Target DNA duplexes were formed by mixing of each strand (10 nmol) in deionized water, heating to 95°C for 3 min and slow cooling to room temperature. All DNAs were purified on 8% native gels containing 1X TBE. DNA bands were visualized by UV shadowing, excised, and eluted by soaking gel pieces in DEPC-treated H_2O. Eluted DNA was ethanol precipitated and dissolved in DEPC-treated H_2O. DNA samples were 5' end labeled with [γ-32P]-ATP using T4 polynucleotide kinase (New England Biolabs) for 30 min at 37°C. PNK was heat denatured at 65°C for 20 min, and unincorporated radiolabel was removed using an Illustra MicroSpin G-25 column (GE Healthcare). Binding assays were performed in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT and 10% glycerol in a total volume of 10 µl. Cas9 D10A/H840A double mutant was programmed with equimolar amounts of pre-annealed tracrRNA:crRNA duplex and titrated from 100 pM to 1 µM. Radiolabeled DNA was added to a final concentration of 20 pM. Samples were incubated for 1 h at 37°C and resolved at 4°C on an 8% native polyacrylamide gel containing 1X TBE and 5 mM $MgCl_2$. Gels were dried and DNA visualized by phosphorimaging.

In silico analysis of DNA and protein sequences

[00526] Vector NTI package (Invitrogen) was used for DNA sequence analysis (Vector NTI) and comparative sequence analysis of proteins (AlignX).

In silico modeling of RNA structure and co-folding

[00527] In silico predictions were performed using the Vienna RNA package algorithms (42, 43). RNA secondary structures and co-folding models were predicted with RNAfold and RNAcofold, respectively and visualized with VARNA (44).

RESULTS

[00528] Bacteria and archaea have evolved RNA mediated adaptive defense systems called clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) that protect organisms from invading viruses and plasmids (1–3). We show that in a subset of these systems, the mature crRNA that is base-paired to trans-activating crRNA (tracrRNA) forms a two-RNA structure that directs the CRISPR-associated protein Cas9 to introduce double-stranded (ds) breaks in target DNA. At sites complementary to the crRNA-guide sequence, the Cas9 HNH nuclease domain cleaves the complementary strand, whereas the Cas9 RuvC-like domain cleaves the noncomplementary strand. The dual-tracrRNA:crRNA, when engineered as a single RNA chimera, also directs sequence-specific Cas9 dsDNA cleavage. These studies reveal a family of endonucleases that use dual-RNAs for site-specific DNA cleavage and highlight the ability to exploit the system for RNA-programmable genome editing.

[00529] CRISPR/Cas defense systems rely on small RNAs for sequence-specific detection and silencing of foreign nucleic acids. CRISPR/Cas systems are composed of cas genes organized in operon(s) and CRISPR array(s) consisting of genome-targeting sequences (called spacers) interspersed with identical repeats (1–3). CRISPR/Cas-mediated immunity occurs in three steps. In the adaptive phase, bacteria and archaea harboring one or more CRISPR loci respond to viral or plasmid challenge by integrating short fragments of foreign sequence (protospacers) into the host chromosome at the proximal end of the CRISPR array (1–3). In the expression and interference phases, transcription of the repeat spacer element into precursor CRISPR RNA (pre-crRNA) molecules followed by enzymatic cleavage yields the short crRNAs that can pair with complementary protospacer sequences of invading viral or plasmid targets (4–11). Target recognition by crRNAs directs the silencing of the foreign sequences by means of Cas proteins that function in complex with the crRNAs (10, 12–20).

[00530] There are three types of CRISPR/Cas systems (21–23). The type I and III systems share some overarching features: specialized Cas endonucleases process the pre-crRNAs, and oncemature, each crRNA assembles into a large multi-Cas protein complex capable of recognizing and cleaving nucleic acids complementary to the crRNA. In contrast, type II systems process precrRNAs by a different mechanism in which a trans-activating crRNA (tracrRNA) complementary to the repeat sequences in pre-crRNA triggers processing by the double-stranded (ds) RNAspecific ribonuclease RNase III in the presence of the Cas9 (formerly Csn1) protein (Figure 15) (4, 24). Cas9 is thought to be the sole protein responsible for crRNA-guided silencing of foreign DNA (25–27).

[00531] We show that in type II systems, Cas9 proteins constitute a family of enzymes that require a base-paired structure formed between the activating tracrRNA and the targeting crRNA to cleave target dsDNA. Site-specific cleavage occurs at locations determined by both base-pairing complementarity between the crRNA and the target protospacer DNA and a short motif [referred to as the protospacer adjacent motif (PAM)] juxtaposed to the complementary region in the target DNA. Our study further demonstrates that the Cas9 endonuclease family can be programmed with single RNA molecules to cleave specificDNA sites, thereby facilitating the development of a simple and versatile RNA-directed system to generate dsDNA breaks for genome targeting and editing.

Cas9 is a DNA endonuclease guided by two RNAs

[00532] Cas9, the hallmark protein of type II systems, has been hypothesized to be involved in both crRNA maturation and crRNA-guided DNA interference (Figure 15) (4, 25–27). Cas9 is involved in crRNA maturation (4), but its direct participation in target DNA destruction has not been investigated. To test whether and how Cas9 might be capable of target DNA cleavage, we

used an overexpression system to purify Cas9 protein derived from the pathogen Streptococcus pyogenes (Figure 16, see supplementary materials and methods) and tested its ability to cleave a plasmid DNA or an oligonucleotide duplex bearing a protospacer sequence complementary to a mature crRNA, and a bona fide PAM.We found that mature crRNA alone was incapable of directing Cas9-catalyzed plasmid DNA cleavage (Figure 10A and Figure 17A). However, addition of tracrRNA, which can pair with the repeat sequence of crRNA and is essential to crRNA maturation in this system, triggered Cas9 to cleave plasmidDNA (Figure 10A and Figure 17A). The cleavage reaction required both magnesium and the presence of a crRNA sequence complementary to the DNA; a crRNA capable of tracrRNAbase pairing but containing a noncognate target DNA-binding sequence did not support Cas9-catalyzed plasmid cleavage (Figure 10A; Figure 17A, compare crRNA-sp2 to crRNA-sp1; and Figure 18A). We obtained similar results with a short linear dsDNA substrate (Figure 10B and Figure 17, B and C). Thus, the trans-activating tracrRNA is a small noncoding RNAwith two critical functions: triggering pre-crRNA processing by the enzyme RNase III (4) and subsequently activating crRNA-guided DNA cleavage by Cas9.

[00533] Cleavage of both plasmid and short linear dsDNA by tracrRNA:crRNA-guided Cas9 is sitespecific (Figure 10, C to E, and Figure 19, A and B). Plasmid DNA cleavage produced blunt ends at a position three base pairs upstream of the PAM sequence (Figure 10, C and E, and Figure 19, A and C) (26). Similarly, within short dsDNA duplexes, the DNA strand that is complementary to the target-binding sequence in the crRNA (the complementary strand) is cleaved at a site three base pairs upstream of the PAM(Figure 10, D and E, and Figure 19, B and C). The noncomplementary DNA strand is cleaved at one or more sites within three to eight base pairs upstream of the PAM. Further investigation revealed that the noncomplementary strand is first cleaved endonucleolytically and subsequently trimmed by a 3′-5′ exonuclease activity (Figure 18B). The cleavage rates by Cas9 under single-turnover conditions ranged from 0.3 to 1 min−1, comparable to those of restriction endonucleases (Figure 20A), whereas incubation of wildtype (WT) Cas9-tracrRNA:crRNA complex with a fivefold molar excess of substrate DNA provided evidence that the dual-RNA–guided Cas9 is a multiple-turnover enzyme (Figure 20B). In contrast to the CRISPR type I Cascade complex (18), Cas9 cleaves both linearized and supercoiled plasmids (Figures 10A and 11A). Therefore, an invading plasmid can, in principle, be cleaved multiple times by Cas9 proteins programmed with different crRNAs.

[00534] Figure 10 (A) Cas9 was programmed with a 42-nucleotide crRNA-sp2 (crRNA containing a spacer 2 sequence) in the presence or absence of 75-nucleotide tracrRNA. The complex was added to circular or XhoI-linearized plasmid DNA bearing a sequence complementary to spacer 2 and a functional PAM. crRNA-sp1, specificity control; M, DNA marker; kbp, kilo–base pair.

See Figure 17A. **(B)** Cas9 was programmed with crRNA-sp2 and tracrRNA (nucleotides 4 to 89). The complex was incubated with double- or single-stranded DNAs harboring a sequence complementary to spacer 2 and a functional PAM (4). The complementary or noncomplementary strands of the DNA were 5′-radiolabeled and annealed with a nonlabeled partner strand. nt, nucleotides. See Figure 17, B and C. **(C)** Sequencing analysis of cleavage products from Figure 10A. Termination of primer extension in the sequencing reaction indicates the position of the cleavage site. The 3′ terminal A overhang (asterisks) is an artifact of the sequencing reaction. See Figure 19, A and C. **(D)** The cleavage products from Figure 10B were analyzed alongside 5′ end-labeled size markers derived from the complementary and noncomplementary strands of the target DNA duplex. M, marker; P, cleavage product. See Figure 19, B and C. **(E)** Schematic representation of tracrRNA, crRNA-sp2, and protospacer 2 DNA sequences. Regions of crRNA complementarity to tracrRNA (overline) and the protospacer DNA (underline) are represented. The PAM sequence is labeled; cleavage sites mapped in (C) and (D) are represented by white-filled arrows (C), a black-filled arrow [(D), complementary strand], and a black bar [(D), noncomplementary strand].

[00535] Figure 15 depicts the type II RNA-mediated CRISPR/Cas immune pathway. The expression and interference steps are represented in the drawing. The type II CRISPR/Cas loci are composed of an operon of four genes encoding the proteins Cas9, Cas1, Cas2 and Csn2, a CRISPR array consisting of a leader sequence followed by identical repeats (black rectangles) interspersed with unique genome-targeting spacers (diamonds) and a sequence encoding the trans-activating tracrRNA. Represented here is the type II CRISPR/Cas locus of S. pyogenes SF370 (Accession number NC_002737) (4). Experimentally confirmed promoters and transcriptional terminator in this locus are indicated (4). The CRISPR array is transcribed as a precursor CRISPR RNA (pre-crRNA) molecule that undergoes a maturation process specific to the type II systems (4). In S. pyogenes SF370, tracrRNA is transcribed as two primary transcripts of 171 and 89 nt in length that have complementarity to each repeat of the pre-crRNA. The first processing event involves pairing of tracrRNA to pre-crRNA, forming a duplex RNA that is recognized and cleaved by the housekeeping endoribonuclease RNase III in the presence of the Cas9 protein. RNase III-mediated cleavage of the duplex RNA generates a 75-nt processed tracrRNA and a 66-nt intermediate crRNAs consisting of a central region containing a sequence of one spacer, flanked by portions of the repeat sequence. A second processing event, mediated by unknown ribonuclease(s), leads to the formation of mature crRNAs of 39 to 42 nt in length consisting of 5'-terminal spacer-derived guide sequence and repeat-derived 3'-terminal sequence. Following the first and second processing events, mature tracrRNA remains paired to the mature crRNAs and bound to the Cas9 protein. In this ternary complex, the dual

tracrRNA:crRNA structure acts as guide RNA that directs the endonuclease Cas9 to the cognate target DNA. Target recognition by the Cas9-tracrRNA:crRNA complex is initiated by scanning the invading DNA molecule for homology between the protospacer sequence in the target DNA and the spacer-derived sequence in the crRNA. In addition to the DNA protospacer-crRNA spacer complementarity, DNA targeting requires the presence of a short motif (NGG, where N can be any nucleotide) adjacent to the protospacer (protospacer adjacent motif - PAM). Following pairing between the dual-RNA and the protospacer sequence, an R-loop is formed and Cas9 subsequently introduces a double-stranded break (DSB) in the DNA. Cleavage of target DNA by Cas9 requires two catalytic domains in the protein. At a specific site relative to the PAM, the HNH domain cleaves the complementary strand of the DNA while the RuvC-like domain cleaves the noncomplementary strand.

[00536] Figure 16 (A) *S. pyogenes* Cas9 was expressed in *E. coli* as a fusion protein containing an N-terminal His6-MBP tag and purified by a combination of affinity, ion exchange and size exclusion chromatographic steps. The affinity tag was removed by TEV protease cleavage following the affinity purification step. Shown is a chromatogram of the final size exclusion chromatography step on a Superdex 200 (16/60) column. Cas9 elutes as a single monomeric peak devoid of contaminating nucleic acids, as judged by the ratio of absorbances at 280 and 260 nm. Inset; eluted fractions were resolved by SDS-PAGE on a 10% polyacrylamide gel and stained with SimplyBlue Safe Stain (Invitrogen). **(B)** SDS-PAGE analysis of purified Cas9 orthologs. Cas9 orthologs were purified as described in Supplementary Materials and Methods. 2.5 μg of each purified Cas9 were analyzed on a 4-20% gradient polyacrylamide gel and stained with SimplyBlue Safe Stain.

[00537] Figure 17 (also see Figure 10). The protospacer 1 sequence originates from *S. pyogenes* SF370 (M1) SPy_0700, target of *S. pyogenes* SF370 crRNAsp1 (*4*). Here, the protospacer 1 sequence was manipulated by changing the PAM from a nonfunctional sequence (TTG) to a functional one (TGG). The protospacer 4 sequence originates from *S. pyogenes* MGAS10750 (M4) MGAS10750_Spy1285, target of *S. pyogenes* SF370 crRNA-sp4 (*4*). **(A)** Protospacer 1 plasmid DNA cleavage guided by cognate tracrRNA:crRNA duplexes. The cleavage products were resolved by agarose gel electrophoresis and visualized by ethidium bromide staining. M, DNA marker; fragment sizes in base pairs are indicated. **(B)** Protospacer 1 oligonucleotide DNA cleavage guided by cognate tracrRNA:crRNA-sp1 duplex. The cleavage products were resolved by denaturating polyacrylamide gel electrophoresis and visualized by phosphorimaging. Fragment sizes in nucleotides are indicated. **(C)** Protospacer 4 oligonucleotide DNA cleavage guided by cognate tracrRNA:crRNA-sp4 duplex. The cleavage products were resolved by denaturating polyacrylamide gel electrophoresis and visualized by phosphorimaging. Fragment

sizes in nucleotides are indicated. **(A, B, C)** Experiments in (A) were performed as in Figure 10A; in (B) and in (C) as in Figure 10B. **(B, C)** A schematic of the tracrRNA:crRNAtarget DNA interaction is shown below. The regions of crRNA complementarity to tracrRNA and the protospacer DNA are overlined and underlined, respectively. The PAM sequence is labeled.

[00538] Figure 18 (also see Figure 10). **(A)** Protospacer 2 plasmid DNA was incubated with Cas9 complexed with tracrRNA:crRNA-sp2 in the presence of different concentrations of Mg^{2+}, Mn^{2+}, Ca^{2+}, Zn^{2+}, Co^{2+}, Ni^{2+} or Cu^{2+}. The cleavage products were resolved by agarose gel electrophoresis and visualized by ethidium bromide staining. Plasmid forms are indicated. **(B)** A protospacer 4 oligonucleotide DNA duplex containing a PAM motif was annealed and gel-purified prior to radiolabeling at both 5' ends. The duplex (10 nM final concentration) was incubated with Cas9 programmed with tracrRNA (nucleotides 23-89) and crRNAsp4 (500 nM final concentration, 1:1). At indicated time points (min), 10 µl aliquots of the cleavage reaction were quenched with formamide buffer containing 0.025% SDS and 5 mM EDTA, and analyzed by denaturing polyacrylamide gel electrophoresis as in Figure 10B. Sizes in nucleotides are indicated.

[00539] Figure 19 (A) Mapping of protospacer 1 plasmid DNA cleavage. Cleavage products from Figure 17A were analyzed by sequencing as in Figure 10C. Note that the 3' terminal A overhang (asterisk) is an artifact of the sequencing reaction. **(B)** Mapping of protospacer 4 oligonucleotide DNA cleavage. Cleavage products from Figure 17C were analyzed by denaturing polyacrylamide gel electrophoresis alongside 5' endlabeled oligonucleotide size markers derived from the complementary and noncomplementary strands of the protospacer 4 duplex DNA. M, marker; P, cleavage product. Lanes 1-2: complementary strand. Lanes 3-4: non-complementary strand. Fragment sizes in nucleotides are indicated. **(C)** Schematic representations of tracrRNA, crRNA-sp1 and protospacer 1 DNA sequences (top) and tracrRNA, crRNAsp4 and protospacer 4 DNA sequences (bottom). tracrRNA:crRNA forms a dual-RNA structure directed to complementary protospacer DNA through crRNA-protospacer DNA pairing. The regions of crRNA complementary to tracrRNA and the protospacer DNA are overlined and underlined, respectively. The cleavage sites in the complementary and noncomplementary DNA strands mapped in (A) (top) and (B) (bottom) are represented with arrows (A and B, complementary strand) and a black bar (B, noncomplementary strand) above the sequences, respectively.

[00540] Figure 20 (A) Single turnover kinetics of Cas9 under different RNA pre-annealing and protein-RNA pre-incubation conditions. Protospacer 2 plasmid DNA was incubated with either Cas9 pre-incubated with pre-annealed tracrRNA:crRNA-sp2 (○), Cas9 not pre-incubated with pre-annealed tracrRNA:crRNA-sp2 (●), Cas9 pre-incubated with not pre-annealed tracrRNA and crRNA-sp2 (□) or Cas9 not pre-incubated with not pre-annealed RNAs (■). The cleavage

activity was monitored in a time-dependent manner and analyzed by agarose gel electrophoresis followed by ethidium bromide staining. The average percentage of cleavage from three independent experiments is plotted against the time (min) and fitted with a nonlinear regression. The calculated cleavage rates (k_{obs}) are shown in the table. The results suggest that the binding of Cas9 to the RNAs is not rate-limiting under the conditions tested. Plasmid forms are indicated. The obtained k_{obs} values are comparable to those of restriction endonucleases which are typically of the order of 1-10 per min (*45-47*). **(B)** Cas9 is a multiple turnover endonuclease. Cas9 loaded with duplexed tracrRNA:crRNA-sp2 (1 nM, 1:1:1 – indicated with gray line on the graph) was incubated with a 5-fold excess of native protospacer 2 plasmid DNA. Cleavage was monitored by withdrawing samples from the reaction at defined time intervals (0 to 120 min) followed by agarose gel electrophoresis analysis (top) and determination of cleavage product amount (nM) (bottom). Standard deviations of three independent experiments are indicated. In the time interval investigated, 1 nM Cas9 was able to cleave ~2.5 nM plasmid DNA.

Each Cas9 nuclease domain cleaves one DNA strand

[00541] Cas9 contains domains homologous to both HNH and RuvC endonucleases (Figure 11A and Figure 3) (21–23, 27, 28). We designed and purified Cas9 variants containing inactivating point mutations in the catalytic residues of either the HNH or RuvC-like domains (Figure 11A and Figure 3) (23, 27). Incubation of these variant Cas9 proteins with native plasmid DNA showed that dual-RNA–guided mutant Cas9 proteins yielded nicked open circular plasmids, whereas the WT Cas9 protein-tracrRNA:crRNA complex produced a linear DNA product (Figures 10A and 11A and figures 17A and 25A). This result indicates that the Cas9 HNH and RuvC-like domains each cleave one plasmid DNA strand. To determine which strand of the target DNA is cleaved by each Cas9 catalytic domain, we incubated the mutant Cas9-tracrRNA:crRNA complexes with short dsDNA substrates in which either the complementary or noncomplementary strand was radiolabeled at its 5′ end. The resulting cleavage products indicated that the Cas9 HNH domain cleaves the complementary DNA strand, whereas the Cas9 RuvC-like domain cleaves the noncomplementaryDNAstrand (Figure 11B and Figure 21B).

[00542] Figure 11(A) (Top) Schematic representation of Cas9 domain structure showing the positions of domain mutations. D10A, Asp10→Ala10; H840A; His840→Ala840. Complexes of WT or nuclease mutant Cas9 proteins with tracrRNA: crRNA-sp2 were assayed for endonuclease activity as in Figure 10A. **(B)** Complexes of WT Cas9 or nuclease domain mutants with tracrRNA and crRNA-sp2 were tested for activity as in Figure 10B.

[00543] Figure 3 The amino-acid sequence of Cas9 from *S. pyogenes* (SEQ ID NO:8) is represented. Cas9/Csn1 proteins from various diverse species have 2 domains that include motifs homologous to both HNH and RuvC endonucleases. (**A**) Motifs 1-4 (motif numbers are marked on left side of

sequence) are shown for *S. pyogenes* Cas9/Csn1. The three predicted RuvC-like motifs (1, 2, 4) and the predicted HNH motif (3) are overlined. Residues Asp10 and His840, which were substituted by Ala in this study are highlighted by an asterisk above the sequence. Underlined residues are highly conserved among Cas9 proteins from different species. Mutations in underlined residues are likely to have functional consequences on Cas9 activity. Note that in the present study coupling of the two nuclease-like activities is experimentally demonstrated (Figure 11 and Figure 21). (**B**) Domains 1 (amino acids 7-166) and 2 (amino acids 731-1003), which include motifs 1-4, are depicted for *S. pyogenes* Cas9/Csn1. Refer to **Table 1**and **Figure 5** for additional information.

[00544] Figure 21 Protospacer DNA cleavage by cognate tracrRNA:crRNA-directed Cas9 mutants containing mutations in the HNH or RuvC-like domain. **(A)** Protospacer 1 plasmid DNA cleavage. The experiment was performed as in Figure 11A. Plasmid DNA conformations and sizes in base pairs are indicated. **(B)** Protospacer 4 oligonucleotide DNA cleavage. The experiment was performed as in Figure 11B. Sizes in nucleotides are indicated.

Dual-RNA requirements for target DNA binding and cleavage

[00545] tracrRNA might be required for targetDNAbinding and/or to stimulate the nuclease activity of Cas9 downstream of target recognition. To distinguish between these possibilities, we used an electrophoretic mobility shift assay to monitor target DNA binding by catalytically inactive Cas9 in the presence or absence of crRNA and/or tracrRNA. Addition of tracrRNA substantially enhanced target DNA binding by Cas9, whereas we observed little specific DNA binding with Cas9 alone or Cas9- crRNA (Figure 22). This indicates that tracrRNA is required for target DNA recognition, possibly by properly orienting the crRNA for interaction with the complementary strand of target DNA. The predicted tracrRNA:crRNA secondary structure includes base pairing between the 22 nucleotides at the 3′ terminus of the crRNA and a segment near the 5′ end of the mature tracrRNA (Figure 10E). This interaction creates a structure in which the 5′-terminal 20 nucleotides of the crRNA, which vary in sequence in different crRNAs, are available for target DNA binding. The bulk of the tracrRNA downstream of the crRNA basepairing region is free to form additional RNA structure(s) and/or to interact with Cas9 or the target DNA site. To determine whether the entire length of the tracrRNA is necessary for sitespecific Cas9-catalyzed DNA cleavage, we tested Cas9-tracrRNA:crRNA complexes reconstituted using full-length mature (42-nucleotide) crRNA and various truncated forms of tracrRNA lacking sequences at their 5′ or 3′ ends. These complexes were tested for cleavage using a short target dsDNA. A substantially truncated version of the tracrRNA retaining nucleotides 23 to 48 of the native sequence was capable of supporting robust dual-RNA–guided Cas9-catalyzed DNA cleavage (Figure 12, A and C, and Figure 23, A and B). Truncation of the crRNA from either end showed

that Cas9-catalyzed cleavage in the presence of tracrRNA could be triggered with crRNAs missing the 3′-terminal 10 nucleotides (Figure 12, B and C). In contrast, a 10-nucleotide deletion from the 5′ end of crRNA abolished DNA cleavage by Cas9 (Figure 12B). We also analyzed Cas9 orthologs from various bacterial species for their ability to support *S. pyogenes* tracrRNA:crRNA-guided DNA cleavage. In contrast to closely related *S. pyogenes* Cas9 orthologs, more distantly related orthologs were not functional in the cleavage reaction (Figure 24). Similarly, S. pyogenes Cas9 guided by tracrRNA:crRNA duplexes originating from more distant systems was unable to cleave DNA efficiently (Figure 24). Species specificity of dual-RNA–guided cleavage of DNA indicates coevolution of Cas9, tracrRNA, and the crRNA repeat, as well as the existence of a still unknown structure and/or sequence in the dual- RNA that is critical for the formation of the ternary complex with specific Cas9 orthologs.

[00546] To investigate the protospacer sequence requirements for type II CRISPR/Cas immunity in bacterial cells, we analyzed a series of protospacer-containing plasmid DNAs harboring single-nucleotide mutations for their maintenance following transformation in S. pyogenes and their ability to be cleaved by Cas9 in vitro. In contrast to point mutations introduced at the 5′ end of the protospacer, mutations in the region close to the PAM and the Cas9 cleavage sites were not tolerated in vivo and resulted in decreased plasmid cleavage efficiency in vitro (Figure 12D). Our results are in agreement with a previous report of protospacer escape mutants selected in the type II CRISPR system from S. thermophilus in vivo (27, 29). Furthermore, the plasmid maintenance and cleavage results hint at the existence of a "seed" region located at the 3′ end of the protospacer sequence that is crucial for the interaction with crRNA and subsequent cleavage by Cas9. In support of this notion, Cas9 enhanced complementary DNA strand hybridization to the crRNA; this enhancement was the strongest in the 3′-terminal region of the crRNA targeting sequence (Figure 25A-C). Corroborating this finding, a contiguous stretch of at least 13 base pairs between the crRNA and the target DNA site proximal to the PAM is required for efficient target cleavage, whereas up to six contiguous mismatches in the 5′-terminal region of the protospacer are tolerated (Figure 12E). These findings are reminiscent of the previously observed seed-sequence requirements for target nucleic acid recognition in Argonaute proteins (30, 31) and the Cascade and Csy CRISPR complexes (13, 14).

[00547] Figure 12 (A) Cas9-tracrRNA: crRNA complexes were reconstituted using 42-nucleotide crRNA-sp2 and truncated tracrRNA constructs and were assayed for cleavage activity as in Figure 10B. **(B)** Cas9 programmed with full-length tracrRNA and crRNA-sp2 truncations was assayed for activity as in (A). **(C)** Minimal regions of tracrRNA and crRNA capable of guiding Cas9-mediated DNA cleavage (shaded region). **(D)** Plasmids containing WT or mutant protospacer 2 sequences with indicated point mutations were cleaved in vitro by programmed

Cas9 as in Figure 10A and used for transformation assays of WT or pre-crRNA–deficient S. pyogenes. The transformation efficiency was calculated as colony-forming units (CFU) per microgram of plasmid DNA. Error bars represent SDs for three biological replicates. **(E)** Plasmids containing WT and mutant protospacer 2 inserts with varying extent of crRNA-target DNA mismatches (bottom) were cleaved in vitro by programmed Cas9 (top). The cleavage reactions were further digested with XmnI. The 1880- and 800-bp fragments are Cas9-generated cleavage products. M, DNA marker.

[00548] Figure 22 Electrophoretic mobility shift assays were performed using protospacer 4 target DNA duplex and Cas9 (containing nuclease domain inactivating mutations D10A and H840) alone or in the presence of crRNA-sp4, tracrRNA (75nt), or both. The target DNA duplex was radiolabeled at both 5' ends. Cas9 (D10/H840A) and complexes were titrated from 1 nM to 1 μM. Binding was analyzed by 8% native polyacrylamide gel electrophoresis and visualized by phosphorimaging. Note that Cas9 alone binds target DNA with moderate affinity. This binding is unaffected by the addition of crRNA, suggesting that this represents sequence nonspecific interaction with the dsDNA. Furthermore, this interaction can be outcompeted by tracrRNA alone in the absence of crRNA. In the presence of both crRNA and tracrRNA, target DNA binding is substantially enhanced and yields a species with distinct electrophoretic mobility, indicative of specific target DNA recognition.

[00549] Figure 23 A fragment of tracrRNA encompassing a part of the crRNA paired region and a portion of the downstream region is sufficient to direct cleavage of protospacer oligonucleotide DNA by Cas9. **(A)** Protospacer 1 oligonucleotide DNA cleavage and **(B)** Protospacer 4 oligonucleotide DNA cleavage by Cas9 guided with a mature cognate crRNA and various tracrRNA fragments. **(A, B)** Sizes in nucleotides are indicated.

[00550] Figure 24 Like Cas9 from *S. pyogenes*, the closely related Cas9 orthologs from the Gram-positive bacteria *L. innocua* and *S. thermophilus* cleave protospacer DNA when targeted by tracrRNA:crRNA from *S. pyogenes*. However, under the same conditions, DNA cleavage by the less closely related Cas9 orthologs from the Gramnegative bacteria *C. jejuni* and *N. meningitidis* is not observed. Spy, *S. pyogenes* SF370 (Accession Number NC_002737); Sth, *S. thermophilus* LMD-9 (STER_1477 Cas9 ortholog; Accession Number NC_008532); Lin, *L. innocua* Clip11262 (Accession Number NC_003212); Cje, *C. jejuni* NCTC 11168 (Accession Number NC_002163); Nme, *N. meningitidis* A Z2491 (Accession Number NC_003116). **(A)** Cleavage of protospacer plasmid DNA. Protospacer 2 plasmid DNA (300 ng) was subjected to cleavage by different Cas9 orthologs (500 nM) guided by hybrid tracrRNA:crRNA-sp2 duplexes (500 nM, 1:1) from different species. To design the RNA duplexes, we predicted tracrRNA sequences from *L. innocua* and *N. meningitidis* based on previously published Northern blot data *(4)*. The

dual-hybrid RNA duplexes consist of speciesspecific tracrRNA and a heterologous crRNA. The heterologous crRNA sequence was engineered to contain *S. pyogenes* DNA-targeting sp2 sequence at the 5' end fused to *L. innocua* or *N. meningitidis* tracrRNA-binding repeat sequence at the 3' end. Cas9 orthologs from *S. thermophilus* and *L. innocua,* but not from *N. meningitidis* or *C. jejuni*, can be guided by *S. pyogenes* tracrRNA:crRNA-sp2 to cleave protospacer 2 plasmid DNA, albeit with slightly decreased efficiency. Similarly, the hybrid *L. innocua* tracrRNA:crRNA-sp2 can guide *S. pyogenes* Cas9 to cleave the target DNA with high efficiency, whereas the hybrid *N. meningitidis* tracrRNA:crRNA-sp2 triggers only slight DNA cleavage activity by *S. pyogenes* Cas9. As controls, *N. meningitidis* and *L. innocua* Cas9 orthologs cleave protospacer 2 plasmid DNA when guided by the cognate hybrid tracrRNA:crRNA-sp2. Note that as mentioned above, the tracrRNA sequence of *N. meningitidis* is predicted only and has not yet been confirmed by RNA sequencing. Therefore, the low efficiency of cleavage could be the result of either low activity of the Cas9 orthologs or the use of a nonoptimally designed tracrRNA sequence. **(B)** Cleavage of protospacer oligonucleotide DNA. 5'-end radioactively labeled complementary strand oligonucleotide (10 nM) pre-annealed with unlabeled noncomplementary strand oligonucleotide (protospacer 1) (10 nM) (left) or 5'-end radioactively labeled noncomplementary strand oligonucleotide (10 nM) pre-annealed with unlabeled complementary strand oligonucleotide (10 nM) (right) (protospacer 1) was subjected to cleavage by various Cas9 orthologs (500 nM) guided by tracrRNA:crRNA-sp1 duplex from *S. pyogenes* (500 nM, 1:1). Cas9 orthologs from *S. thermophilus* and *L. innocua*, but not from *N. meningitidis* or *C. jejuni* can be guided by *S. pyogenes* cognate dual-RNA to cleave the protospacer oligonucleotide DNA, albeit with decreased efficiency. Note that the cleavage site on the complementary DNA strand is identical for all three orthologs. Cleavage of the noncomplementary strand occurs at distinct positions. **(C)** Amino acid sequence identity of Cas9 orthologs. *S. pyogenes*, *S. thermophilus* and *L. innocua* Cas9 orthologs share high percentage of amino acid identity. In contrast, the *C. jejuni* and *N. meningitidis* Cas9 proteins differ in sequence and length (~300-400 amino acids shorter). **(D)** Co-foldings of engineered species-specific heterologous crRNA sequences with the corresponding tracrRNA orthologs from *S. pyogenes* (experimentally confirmed, (*4*)), *L. innocua* (predicted) or *N. meningitidis* (predicted). tracrRNAs; crRNA spacer 2 fragments; and crRNA repeat fragments are traced and labeled. *L. innocua* and *S. pyogenes* hybrid tracrRNA:crRNA-sp2 duplexes share very similar structural characteristics, albeit distinct from the *N. meningitidis* hybrid tracrRNA:crRNA. Together with the cleavage data described above in (A) and (B), the co-folding predictions would indicate that the species-specificity cleavage of target DNA by Cas9-tracrRNA:crRNA is dictated by a still unknown structural feature in the tracrRNA:crRNA duplex that is recognized specifically by a

cognate Cas9 ortholog. It was predicted that the species-specificity of cleavage observed in (A) and (B) occurs at the level of binding of Cas9 to dual-tracrRNA:crRNA. Dual-RNA guided Cas9 cleavage of target DNA can be species specific. Depending on the degree of diversity/evolution among Cas9 proteins and tracrRNA:crRNA duplexes, Cas9 and dual- RNA orthologs are partially interchangeable.

[00551] Figure 25 A series of 8-nucleotide DNA probes complementary to regions in the crRNA encompassing the DNA-targeting region and tracrRNA-binding region were analyzed for their ability to hybridize to the crRNA in the context of a tracrRNA:crRNA duplex and the Cas9-tracrRNA:crRNA ternary complex. **(A)** Schematic representation of the sequences of DNA probes used in the assay and their binding sites in crRNA-sp4. **(B-C)** Electrophoretic mobility shift assays of target DNA probes with tracrRNA:crRNA-sp4 or Cas9-tracrRNA:crRNA-sp4. The tracrRNA(15-89) construct was used in the experiment. Binding of the duplexes or complexes to target oligonucleotide DNAs was analyzed on a 16% native polyacrylamide gel and visualized by phosphorimaging.

A short sequence motif dictates R-loop formation

[00552] In multiple CRISPR/Cas systems, recognition of self versus nonself has been shown to involve a short sequence motif that is preserved in the foreign genome, referred to as the PAM(27, 29, 32–34). PAMmotifs are only a few base pairs in length, and their precise sequence and position vary according to the CRISPR/Cas system type (32). In the S. pyogenes type II system, the PAM conforms to an NGG consensus sequence, containing two G:C base pairs that occur one base pair downstream of the crRNA binding sequence, within the target DNA (4). Transformation assays demonstrated that the GG motif is essential for protospacer plasmid DNA elimination by CRISPR/Cas in bacterial cells (Figure 26A), consistent with previous observations in S. thermophilus (27). The motif is also essential for in vitro protospacer plasmid cleavage by tracrRNA:crRNA-guided Cas9 (Figure 26B). To determine the role of the PAM in target DNA cleavage by the Cas9-tracrRNA: crRNA complex, we tested a series of dsDNA duplexes containing mutations in the PAM sequence on the complementary or noncomplementary strands, or both (Figure 13A). Cleavage assays using these substrates showed that Cas9- catalyzed DNA cleavage was particularly sensitive to mutations in the PAM sequence on the noncomplementary strand of the DNA, in contrast to complementary strand PAM recognition by type I CRISPR/Cas systems (18, 34). Cleavage of target single-stranded DNAs was unaffected by mutations of the PAM motif. This observation suggests that the PAM motif is required only in the context of target dsDNA and may thus be required to license duplex unwinding, strand invasion, and the formation of an R-loop structure. When we used a different crRNA-target DNA pair (crRNA-sp4 and protospacer 4 DNA), selected due to the presence of a canonical PAM not present in the

protospacer 2 target DNA, we found that both G nucleotides of the PAM were required for efficient Cas9-catalyzed DNA cleavage (Figure 13B and Figure 26C). To determine whether the PAM plays a direct role in recruiting the Cas9-tracrRNA:crRNA complex to the correct target DNA site, we analyzed binding affinities of the complex for target DNA sequences by native gel mobility shift assays (Figure 13C). Mutation of either G in the PAM sequence substantially reduced the affinity of Cas9-tracrRNA: crRNA for the target DNA. This finding illustrates a role for the PAM sequence in target DNA binding by Cas9.

[00553] Figure 13 (A) Dual RNA-programmed Cas9 was tested for activity as in Figure 10B. WT and mutant PAM sequences in target DNAs are indicated with lines. **(B)** Protospacer 4 target DNA duplexes (labeled at both 5′ ends) containing WT and mutant PAM motifs were incubated with Cas9 programmed with tracrRNA:crRNA-sp4 (nucleotides 23 to 89). At the indicated time points (in minutes), aliquots of the cleavage reaction were taken and analyzed as in Figure 10B. **(C)** Electrophoretic mobility shift assays were performed using RNA-programmed Cas9 (D10A/H840A) and protospacer 4 target DNA duplexes [same as in (B)] containing WT and mutated PAM motifs. The Cas9 (D10A/H840A)–RNA complex was titrated from 100 pM to 1 mM.

[00554] Figure 26 (A) Mutations of the PAM sequence in protospacer 2 plasmid DNA abolish interference of plasmid maintenance by the Type II CRISPR/Cas system in bacterial cells. Wild-type protospacer 2 plasmids with a functional or mutated PAM were transformed into wild-type (strain SF370, also named EC904) and pre-crRNA-deficient mutant (EC1479) *S. pyogenes* as in Figure 12D. PAM mutations are not tolerated by the Type II CRISPR/Cas system in vivo. The mean values and standard deviations of three biological replicates are shown. **(B)** Mutations of the PAM sequence in protospacer plasmid DNA abolishes cleavage by Cas9-tracrRNA:crRNA. Wild type protospacer 2 plasmid with a functional or mutated PAM were subjected to Cas9 cleavage as in Figure 10A. The PAM mutant plasmids are not cleaved by the Cas9-tracrRNA:crRNA complex. **(C)** Mutations of the canonical PAM sequence abolish interference of plasmid maintenance by the Type II CRISPR/Cas system in bacterial cells. Wild-type protospacer 4 plasmids with a functional or mutated PAM were cleaved with Cas9 programmed with tracrRNA and crRNA-sp2. The cleavage reactions were carried out in the presence of the XmnI restriction endonuclease to visualize the Cas9 cleavage products as two fragments (~1880 and ~800 bp). Fragment sizes in base pairs are indicated.

Cas9 can be programmed with a single chimeric RNA

[00555] Examination of the likely secondary structure of the tracrRNA:crRNA duplex (Figures 10E and 12C) suggested the possibility that the features required for site-specific Cas9-catalyzed DNA cleavage could be captured in a single chimeric RNA. Although the tracrRNA:crRNA target-

selection mechanism works efficiently in nature, the possibility of a single RNA-guided Cas9 is appealing due to its potential utility for programmed DNA cleavage and genome editing (Figure 1A-B). We designed two versions of a chimeric RNA containing a target recognition sequence at the 5′ end followed by a hairpin structure retaining the base-pairing interactions that occur between the tracrRNA and the crRNA (Figure 14A). This single transcript effectively fuses the 3′ end of crRNA to the 5′ end of tracrRNA, thereby mimicking the dual-RNA structure required to guide site-specific DNA cleavage by Cas9. In cleavage assays using plasmid DNA, we observed that the longer chimeric RNAwas able to guide Cas9-catalyzed DNA cleavage in a manner similar to that observed for the truncated tracrRNA:crRNA duplex (Figure 14A and Figure 27, A and C). The shorter chimeric RNA did not work efficiently in this assay, confirming that nucleotides that are 5 to 12 positions beyond the tracrRNA:crRNA base-pairing interaction are important for efficient Cas9 binding and/or target recognition. We obtained similar results in cleavage assays using short dsDNA as a substrate, further indicating that the position of the cleavage site in target DNA is identical to that observed using the dual tracrRNA:crRNA as a guide (Figure 14B and Figure 27, B and C). Finally, to establish whether the design of chimeric RNA might be universally applicable, we engineered five different chimeric guide RNAs to target a portion of the gene encoding the green-fluorescent protein (GFP) (Figure 28, A to C) and tested their efficacy against a plasmid carrying the GFP coding sequence in vitro. In all five cases, Cas9 programmed with these chimeric RNAs efficiently cleaved the plasmid at the correct target site (Figure 14C and Figure 28D), indicating that rational design of chimeric RNAs is robust and could, in principle, enable targeting of any DNA sequence of interest with few constraints beyond the presence of a GG dinucleotide adjacent to the targeted sequence.

[00556] Figure 1 A DNA-targeting RNA comprises a single stranded "DNA-targeting segment" and a "protein-binding segment," which comprises a stretch of double stranded RNA. (A) A DNA-targeting RNA can comprise two separate RNA molecules (referred to as a "double-molecule" or "two-molecule" DNA-targeting RNA). A double-molecule DNA-targeting RNA comprises a "targeter-RNA" and an "activator-RNA." (B) A DNA-targeting RNA can comprise a single RNA molecule (referred to as a "single-molecule" DNA-targeting RNA). A single-molecule DNA-targeting RNA comprises "linker nucleotides."

[00557] Figure 14 (A) A plasmid harboring protospacer 4 target sequence and a WT PAM was subjected to cleavage by Cas9 programmed with tracrRNA(4-89):crRNA-sp4 duplex or in vitro–transcribed chimeric RNAs constructed by joining the 3′ end of crRNA to the 5′ end of tracrRNA with a GAAA tetraloop. Cleavage reactions were analyzed by restriction mapping with XmnI. Sequences of chimeric RNAs A and B are shown with DNA-targeting (underline), crRNA

repeat-derived sequences (overlined), and tracrRNA-derived (dashed underlined) sequences. **(B)** Protospacer 4 DNA duplex cleavage reactions were performed as in Figure 10B. **(C)** Five chimeric RNAs designed to target the GFP gene were used to program Cas9 to cleave a GFP gene–containing plasmid. Plasmid cleavage reactions were performed as in Figure 12E, except that the plasmid DNA was restriction mapped with AvrII after Cas9 cleavage.

[00558] Figure 27 (A) A single chimeric RNA guides Cas9-catalyzed cleavage of cognate protospacer plasmid DNA (protospacer 1 and protospacer 2). The cleavage reactions were carried out in the presence of the XmnI restriction endonuclease to visualize the Cas9 cleavage products as two fragments (~1880 and ~800 bp). Fragment sizes in base pairs are indicated. **(B)** A single chimeric RNA guides Cas9-catalyzed cleavage of cognate protospacer oligonucleotide DNA (protospacer 1 and protospacer 2). Fragment sizes in nucleotides are indicated. **(C)** Schematic representations of the chimeric RNAs used in the experiment. Sequences of chimeric RNAs A and B are shown with the 5' protospacer DNA-targeting sequence of crRNA (underlined), the tracrRNA-binding sequence of crRNA (overlined) and tracrRNA-derived sequence (dashed underlined).

[00559] Figure 28 (A) Schematic representation of the GFP expression plasmid pCFJ127. The targeted portion of the GFP open reading frame is indicated with a black arrowhead. **(B)** Close-up of the sequence of the targeted region. Sequences targeted by the chimeric RNAs are shown with gray bars. PAM dinucleotides are boxed. A unique SalI restriction site is located 60 bp upstream of the target locus. **(C)** Left: Target DNA sequences are shown together with their adjacent PAM motifs. Right: Sequences of the chimeric guide RNAs. **(D)** pCFJ127 was cleaved by Cas9 programmed with chimeric RNAs GFP1-5, as indicated. The plasmid was additionally digested with SalI and the reactions were analyzed by electrophoresis on a 3% agarose gel and visualized by staining with SYBR Safe.

Conclusions

[00560] A DNA interference mechanism was identified, involving a dual-RNA structure that directs a Cas9 endonuclease to introduce site-specific double-stranded breaks in target DNA. The tracrRNA:crRNA-guided Cas9 protein makes use of distinct endonuclease domains (HNH and RuvC-like domains) to cleave the two strands in the target DNA. Target recognition by Cas9 requires both a seed sequence in the crRNA and a GG dinucleotide-containing PAM sequence adjacent to the crRNA-binding region in the DNA target.We further show that the Cas9 endonuclease can be programmed with guide RNA engineered as a single transcript to target and cleave any dsDNA sequence of interest. The system is efficient, versatile, and programmable by changing the DNA target-binding sequence in the guide chimeric RNA. Zinc-finger nucleases and transcription-activator–like effector nucleases have attracted considerable interest as

artificial enzymes engineered to manipulate genomes (35–38). This represents alternative methodology based on RNA-programmed Cas9 that facilitates gene-targeting and genome-editing applications.

References Cited

[00561] 1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, Nature 482, 331 (2012).
2. D. Bhaya, M. Davison, R. Barrangou, Annu. Rev. Genet. 45, 273 (2011).
3. M. P. Terns, R. M. Terns, Curr. Opin. Microbiol. 14, 321 (2011).
4. E. Deltcheva et al., Nature 471, 602 (2011).
5. J. Carte, R. Wang, H. Li, R. M. Terns, M. P. Terns, Genes Dev. 22, 3489 (2008).
6. R. E. Haurwitz, M. Jinek, B. Wiedenheft, K. Zhou, J. A. Doudna, Science 329, 1355 (2010).
7. R. Wang, G. Preamplume, M. P. Terns, R. M. Terns, H. Li, Structure 19, 257 (2011).
8. E. M. Gesner, M. J. Schellenberg, E. L. Garside, M. M. George, A. M. Macmillan, Nat. Struct. Mol. Biol. 18, 688 (2011).
9. A. Hatoum-Aslan, I. Maniv, L. A. Marraffini, Proc. Natl. Acad. Sci. U.S.A. 108, 21218 (2011).
10. S. J. J. Brouns et al., Science 321, 960 (2008).
11. D. G. Sashital, M. Jinek, J. A. Doudna, Nat. Struct. Mol. Biol. 18, 680 (2011).
12. N. G. Lintner et al., J. Biol. Chem. 286, 21643 (2011).
13. E. Semenova et al., Proc. Natl. Acad. Sci. U.S.A. 108, 10098 (2011).
14. B. Wiedenheft et al., Proc. Natl. Acad. Sci. U.S.A. 108, 10092 (2011).
15. B. Wiedenheft et al., Nature 477, 486 (2011).
16. C. R. Hale et al., Cell 139, 945 (2009).
17. J. A. L. Howard, S. Delmas, I. Ivančić-Baće, E. L. Bolt, Biochem. J. 439, 85 (2011).
18. E. R. Westra et al., Mol. Cell 46, 595 (2012).
19. C. R. Hale et al., Mol. Cell 45, 292 (2012).
20. J. Zhang et al., Mol. Cell 45, 303 (2012).
21. K. S. Makarova et al., Nat. Rev. Microbiol. 9, 467 (2011).
22. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, Biol. Direct 1, 7 (2006).
23. K. S. Makarova, L. Aravind, Y. I. Wolf, E. V. Koonin, Biol. Direct 6, 38 (2011).
24. S. Gottesman, Nature 471, 588 (2011).
25. R. Barrangou et al., Science 315, 1709 (2007).
26. J. E. Garneau et al., Nature 468, 67 (2010).
27. R. Sapranauskas et al., Nucleic Acids Res. 39, 9275 (2011).
28. G. K. Taylor, D. F. Heiter, S. Pietrokovski, B. L. Stoddard, Nucleic Acids Res. 39, 9705 (2011).
29. H. Deveau et al., J. Bacteriol. 190, 1390 (2008).
30. B. P. Lewis, C. B. Burge, D. P. Bartel, Cell 120, 15 (2005).
31. G. Hutvagner, M. J. Simard, Nat. Rev. Mol. Cell Biol. 9, 22 (2008).
32. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Microbiology 155, 733 (2009).
33. L. A. Marraffini, E. J. Sontheimer, Nature 463, 568 (2010).
34. D. G. Sashital, B. Wiedenheft, J. A. Doudna, Mol. Cell 46, 606 (2012).

35. M. Christian et al., Genetics 186, 757 (2010).
36. J. C. Miller et al., Nat. Biotechnol. 29, 143 (2011).
37. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, Nat. Rev. Genet. 11, 636 (2010).
38. D. Carroll, Gene Ther. 15, 1463 (2008).
39. J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1989).
40. M. G. Caparon, J. R. Scott, Genetic manipulation of pathogenic streptococci. Methods Enzymol. 204, 556 (1991). doi:10.1016/0076-6879(91)04028-M Medline
41. C. Frøkjær-Jensen et al., Single-copy insertion of transgenes in Caenorhabditis elegans. Nat. Genet. 40, 1375 (2008). doi:10.1038/ng.248 Medline
42. R. B. Denman, Using RNAFOLD to predict the activity of small catalytic RNAs. Biotechniques 15, 1090 (1993). Medline
43. I. L. Hofacker, P. F. Stadler, Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics 22, 1172 (2006). doi:10.1093/bioinformatics/btl023 Medline
44. K. Darty, A. Denise, Y. Ponty, VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics 25, 1974 (2009). doi:10.1093/bioinformatics/btp250 Medline

Example 2: RNA-programmed genome editing in human cells

[00562] Data provided below demonstrate that Cas9 can be expressed and localized to the nucleus of human cells, and that it assembles with single-guide RNA ("sgRNA"; encompassing the features required for both Cas9 binding and DNA target site recognition) in a human cell. These complexes can generate double stranded breaks and stimulate non-homologous end joining (NHEJ) repair in genomic DNA at a site complementary to the sgRNA sequence, an activity that requires both Cas9 and the sgRNA. Extension of the RNA sequence at its 3' end enhances DNA targeting activity in living cells. Further, experiments using extracts from transfected cells show that sgRNA assembly into Cas9 is the limiting factor for Cas9-mediated DNA cleavage. These results demonstrate that RNA-programmed genome editing works in living cells and *in vivo*.

MATERIALS AND METHODS

Plasmid design and construction

[00563] The sequence encoding *Streptococcus pyogenes* Cas9 (residues 1-1368) fused to an HA epitope (amino acid sequence DAYPYDVPDYASL (SEQ ID NO:274)), a nuclear localization signal (amino acid sequence PKKKRKVEDPKKKRKVD (SEQ ID NO:275)) was codon optimized for human expression and synthesized by GeneArt. The DNA sequence is SEQ ID NO:276 and the protein sequence is SEQ ID NO:277. Ligation-independent cloning (LIC) was used to insert this sequence into a pcDNA3.1-derived GFP and mCherry LIC vectors (vectors 6D and 6B, respectively, obtained from the UC Berkeley MacroLab), resulting in a Cas9-HA-NLS-GFP and

Cas9-HA-NLS-mCherry fusions expressed under the control of the CMV promoter. Guide sgRNAs were expressed using expression vector pSilencer 2.1-U6 puro (Life Technologies) and pSuper (Oligoengine). RNA expression constructs were generated by annealing complementary oligonucleotides to form the RNA-coding DNA sequence and ligating the annealed DNA fragment between the BamHI and HindIII sites in pSilencer 2.1-U6 puro and BglII and HindIII sites in pSuper.

Cell culture conditions and DNA transfections

[00564] HEK293T cells were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) in a 37°C humidified incubator with 5% CO_2. Cells were transiently transfected with plasmid DNA using either X-tremeGENE DNA Transfection Reagent (Roche) or Turbofect Transfection Reagent (Thermo Scientific) with recommended protocols. Briefly, HEK293T cells were transfected at 60-80% confluency in 6-well plates using 0.5 μg of the Cas9 expression plasmid and 2.0 μg of the RNA expression plasmid. The transfection efficiencies were estimated to be 30-50% for Tubofect (Figures 29E and 37A-B) and 80-90% for X-tremegene (Figure 31B), based on the fraction of GFP-positive cells observed by fluorescence microscopy. 48 hours post transfection, cells were washed with phosphate buffered saline (PBS) and lysed by applying 250 μl lysis buffer (20 mM Hepes pH 7.5, 100 mM potassium chloride (KCl), 5 mM magnesium chloride ($MgCl_2$), 1 mM dithiothreitol (DTT), 5% glycerol, 0.1% Triton X-100, supplemented with Roche Protease Inhibitor cocktail) and then rocked for 10 min at 4°C. The resulting cell lysate was divided into aliquots for further analysis. Genomic DNA was isolated from 200 μl cell lysate using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol.

Western blot analysis of Cas9 expression

[00565] HEK293T , transfected with the Cas9-HA-NLS-GFP expression plasmid, were harvested and lysed 48 hours post transfection as above. 5 ul of lysate were eletrophoresed on a 10% SDS polyacrylamide gel, blotter onto a PVDF membrane and probed with HRP-conjugated anti-HA antibody (Sigma, 1:1000 dilution in 1x PBS).

Surveyor assay

[00566] The Surveyor assay was performed as previously described [10,12,13]. Briefly, the human clathrin light chain A (CLTA) locus was PCR amplified from 200 ng of genomic DNA using a high fidelity polymerase, Herculase II Fusion DNA Polymerase (Agilent Technologies) and forward primer 5'-GCAGCAGAAGAAGCCTTTGT-3' (SEQ ID NO://) and reverse primer 5'-TTCCTCCTCTCCCTCCTCTC-3' (SEQ ID NO://). 300 ng of the 360 bp amplicon was then denatured by heating to 95°C and slowly reannealed using a heat block to randomly rehybridize wild type and mutant DNA strands. Samples were then incubated with Cel-1 nuclease (Surveyor

Kit, Transgenomic) for 1 hour at 42°C. Cel-1 recognizes and cleaves DNA helices containing mismatches (wild type:mutant hybridization). Cel-1 nuclease digestion products were separated on a 10% acrylamide gel and visualized by staining with SYBR Safe (Life Technologies). Quantification of cleavage bands was performed using ImageLab software (Bio-Rad). The percent cleavage was determined by dividing the average intensity of cleavage products (160-200 bps) by the sum of the intensities of the uncleaved PCR product (360 bp) and the cleavage product.

***In vitro* transcription**

[00567] Guide RNA was *in vitro* transcribed using recombinant T7 RNA polymerase and a DNA template generated by annealing complementary synthetic oligonucleotides as previously described [14]. RNAs were purified by electrophoresis on 7M urea denaturing acrylamide gel, ethanol precipitated, and dissolved in DEPC-treated water.

Northern blot analysis

[00568] RNA was purified from HEK293T cells using the mirVana small-RNA isolation kit (Ambion). For each sample, 800 ng of RNA were separated on a 10% urea-PAGE gel after denaturation for 10 min at 70°C in RNA loading buffer (0.5X TBE (pH7.5), 0.5 mg/ml bromophenol blue, 0.5 mg xylene cyanol and 47% formamide). After electrophoresis at 10W in 0.5X TBE buffer until the bromophenol blue dye reached the bottom of the gel, samples were electroblotted onto a Nytran membrane at 20 volts for 1.5 hours in 0.5X TBE. The transferred RNAs were cross-linked onto the Nytran membrane in UV-Crosslinker (Strategene) and were pre-hybridized at 45°C for 3 hours in a buffer containing 40% formamide, 5X SSC, 3X Dernhardt's (0.1% each of ficoll, polyvinylpyrollidone, and BSA) and 200 µg/ml Salmon sperm DNA. The pre-hybridized membranes were incubated overnight in the prehybridization buffer supplemented with 5'-^{32}P-labeled antisense DNA oligo probe at 1 million cpm/ml. After several washes in SSC buffer (final wash in 0.2X SCC), the membranes were imaged phosphorimaging.

***In vitro* cleavage assay**

[00569] Cell lysates were prepared as described above and incubated with CLTA-RFP donor plasmid [10]. Cleavage reactions were carried out in a total volume of 20 µl and contained 10 µl lysate, 2 µl of 5x cleavage buffer (100 mM HEPES pH 7.5, 500 mM KCl, 25 mM $MgCl_2$, 5 mM DTT, 25% glycerol) and 300 ng plasmid. Where indicated, reactions were supplemented with 10 pmol of *in vitro* transcribed CLTA1 sgRNA. Reactions were incubated at 37°C for one hour and subsequently digested with 10 U of XhoI (NEB) for an additional 30 min at 37°C. The reactions were stopped by the addition of Proteinase K (Thermo Scientific) and incubated at 37°C for 15 min. Cleavage products were analyzed by electrophoresis on a 1% agarose gel and stained with

SYBR Safe. The presence of ~2230 and ~3100 bp fragments is indicative of Cas9-mediated cleavage.

RESULTS

[00570] To test whether Cas9 could be programmed to cleave genomic DNA in living cells, Cas9 was co-expressed together with an sgRNA designed to target the human clathrin light chain (CLTA) gene. The CLTA genomic locus has previously been targeted and edited using ZFNs [10]. We first tested the expression of a human-codon-optimized version of the *Streptococcus pyogenes* Cas9 protein and sgRNA in human HEK293T cells. The 160 kDa Cas9 protein was expressed as a fusion protein bearing an HA epitope, a nuclear localization signal (NLS), and green fluorescent protein (GFP) attached to the C-terminus of Cas9 (Figure 29A). Analysis of cells transfected with a vector encoding the GFP-fused Cas9 revealed abundant Cas9 expression and nuclear localization (Figure 29B). Western blotting confirmed that the Cas9 protein is expressed largely intact in extracts from these cells (Figure 29A). To program Cas9, we expressed sgRNA bearing a 5'-terminal 20-nucleotide sequence complementary to the target DNA sequence, and a 42-nucleotide 3'-terminal stem loop structure required for Cas9 binding (Figure 29C). This 3'-terminal sequence corresponds to the minimal stem-loop structure that has previously been used to program Cas9 *in vitro* [8]. The expression of this sgRNA was driven by the human U6 (RNA polymerase III) promoter [11]. Northern blotting analysis of RNA extracted from cells transfected with the U6 promoter-driven sgRNA plasmid expression vector showed that the sgRNA is indeed expressed, and that their stability is enhanced by the presence of Cas9 (Figure 29D).

[00571] Figure 29 demonstrates that co-expression of Cas9 and guide RNA in human cells generates double-strand DNA breaks at the target locus. **(A)** Top; schematic diagram of the Cas9-HA-NLS-GFP expression construct. Bottom; lysate from HEK293T cells transfected with the Cas9 expression plasmid was analyzed by Western blotting using an anti-HA antibody. **(B)** Fluorescence microscopy of HEK293T cells expressing Cas9-HA-NLS-GFP. **(C)** Design of a single-guide RNA (sgRNA, i.e., a single-molecule DNA-targeting RNA) targeting the human CLTA locus. Top; schematic diagram of the sgRNA target site in exon 7 of the human CLTA gene. The target sequence that hybridizes to the guide segment of CLTA1 sgRNA is indicated by "CLTA1 sgRNA." The GG di-nucleotide protospacer adjacent motif (PAM) is marked by an arrow. Black lines denote the DNA binding regions of the control ZFN protein. The translation stop codon of the CLTA open reading frame is marked with a dotted line for reference. Middle; schematic diagram of the sgRNA expression construct. The RNA is expressed under the control of the U6 Pol III promoter and a poly(T) tract that serves as a Pol III transcriptional terminator signal. Bottom; sgRNA-guided cleavage of target DNA by Cas9. The sgRNA consists of a 20-nt

5'-terminal guide segment followed by a 42-nt stem-loop structure required for Cas9 binding. Cas9-mediated cleavage of the two target DNA strands occurs upon unwinding of the target DNA and formation of a duplex between the guide segment of the sgRNA and the target DNA. This is dependent on the presence of a PAM motif (appropriate for the Cas9 being used, e.g., GG dinucleotide, see Example 1 above) downstream of the target sequence in the target DNA. Note that the target sequence is inverted relative to the upper diagram. **(D)** Northern blot analysis of sgRNA expression in HEK239T cells. **(E)** Surveyor nuclease assay of genomic DNA isolated from HEK293T cells expressing Cas9 and/or CLTA sgRNA. A ZFN construct previously used to target the CLTA locus [10] was used as a positive control for detecting DSB-induced DNA repair by non-homologous end joining.

[00572] Next we investigated whether site-specific DSBs are generated in HEK293T cells transfected with Cas9-HA-NLS-mCherry and the CLTA1 sgRNA. To do this, we probed for minor insertions and deletions in the locus resulting from imperfect repair by DSB-induced NHEJ using the Surveyor nuclease assay [12]. The region of genomic DNA targeted by Cas9:sgRNA is amplified by PCR and the resulting products are denatured and reannealed. The rehybridized PCR products are incubated with the mismatch recognition endonuclease Cel-1 and resolved on an acrylamide gel to identify Cel-1 cleavage bands. As DNA repair by NHEJ is typically induced by a DSB, a positive signal in the Surveyor assay indicates that genomic DNA cleavage has occurred. Using this assay, we detected cleavage of the CLTA locus at a position targeted by the CLTA1 sgRNA (Figure 29E). A pair of ZFNs that target a neighboring site in the CLTA locus provided a positive control in these experiments [10].

[00573] To determine if either Cas9 or sgRNA expression is a limiting factor in the observed genome editing reactions, lysates prepared from the transfected cells were incubated with plasmid DNA harboring a fragment of the CLTA gene targeted by the CLTA1 sgRNA. Plasmid DNA cleavage was not observed upon incubation with lysate prepared from cells transfected with the Cas9-HA-NLS-GFP expression vector alone, consistent with the Surveyor assay results. However, robust plasmid cleavage was detected when the lysate was supplemented with *in vitro* transcribed CLTA1 sgRNA (Figure 30A). Furthermore, lysate prepared from cells transfected with both Cas9 and sgRNA expression vectors supported plasmid cleavage, while lysates from cells transfected with the sgRNA-encoding vector alone did not (Figure 30A). These results suggest that a limiting factor for Cas9 function in human cells could be assembly with the sgRNA. We tested this possibility directly by analyzing plasmid cleavage in lysates from cells transfected as before in the presence and absence of added exogenous sgRNA. Notably, when exogenous sgRNA was added to lysate from cells transfected with both the Cas9 and sgRNA expression vectors, a substantial increase in DNA cleavage activity was observed (Figure 30B). This result

indicates that the limiting factor for Cas9 function in HEK293T cells is the expression of the sgRNA or its loading into Cas9.

[00574] **Figure 30** demonstrates that cell lysates contain active Cas9:sgRNA and support site-specific DNA cleavage. **(A)** Lysates from cells transfected with the plasmid(s) indicated at left were incubated with plasmid DNA containing a PAM and the target sequence complementary to the CLTA1 sgRNA; where indicated, the reaction was supplemented with 10 pmol of in vitro transcribed CLTA1 sgRNA; secondary cleavage with XhoI generated fragments of ~2230 and ~3100 bp fragments indicative of Cas9-mediated cleavage. A control reaction using lysate from cells transfected with a ZFN expression construct shows fragments of slightly different size reflecting the offset of the ZFN target site relative to the CLTA1 target site. **(B)** Lysates from cells transfected with Cas9-GFP expression plasmid and, where indicated, the CLTA1 sgRNA expression plasmid, were incubated with target plasmid DNA as in (A) in the absence or presence of in vitro-transcribed CLTA1 sgRNA.

[00575] As a means of enhancing the Cas9:sgRNA assembly *in living cells*, we next tested the effect of extending the presumed Cas9-binding region of the guide RNA. Two new versions of the CLTA1 sgRNA were designed to include an additional six or twelve base pairs in the helix that mimics the base-pairing interactions between the crRNA and tracrRNA (Figure 31A). Additionally, the 3'-end of the guide RNA was extended by five nucleotides based on the native sequence of the *S. pyogenes* tracrRNA [9]. Vectors encoding these 3' extended sgRNAs under the control of either the U6 or H1 Pol III promoters were transfected into cells along with the Cas9-HA-NLS-GFP expression vector and site-specific genome cleavage was tested using the Surveyor assay (Figure 31B). The results confirmed that cleavage required both Cas9 and the CLTA1 sgRNA, but did not occur when either Cas9 or the sgRNA were expressed alone. Furthermore, we observed substantially increased frequencies of NHEJ, as detected by Cel-1 nuclease cleavage, while the frequency of NHEJ mutagenesis obtained with the control ZFN pair was largely unchanged.

[00576] **Figure 31** demonstrates that 3' extension of sgRNA constructs enhances site-specific NHEJ-mediated mutagenesis. **(A)** The construct for CLTA1 sgRNA expression (top) was designed to generate transcripts containing the original Cas9-binding sequence (v1.0), or dsRNA duplexes extended by 4 base pairs (v2.1) or 10 base pairs (v2.2). **(B)** Surveyor nuclease assay of genomic DNA isolated from HEK293T cells expressing Cas9 and/or CLTA sgRNA v1.0, v2.1 or v2.2. A ZFN construct previously used to target the CLTA locus [10] was used as a positive control for detecting DSB-induced DNA repair by non-homologous end joining.

[00577] The results thus provide the framework for implementing Cas9 as a facile molecular tool for diverse genome editing applications. A powerful feature of this system is the potential to

program Cas9 with multiple sgRNAs in the same cell, either to increase the efficiency of targeting at a single locus, or as a means of targeting several loci simultaneously. Such strategies would find broad application in genome-wide experiments and large-scale research efforts such as the development of multigenic disease models.

Example 3: The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems

[00578] We searched for all putative type II CRISPR-Cas loci currently existing in publicly available bacterial genomes by screening for sequences homologous to Cas9, the hallmark protein of the type II system. We constructed a phylogenetic tree from a multiple sequence alignment of the identified Cas9 orthologues. The CRISPR repeat length and gene organization of cas operons of the associated type II systems were analyzed in the different Cas9 subclusters. A subclassification of type II loci was proposed and further divided into subgroups based on the selection of 75 representative Cas9 orthologues. We then predicted tracrRNA sequences mainly by retrieving CRISPR repeat sequences and screening for anti-repeats within or in the vicinity of the cas genes and CRISPR arrays of selected type II loci. Comparative analysis of sequences and predicted structures of chosen tracrRNA orthologues was performed. Finally, we determined the expression and processing profiles of tracrRNAs and crRNAs from five bacterial species.

MATERIALS AND METHODS

Bacterial strains and culture conditions

[00579] The following media were used to grow bacteria on plates: TSA (trypticase soy agar, Trypticase™ Soy Agar (TSA II) BD BBL, Becton Dickinson) supplemented with 3% sheep blood for S. mutans (UA159), and BHI (brain heart infusion, BD Bacto™ Brain Heart Infusion, Becton Dickinson) agar for L. innocua (Clip11262). When cultivated in liquid cultures, THY medium (Todd Hewitt Broth (THB, Bacto, Becton Dickinson) supplemented with 0.2% yeast extract (Servabacter®) was used for S. mutans, BHI broth for L. innocua, BHI liquid medium containing 1% vitamin-mix VX (Difco, Becton Dickinson) for N. meningitidis (A Z2491), MH (Mueller Hinton Broth, Oxoid) Broth including 1% vitamin-mix VX for C. jejuni (NCTC 11168; ATCC 700819) and TSB (Tryptic Soy Broth, BD BBL™ Trypticase™ Soy Broth) for F. novicida (U112). S. mutans was incubated at 37°C, 5% CO2 without shaking. Strains of L. innocua, N. meningitidis and F. novicida were grown aerobically at 37°C with shaking. C. jejuni was grown at 37°C in microaerophilic conditions using campygen (Oxoid) atmosphere. Bacterial cell growth was followed by measuring the optical density of cultures at 620 nm (OD_{620} nm) at regular time intervals using a microplate reader (BioTek PowerWave™).

Sequencing of bacterial small RNA libraries.

[00580] C. jejuni NCTC 11168 (ATCC 700819), F. novicida U112, L. innocua Clip11262, N. meningitidis A Z2491 and S. mutans UA159 were cultivated until mid-logarithmic growth phase and total RNA was extracted with TRIzol (Sigma-Aldrich). 10 μg of total RNA from each strain were treated with TURBO™ DNase (Ambion) to remove any residual genomic DNA. Ribosomal RNAs were removed by using the Ribo-Zero™ rRNA Removal Kits® for Gram-positive or Gram-negative bacteria (Epicentre) according to the manufacturer's instructions. Following purification with the RNA Clean & Concentrator™-5 kit (Zymo Research), the libraries were prepared using ScriptMiner™ Small RNA-Seq Library Preparation Kit (Multiplex, Illumina® compatible) following the manufacturer's instructions. RNAs were treated with the Tobacco Acid Pyrophosphatase (TAP) (Epicentre). Columns from RNA Clean & Concentrator™-5 (Zymo Research) were used for subsequent RNA purification and the Phusion® High-Fidelity DNA Polymerase (New England Biolabs) was used for PCR amplification. Specific userdefined barcodes were added to each library (RNA-Seq Barcode Primers (Illumina®- compatible) Epicentre) and the samples were sequenced at the Next Generation Sequencing (CSF NGS Unit; on the web at "csf." followed by "ac.at") facilty of the Vienna Biocenter, Vienna, Austria (Illumina single end sequencing).

Analysis of tracrRNA and crRNA sequencing data

[00581] The RNA sequencing reads were split up using the illumina2bam tool and trimmed by (i) removal of Illumina adapter sequences (cutadapt 1.0) and (ii) removal of 15 nt at the 3' end to improve the quality of reads. After removal of reads shorter than 15 nt, the cDNA reads were aligned to their respective genome using Bowtie by allowing 2 mismatches: C. jejuni (GenBank: NC_002163), F. novicida (GenBank: NC_008601), N. meningitidis (GenBank: NC_003116), L. innocua (GenBank: NC_003212) and S. mutans (GenBank: NC_004350). Coverage of the reads was calculated at each nucleotide position separately for both DNA strands using BEDTools-Version-2.15.0. A normalized wiggle file containing coverage in read per million (rpm) was created and visualized using the Integrative Genomics Viewer (IGV) tool ("www." followed by "broadinstitute.org/igv/") (Figure 36). Using SAMTools flagstat[80] the proportion of mapped reads was calculated on a total of mapped 9914184 reads for C. jejuni, 48205 reads for F. novicida, 13110087 reads for N. meningitidis, 161865 reads L. innocua and 1542239 reads for S. mutans. A file containing the number of reads starting (5') and ending (3') at each single nucleotide position was created and visualized in IGV. For each tracrRNA orthologue and crRNA, the total number of reads retrieved was calculated using SAMtools.

Cas9 sequence analysis, multiple sequence alignment and guide tree construction

[00582] Position-Specific Iterated (PSI)-BLAST program was used to retrieve homologues of the Cas9 family in the NCBI non redundant database. Sequences shorter than 800 amino acids were discarded. The BLASTClust program set up with a length coverage cutoff of 0.8 and a score coverage threshold (bit score divided by alignment length) of 0.8 was used to cluster the remaining sequences (Figure 38). This procedure produced 78 clusters (48 of those were represented by one sequence only). One (or rarely a few representatives) were selected from each cluster and multiple alignment for these sequences was constructed using the MUSCLE program with default parameters, followed by a manual correction on the basis of local alignments obtained using PSI-BLAST and HHpred programs. A few more sequences were unalignable and also excluded from the final alignments. The confidently aligned blocks with 272 informative positions were used for maximum likelihood tree reconstruction using the FastTree program with the default parameters: JTT evolutionary model, discrete gamma model with 20 rate categories. The same program was used to calculate the bootstrap values.

[00583] **Figure 38** depicts sequences that were grouped according to the BLASTclust clustering program. Only sequences longer than 800 amino acids were selected for the BLASTclust analysis (see Materials and Methods). Representative strains harboring *cas9* orthologue genes were used. Some sequences did not cluster, but were verified as Cas9 sequences due to the presence of conserved motifs and/or other *cas* genes in their immediate vicinity.

Analysis of CRISPR-Cas loci

[00584] The CRISPR repeat sequences were retrieved from the CRISPRdb database or predicted using the CRISPRFinder tool (Grissa I et al., BMC Bioinformatics 2007; 8:172; Grissa I et al., Nucleic Acids Res 2007). The cas genes were identified using the BLASTp algorithm and/or verified with the KEGG database (on the web at "www." followed by kegg.jp/).

In silico prediction and analysis of tracrRNA orthologues

[00585] The putative antirepeats were identified using the Vector NTI® software (Invitrogen) by screening for additional, degenerated repeat sequences that did not belong to the repeat-spacer array on both strands of the respective genomes allowing up to 15 mismatches. The transcriptional promoters and rho-independent terminators were predicted using the BDGP Neural Network Promoter Prediction program ("www." followed by fruitfly.org/seq_tools/promoter.html) and the TransTermHP software, respectively. The multiple sequence alignments were performed using the MUSCLE program with default parameters. The alignments were analyzed for the presence of conserved structure motifs using the RNAalifold algorithm of the Vienna RNA package 2.0.

RESULTS

Type II CRISPR-Cas systems are widespread in bacteria.

[00586] In addition to the tracrRNA-encoding DNA and the repeat-spacer array, type II CRISPR-Cas loci are typically composed of three to four cas genes organized in an operon (Figure 32A-B). Cas9 is the signature protein characteristic for type II and is involved in the steps of expression and interference. Cas1 and Cas2 are core proteins that are shared by all CRISPR-Cas systems and are implicated in spacer acquisition. Csn2 and Cas4 are present in only a subset of type II systems and were suggested to play a role in adaptation. To retrieve a maximum number of type II CRISPR-Cas loci, containing tracrRNA, we first screened publicly available genomes for sequences homologous to already annotated Cas9 proteins. 235 Cas9 orthologues were identified in 203 bacterial species. A set of 75 diverse sequences representative of all retrieved Cas9 orthologues were selected for further analysis (Figure 32, Figure 38, and Materials and Methods).

[00587] **Figure 32** depicts (A) a phylogenetic tree of representative Cas9 sequences from various organisms as well as (B) representative Cas9 locus architecture. Bootstrap values calculated for each node are indicated. Same color branches represent selected subclusters of similar Cas9 orthologues. CRISPR repeat length in nucleotides, average Cas9 protein size in amino acids (aa) and consensus locus architecture are shown for every subcluster. *- gi|116628213 **- gi|116627542 †- gi|34557790 ‡- gi|34557932. Type II-A is characterized by *cas9- csx12*, *cas1*, *cas2*, *cas4*. Type II-B is characterized by *cas9*, *cas1*, *cas2* followed by a *csn2* variant. Type II-C is characterized by a conserved *cas9*, *cas1*, *cas2* operon (See also Figure 38).

[00588] Next, we performed a multiple sequence alignment of the selected Cas9 orthologues. The comparative analysis revealed high diversities in amino acid composition and protein size. The Cas9 orthologues share only a few identical amino acids and all retrieved sequences have the same domain architecture with a central HNH endonuclease domain and splitted RuvC/RNaseH domain. The lengths of Cas9 proteins range from 984 (Campylobacter jejuni) to 1629 (Francisella novicida) amino acids with typical sizes of ~1100 or ~1400 amino acids. Due to the high diversity of Cas9 sequences, especially in the length of the inter-domain regions, we selected only well-aligned, informative positions of the prepared alignment to reconstruct a phylogenetic tree of the analyzed sequences (Figure 32 and Materials and Methods). Cas9 orthologues grouped into three major, monophyletic clusters with some outlier sequences. The observed topology of the Cas9 tree is well in agreement with the current classification of type II loci, with previously defined type II-A and type II-B forming separate, monophyletic clusters. To further characterize the clusters, we examined in detail the cas operon compositions and CRISPR repeat sequences of all listed strains.

Cas9 subclustering reflects diversity in type II CRISPR-Cas loci architecture

[00589] A deeper analysis of selected type II loci revealed that the clustering of Cas9 orthologue sequences correlates with the diversity in CRISPR repeat length. For most of the type II CRISPR-Cas systems, the repeat length is 36 nucleotides (nt) with some variations for two of the Cas9 tree subclusters. In the type II-A cluster (Figure 32) that comprises loci encoding the long Cas9 orthologue, previously named Csx12, the CRISPR repeats are 37 nt long. The small subcluster composed of sequences from bacteria belonging to the Bacteroidetes phylum (Figure 32) is characterized by unusually long CRISPR repeats, up to 48 nt in size. Furthermore, we noticed that the subclustering of Cas9 sequences correlates with distinct cas operon architectures, as depicted in Figure 32. The third major cluster (Figure 32) and the outlier loci (Figure 32), consist mainly of the minimum operon composed of the cas9, cas1 and cas2 genes, with an exception of some incomplete loci that are discussed later. All other loci of the two first major clusters are associated with a fourth gene, mainly cas4, specific to type II-A or csn2- like, specific to type II-B (Figure 32). We identified genes encoding shorter variants of the Csn2 protein, Csn2a, within loci similar to type II-B S. pyogenes CRISPR01 and S. thermophilus CRISPR3 (Figure 32). The longer variant of Csn2, Csn2b, was found associated with loci similar to type II-B S. thermophilus CRISPR1 (Figure 32). Interestingly, we identified additional putative cas genes encoding proteins with no obvious sequence similarity to previously described Csn2 variants. One of those uncharacterized proteins is exclusively associated with type II-B loci of Mycoplasma species (Figure 32 and Figure 33). Two others were found encoded in type II-B loci of Staphylococcus species (Figure 33). In all cases the cas operon architecture diversity is thus consistent with the subclustering of Cas9 sequences. These characteristics together with the general topology of the Cas9 tree divided into three major, distinct, monophyletic clusters, led us to propose a new, further division of the type II CRISPR-Cas system into three subtypes. Type II-A is associated with Csx12- like Cas9 and Cas4, type II-B is associated with Csn2-like and type II-C only contains the minimal set of the cas9, cas1 and cas2 genes, as depicted in Figure 32.

[00590] **Figure 33** depicts the architecture of type II CRISPR-Cas from selected bacterial species. The vertical bars group the loci that code for Cas9 orthologues belonging to the same tree subcluster (compare with **Figure 32**). Horizontal black bar, leader sequence; black rectangles and diamonds, repeat-spacer array. Predicted anti-repeats are represented by arrows indicating the direction of putative tracrRNA orthologue transcription. Note that for the loci that were not verified experimentally, the CRISPR repeat-spacer array is considered here to be transcribed from the same strand as the *cas* operon. The transcription direction of the putative tracrRNA orthologue is indicated accordingly.

***In silico* predictions of novel tracrRNA orthologues**

[00591] Type II loci selected earlier based on the 75 representative Cas9 orthologues were screened for the presence of putative tracrRNA orthologues. Our previous analysis performed on a restricted number of tracrRNA sequences revealed that neither the sequences of tracrRNAs nor their localization within the CRISPR-Cas loci seemed to be conserved. However, as mentioned above, tracrRNAs are also characterized by an anti-repeat sequence capable of base-pairing with each of the pre-crRNA repeats to form tracrRNA:precrRNA repeat duplexes that are cleaved by RNase III in the presence of Cas9. To predict novel tracrRNAs, we took advantage of this characteristic and used the following workflow: (i) screen for potential anti-repeats (sequence base-pairing with CRISPR repeats) within the CRISPR-Cas loci, (ii) select anti-repeats located in the intergenic regions, (iii) validate CRISPR anti-repeat:repeat base-pairing, and (iv) predict promoters and Rho-independent transcriptional terminators associated to the identified tracrRNAs.

[00592] To screen for putative anti-repeats, we retrieved repeat sequences from the CRISPRdb database or, when the information was not available, we predicted the repeat sequences using the CRISPRfinder software. In our previous study, we showed experimentally that the transcription direction of the repeat-spacer array compared to that of the cas operon varied among loci. Here RNA sequencing analysis confirmed this observation. In some of the analyzed loci, namely in F. novicida, N. meningitidis and C. jejuni, the repeat-spacer array is transcribed in the opposite direction of the cas operon (see paragraph 'Deep RNA sequencing validates expression of novel tracrRNA orthologues' and Figures 33 and 34) while in S. pyogenes, S. mutans, S. thermophilus and L. innocua, the array and the cas operon are transcribed in the same direction. These are the only type II repeat-spacer array expression data available to date. To predict the transcription direction of other repeat-spacer arrays, we considered the previous observation according to which the last repeats of the arrays are usually mutated. This remark is in agreement with the current spacer acquisition model, in which typically the first repeat of the array is duplicated upon insertion of a spacer sequence during the adaptation phase. For 37 repeat spacer arrays, we were able to identify the mutated repeat at the putative end of the arrays. We observed that the predicted orientation of transcription for the N. meningitidis and C. jejuni repeat-spacer array would be opposite to the orientation determined experimentally (RNA sequencing and Northern blot analysis). As the predicted orientation is not consistent within the clusters and as in most of the cases we could detect potential promoters on both ends of the arrays, we considered transcription of the repeat-spacer arrays to be in the same direction as transcription of the cas operon, if not validated otherwise.

[00593] **Figure 34** depicts tracrRNA and pre-crRNA co-processing in selected type II CRISPR Cas systems. CRISPR loci architectures with verified positions and directions of tracrRNA and pre-crRNA transcription are shown. Top sequences, pre-crRNA repeats; bottom sequences, tracrRNA sequences base-pairing with crRNA repeats. Putative RNA processing sites as revealed by RNA sequencing are indicated with arrowheads. For each locus, arrowhead sizes represent relative amounts of the retrieved 5' and 3' ends (see also **Figure 37**).

[00594] **Figure 37** lists all tracrRNA orthologues and mature crRNAs retrieved by sequencing for the bacterial species studied, including coordinates (region of interest) and corresponding cDNA sequences (5' to 3'). The arrows represent the transcriptional direction (strand). Number of cDNA reads (calculated using SAMtools), coverage numbers (percentage of mapped reads) and predominant ends associated with each transcript are indicated. Numbers of reads starting or stopping at each nucleotide position around the 5' and 3' ends of each transcript are displayed. The sizes of each crRNA mature forms are indicated. The number allocated to each crRNA species corresponds to the spacer sequence position in the pre-crRNA, according to the CRISPRdb. The number allocated to each tracrRNA species corresponds to different forms of the same transcript.

[00595] We then screened the selected CRISPR-Cas loci including sequences located 1 kb upstream and downstream on both strands for possible repeat sequences that did not belong to the repeat-spacer array, allowing up to 15 mismatches. On average, we found one to three degenerated repeat sequences per locus that would correspond to anti-repeats of tracrRNA orthologues and selected the sequences located within the intergenic regions. The putative anti-repeats were found in four typical localizations: upstream of the cas9 gene, in the region between cas9 and cas1, and upstream or downstream of the repeat-spacer array (Figure 33). For every retrieved sequence, we validated the extent of base-pairing formed between the repeat and anti-repeat (Figure 44) by predicting the possible RNA:RNA interaction and focusing especially on candidates with longer and perfect complementarity region forming an optimal double-stranded structure for RNase III processing. To predict promoters and transcriptional terminators flanking the anti-repeat, we set the putative transcription start and termination sites to be included within a region located maximally 200 nt uptream and 100 nt downstream of the anti-repeat sequence, respectively, based on our previous observations[26]. As mentioned above, experimental information on the transcriptional direction of most repeat-spacer arrays of type II systems is lacking. The in silico promoter prediction algorithms often give false positive results and point to putative promoters that would lead to the transcription of repeat-spacer arrays from both strands. In some cases we could not predict transcriptional terminators, even though the tracrRNA orthologue expression could be validated experimentally, as exemplified by the C. jejuni locus

(see paragraph 'Deep RNA sequencing validates expression of novel tracrRNA orthologues'). We suggest to consider promoter and transcriptional terminator predictions only as a supportive, but not essential, step of the guideline described above.

[00596] Figure 44 depicts predicted pre-crRNA repeat:tracrRNA anti-repeat basepairing in selected bacterial species. [b]The CRISPR loci belong to the type II (Nmeni/CASS4) CRISPR-Cas system. Nomenclature is according to the CRISPR database (CRISPRdb). Note that S. thermophilus LMD-9 and W. succinogenes contain two type II loci. [c]Upper sequence, pre-crRNA repeat consensus sequence (5' to 3'); lower sequence, tracrRNA homologue sequence annealing to the repeat (anti-repeat; 3' to 5'). Note that the repeat sequence given is based on the assumption that the CRISPR repeat-spacer array is transcribed from the same strand as the cas operon. For the sequences that were validated experimentally in this study, RNA sequencing data were taken into account to determine the base-pairing. See Figure 33. [d]Two possible anti-repeats were identified in the F. tularensis subsp. novicida, W. succinogenes and gamma proteobacterium HTCC5015 type II-A loci. Upper sequence pairing, anti-repeat within the putative leader sequence; lower sequence pairing, anti-repeat downstream of the repeat spacer array. See Figure 33. [e]Two possible anti-repeats were identified in the S. wadsworthensis type II-A locus. Upper sequence pairing, anti-repeat; lower sequence pairing, anti-repeat within the putative leader sequence See Figure 33. [f]Two possible anti-repeats were identified in the L. gasseri type II-B locus. Upper sequence pairing, anti-repeat upstream of cas9; lower sequence pairing, anti-repeat between the cas9 and cas1 genes. See Figure 33. [g]Two possible anti-repeats were identified in the C. jejuni type II-C loci. Upper sequence pairing, anti-repeat upstream of cas9; lower sequence pairing, anti-repeat downstream of the repeat-spacer array. See Figure 33. [h]Two possible anti-repeats were identified in the R. rubrum type II-C locus. Upper sequence pairing, antirepeat downstream of the repeat-spacer array; lower sequence pairing, anti-repeat upstream of cas1. See Figure 33.

A plethora of tracrRNA orthologues

[00597] We predicted putative tracrRNA orthologues for 56 of the 75 loci selected earlier. The results of predictions are depicted in Figure 33. As already mentioned, the direction of tracrRNA transcription indicated in this figure is hypothetical and based on the indicated direction of repeat-spacer array transcription. As previously stated, sequences encoding putative tracrRNA orthologues were identified upstream, within and downstream of the cas operon, as well as downstream of the repeat spacer arrays, including the putative leader sequences, commonly found in type II-A loci (Figure 33). However, we observed that anti-repeats of similar localization within CRISPR-Cas loci can be transcribed in different directions (as observed when comparing e.g. Lactobacillus rhamnosus and Eubacterium rectale or Mycoplasma mobile and S.

pyogenes or N. meningitidis) (Figure 33). Notably, loci grouped within a same subcluster of the Cas9 guide tree share a common architecture with respect to the position of the tracrRNA-encoding gene. We identified anti-repeats around the repeat-spacer array in type II-A loci, and mostly upstream of the cas9 gene in types II-B and II-C with several notable exceptions for the putative tracrRNA located between cas9 and cas1 in three distinct subclusters of type II-B.

Some type II CRISPR-Cas loci have defective repeat-spacer arrays and/or tracrRNA orthologues

[00598] For six type II loci (Fusobacterium nucleatum, Aminomonas paucivorans, Helicobacter mustelae, Azospirillum sp., Prevotella ruminicola and Akkermansia muciniphila), we identified potential anti-repeats with weak base-pairing to the repeat sequence or located within the open reading frames. Notably, in these loci, a weak anti-repeat within the open reading frame of the gene encoding a putative ATPase in A. paucivorans, a strong anti-repeat within the first 100 nt of the cas9 gene in Azospirillum sp. B510 and a strong anti-repeat overlapping with both cas9 and cas1 in A. muciniphila were identified (Figure 33). For twelve additional loci (Peptoniphilus duerdenii, Coprococcus catus, Acidaminococcus intestini, Catenibacterium mitsuokai, Staphylococcus pseudintermedius, Ilyobacter polytropus, Elusimicrobium minutum, Bacteroides fragilis, Acidothermus cellulolyticus, Corynebacterium diphteriae, Bifidobacterium longum and Bifidobacterium dentium), we could not detect any putative anti-repeat. There is no available information on pre-crRNA expression and processing in these CRISPR-Cas loci. Thus, the functionality of type II systems in the absence of a clearly defined tracrRNA orthologue remains to be addressed. For seven analyzed loci we could not identify any repeat spacer array (Parasutterella excrementihominis, Bacillus cereus, Ruminococcus albus, Rhodopseudomonas palustris, Nitrobacter hamburgensis, Bradyrhizobium sp. and Prevotella micans) (Figure 33) and in three of those (Bradyrhizobium sp. BTAi1, N. hamburgensis and B. cereus) we detected cas9 as a single gene with no other cas genes in the vicinity. For these three loci, we failed to predict any small RNA sequence upstream or downstream of the cas9 gene. In the case of R. albus and P. excrementihominis, the genomic contig containing cas9 is too short to allow prediction of the repeat spacer array.

Deep RNA sequencing validates expression of novel tracrRNA orthologues

[00599] To verify the in silico tracrRNA predictions and determine tracrRNA:pre-crRNA coprocessing patterns, RNAs from selected Gram-positive (S. mutans and L. innocua) and Gram-negative (N. meningitidis, C. jejuni and F. novicida) bacteria were analyzed by deep sequencing. Sequences of tracrRNA orthologues and processed crRNAs were retrieved (Figure 36 and Figure 37). Consistent with previously published differential tracrRNA sequencing data in S. pyogenes[26], tracrRNA orthologues were highly represented in the libraries, ranging from 0.08 to 6.2% of

total mapped reads. Processed tracrRNAs were also more abundant than primary transcripts, ranging from 66% to more than 95% of the total amount of tracrRNA reads (Figure 36 and Figure 37).

[00600] **Figure 36** depicts the expression of bacterial tracrRNA orthologues and crRNAs revealed by deep RNA sequencing. Expression profiles of tracrRNA orthologues and crRNAs of selected bacterial strains are represented along the corresponding genomes by bar charts (Images captured from the Integrative Genomics Viewer (IGV) tool). Campylobacter jejuni (GenBank: NC_002163), Francisella novicida (GenBank: NC_008601), Neisseria meningitidis (GenBank: NC_003116), Listeria innocua (GenBank: NC_003212) and Streptococcus mutans (GenBank: NC_004350). Genomic coordinates are given. [a]Sequence coverage calculated using BEDTools-Version-2.15.0 (Scale given in reads per million). [b]Distribution of reads starting (5') and ending (3') at each nucleotide position are indicated (Scale given in numbers of reads). Upper panels correspond to transcripts from the positive strand and lower panels correspond to transcripts from the negative strand. The negative coverage values and peaks presented below the axes indicate transcription from the negative strand of the genome. Predominant 5'- and 3'-ends of the reads are plotted for all RNAs. Note that given the low quality of L. innocua cDNA library, the reads are shortened for crRNAs, and an accumulation of the reads at the 3' end of tracrRNA is observed, presumably due to RNA degradation.

[00601] To assess the 5' ends of tracrRNA primary transcripts, we analyzed the abundance of all 5' end reads of tracrRNA and retrieved the most prominent reads upstream or in the vicinity of the 5' end of the predicted anti-repeat sequence. The 5' ends of tracrRNA orthologues were further confirmed using the promoter prediction algorithm. The identified 5' ends of tracrRNAs from S. mutans, L. innocua and N. meningitidis correlated with both in silico predictions and Northern blot analysis of tracrRNA expression[26]. The most prominent 5' end of C. jejuni tracrRNA was identified in the middle of the anti-repeat sequence. Five nucleotides upstream, an additional putative 5' end correlating with the in silico prediction and providing longer sequence of interaction with the CRISPR repeat sequence was detected. We retrieved relatively low amount of reads from the F. novicida library that corresponded almost exclusively to processed transcripts. Analysis of the very small amount of reads of primary transcripts provided a 5' end that corresponded to the strong in silico promoter predictions. Northern blot probing of F. novicida tracrRNA further confirmed the validity of the predictions showing the low abundance of transcripts of around 90 nt in length. The results are listed in **Table 2**. For all examined species, except N. meningitidis, primary tracrRNA transcripts were identified as single small RNA species of 75 to 100 nt in length. In the case of N. meningitidis, we found a predominant

primary tracrRNA form of ~110 nt and a putative longer transcript of ~170 nt represented by a very low amount of reads and detected previously as a weak band by Northern blot analysis.

[00602] Table 2. Selected tracrRNA orthologues

Strains[a]	Transcript	5'-end[b]			3'-end[c]	Length (nt)
		RNA-seq		Predicted		
		First read	Most prominent			
S. pyogenes SF370	*primary*	-	854 546	-	854 378	171
	primary	-	854 464	-		89
	processed	-	854 450	-		~75
C. jejuni NCTC 11168	*primary*	1 455 497	1 455 502	1 455 497	1 455 570	~75
	processed	-	1 455 509	-		~60
L. innocua Clip11262	*primary*	2 774 774	2 774 774	2 774 773	2 774 863	~90
	processed	-	2 774 788	-		~75
S. mutans UA159	*primary*	1 335 040	1 335 040	1 355 039	1 335 141	~100
	processed	-	1 335 054 1 335 062	-		~85 ~80
N. meningitidis A Z2491	*primary*	614 158	614 162	614 154	614 333	~175
	primary	614 223	614 225	614 223		~110
	processed	-	614 240	-		~90
F. novicida U112	*primary*	817 144	-	817 145 817 154	817 085	~80
	processed	-	817 138 817 128	-		~75 ~65
S. thermophilus LMD-9	*primary*	-	-	1 384 330	1 384 425	~95
	primary	-	-	646 654	646 762	~110
P. multocida Pm70	*primary*	-	-	1 327 287	1 327 396	~110
M. mobile 163K	*primary*	-	-	49 470	49 361	~110

[a]tracrRNA orthologues of *S. thermophilus*, *P. multocida* and *M. mobile* were predicted *in silico*.

[b]RNA-seq, revealed by RNA sequencing (Table S3); first read, first 5'-end position retrieved by sequencing; most prominent, abundant 5'-end according to RNA-seq data; predicted, *in silico* prediction of transcription start site; underlined, 5'-end chosen for the primary tracrRNA to be aligned.

[c]Estimated 3' end according to RNA-seq data and transcriptional terminator prediction.

tracrRNA and pre-crRNA co-processing sites lie in the anti-repeat:repeat region.

[00603] We examined the processed tracrRNA transcripts by analyzing abundant tracrRNA 5' ends within the predicted anti-repeat sequence and abundant mature crRNA 3' ends (Figure 34 and 45). In all species, we identified the prominent 5' ends of tracrRNA orthologues that could result from co-processing of the tracrRNA:pre-crRNA repeat duplexes by RNase III. We also identified the processed 5'-ends of crRNAs that most probably result from a second maturation event by putative trimming, consistently with previous observations. Noteworthy, in the closely related RNA pairs of S. pyogenes, S. mutans and L. innocua, we observed the same processing

site around the G:C basepair in the middle of the anti-repeat sequence. In both S. mutans and L. innocua, we detected additional prominent tracrRNA 5' ends and crRNA 3' ends that could suggest further trimming of the tracrRNA:crRNA duplex, with 3'-end of crRNA being shortened additionally to the already mentioned 5'-end trimming, following the RNase III-catalyzed first processing event. Similarly, in C. jejuni we found only a small amount of crRNA 3' ends that would fit to the RNase III processing patterns and retrieved the corresponding 5' ends of processed tracrRNA. Thus, the putative trimming of tracrRNA:crRNA duplexes after initial cleavage by RNase III would result in a shorter repeat-derived part in mature crRNAs, producing shorter tracrRNA:crRNA duplexes stabilized by a triple G:C base-pairing for interaction with the endonuclease Cas9 and subsequent cleavage of target DNAs. The N. meningitidis RNA duplex seems to be processed at two primary sites further to the 3' end of the CRISPR repeat, resulting in a long repeat-derived part in mature crRNA and stable RNA:RNA interaction despite the central bulge within the duplex. Interestingly, the tracrRNA:pre-crRNA duplex of F. novicida seems to be cleaved within the region of low complementarity and some of the retrieved abundant 5' ends of tracrRNA suggest its further trimming without concomitant trimming of crRNA. Differences in primary transcript sizes and in the location of processing sites result in various lengths of processed tracrRNAs ranging from ~65 to 85 nt. The coordinates and sizes of the prominent processed tracrRNA transcripts are shown in **Table 2** and **Figure 37**. The observed processing patterns of tracrRNA and crRNA are well in agreement with the previously proposed model of two maturation events. The putative further trimming of some of the tracrRNA 5'-ends and crRNA 3'-ends could stem from the second maturation event or alternatively, be an artifact of the cDNA library preparation or RNA sequencing. The nature of these processings remains to be investigated further.

Sequences of tracrRNA orthologues are highly diverse

[00604] Sequences similarities of selected tracrRNA orthologues were also determined. We performed multiple sequence alignments of primary tracrRNA transcripts of S. pyogenes (89 nt form only), S. mutans, L. innocua and N. meningitidis (110 nt form only), S. thermophilus, P. multocida and M. mobile (**Table 2, Figure 35**). We observed high diversity in tracrRNA sequences but significant conservation of sequences from closely related CRISPR-Cas loci. tracrRNAs from L. innocua, S. pyogenes, S. mutans and S. thermophiles share on average 77% identity and tracrRNAs from N. meningitidis and P. multocida share 82% identity according to pairwise alignments. The average identity of the analyzed tracrRNA sequences is 56%, comparable to the identity of random RNA sequences. This observation further confirms that the prediction of tracrRNA orthologues based on sequence similarity can be performed only in the case of closely related loci. We also sought for possible tracrRNA structure conservation but could not find any

significant similarity except one co-variation and conserved transcriptional terminator structure (**Figure 35**).

[00605] **Figure 35** depicts sequence diversity of tracrRNA orthologues. tracrRNA sequence multiple alignment. S. thermophilus and S. thermophilus2, tracrRNA associated with SEQ ID NO:41 and SEQ ID NO:40 Cas9 orthologues, accordingly. Black, highly conserved; dark grey, conserved; light grey, weakly conserved. Predicted consensus structure is depicted on the top of the alignment. Arrows indicate the nucleotide covariations. S. pyogenes SF370, S. mutans UA159, L. innocua Clip11262, C. jejuni NCTC 11168, F. novicida U112 and N. meningitidis A Z2491 tracrRNAs were validated by RNA sequencing and Northern blot analysis. S. thermophiles LMD-9 tracrRNA was validated by Northern blot analysis. P. multocida Pm70 tracrRNA was predicted from high similarity of the CRISPR-Cas locus with that of N. meningitidis A Z2491. M. mobile 163K tracrRNA was predicted in silico from strong predictions of transcriptional promoter and terminator.

Example 4: Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression

[00606] Targeted gene regulation on a genome-wide scale is a powerful strategy for interrogating, perturbing and engineering cellular systems. The inventors have developed a new method for controlling gene expression, based on Cas9, an RNA-guided DNA endonuclease from a Type II CRISPR system. This example demonstrates that a catalytically dead Cas9, lacking endonuclease activity, when co-expressed with a guide RNA, generates a DNA recognition complex that can specifically interfere with transcriptional elongation, RNA polymerase binding, or transcription factor binding. This system, called CRISPR interference (CRISPRi), can efficiently repress expression of targeted genes in Escherichia coli with no detectable off-target effects. CRISPRi can be used to repress multiple target genes simultaneously, and its effects are reversible. In addition, the system can be adapted for gene repression in mammalian cells. This RNA-guided DNA recognition platform provides a simple approach for selectively perturbing gene expression on a genome-wide scale.

MATERIALS AND METHODS

Strains and Media

[00607] The *Escherichia coli* K-12 strain MG1655 was used as the host strain for the in vivo fluorescence measurements. An *E. coli* MG1655-derived strain that endogenously expresses a variant of RNAP with a 3x-FLAG epitope tag attached to the C-terminal end of the RpoC subunit was used for all sequencing experiments. EZ rich defined media (EZ-RDM, Teknoka) was used as the growth media for in vivo fluorescence assays. Genetic transformation and

verification of transformation were done using standard protocols, using AmpR, CmR, or KanR genes as selectable markers.

Plasmid Construction and E. coli Genome Cloning

[00608] The Cas9 and dCas9 genes were cloned from the previous described vector pMJ806 and pMJ841, respectively. The genes were PCR amplified and inserted into a vector containing an anhydrotetracycline (aTc)-inducible promoter PLtetO-1, a chloramphenicol selectable marker and a p15A replication origin. The sgRNA template was cloned into a vector containing a minimal synthetic promoter (J23119) with an annotated transcription start site, an ampicillin selectable marker and a ColE1 replication origin. Inverse PCR was used to generate sgRNA cassettes with new 20-bp complementary regions. To insert fluorescent reporter genes into E. coli genomes, the fluorescence gene was first cloned onto an entry vector, which was then PCR amplified to generate linearized DNA fragments that contained nsfA 5'/3' UTR sequences, the fluorescent gene and a KanR selectable marker. The E. coli MG1655 strain was transformed with a temperature-sensitive plasmid pKD46 that contained λ-Red recombination proteins (Exo, Beta and Gama). Cell cultures were grown at 30 °C to an OD (600 nm) of ~ 0.5, and 0.2 % arabinose was added to induce expression of the λ-Red recombination proteins for 1 h. The cells were harvested at 4°C, and used for transformation of the linearized DNA fragments by electroporation. Cells that contain correct genome insertions were selected by using 50 μg/mL Kanamycin.

Flow cytometry and analysis

[00609] Strains were cultivated in EZ-RDM containing 100 μg/mL carbenicillin and 34 μg/mL chloramphenicol in 2mL 96-well deep well plates (Costar 3960) overnight at 37 C and 1200 r.p.m. One-μL of this overnight culture was then added to 249 μL of fresh EZ-RDM with the same antibiotic concentrations with 2 μM aTc supplemented to induce production of the dCas9 protein. When cells were grown to mid-log phase (~4h), the levels of fluorescence protein were determined using the LSRII flow cytometer (BD Biosciences) equipped with a high-throughput sampler. Cells were sampled with a low flow rate until at least 20,000 cells had been collected. Data were analyzed using FCS Express (De Novo Software) by gating on a polygonal region containing 60% cell population in the forward scatter-side scatter plot. For each experiment, triplicate cultures were measured, and their standard deviation was indicated as the error bar.

B-galactosidase assay

[00610] To perform β-galactosidase assay, 1 μL of overnight culture prepared as above was added to 249 μL of fresh EZ-RDM with the same antibiotic concentrations with 2 μM aTc, with or without 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were grown to mid-log

phase. The LacZ activity of 100 uL of this culture was measured using the yeast β-galactosidase assay kit (Pierce) following the instructions.

Extraction and purification of total RNA

[00611] For each sample, a monoclonal culture of *E. coli* was grown at 37˚C from an OD (600 nm) 0.1 in 500 mL of EZ-RDM to early log-phase (OD 0.45 ± 0.05), at which point the cells were harvested by filtration over 0.22 μm nitrocellulose filters (GE) and frozen in liquid nitrogen to simultaneously halt all transcriptional progress. Frozen cells (100 μg) were pulverized on a Qiagen TissueLyser II mixer mill 6 times at 15 Hz for 3 min in the presence of 500 μL frozen lysis buffer (20 mM Tris pH 8, 0.4% Triton X-100, 0.1% NP-40, 100 mM NH_4Cl, 50 U/mL SUPERase•In (Ambion) and 1x protease inhibitor cocktail (Complete, EDTA-free, Roche), supplemented with 10 mM $MnCl_2$ and 15 μM Tagetin transcriptional inhibitor (Epicentre).

[00612] The lysate was resuspended on ice by pipetting. RQ1 DNase I (110 U total, Promega) was added and incubated for 20 min on ice. The reaction was quenched with EDTA (25 mM final) and the lysate clarified at 4°C by centrifugation at 20,000g for 10 min. The lysate was loaded onto a PD MiniTrap G-25 column (GE Healthcare) and eluted with lysis buffer supplemented with 1 mM EDTA.

Total mRNA purification

[00613] Total RNA was purified from the clarified lysate using the miRNeasy kit (Qiagen). 1 μg of RNA in 20 μL of 10 mM Tris pH 7 was mixed with an equal volume of 2x alkaline fragmentation solution (2 mM EDTA, 10 mM Na_2CO_3, 90 mM $NaHCO_3$, pH 9.3) and incubated for ~25 min at 95 °C to generate fragments ranging from 30-100 nt. The fragmentation reaction was stopped by adding 0.56 mL of ice-cold precipitation solution (300 mM NaOAc pH 5.5 plus GlycoBlue (Ambion)), and the RNA was purified by a standard isopropanol precipitation. The fragmented mRNA was then dephosphorylated in a 50 μL reaction with 25 U T4 PNK (NEB) in 1x PNK buffer (without ATP) plus 0.5 U SUPERase•In, and precipitated with GlycoBlue via standard isopropanol precipitation methods.

Nascent RNA purification

[00614] For nascent RNA purification, the clarified lysate was added to 0.5 mL anti-FLAG M2 affinity gel (Sigma Aldrich) as described previously. The affinity gel was washed twice with lysis buffer supplemented with 1 mM EDTA before incubation with the clarified lysate at 4°C for 2.5 h with nutation. The immunoprecipitation was washed 4 × 10 ml with lysis buffer supplemented with 300 mM KCl, and bound RNAP was eluted twice with lysis buffer supplemented with 1 mM EDTA and 2 mg/mL 3x-FLAG peptide (Sigma Aldrich). Nascent RNA was purified from the eluate using the miRNeasy kit (Qiagen) and converted to DNA using a previously established library generation protocol.

DNA library preparation and DNA sequencing

[00615] The DNA library was sequencing on an Illumina HiSeq 2000. Reads were processed using the HTSeq Python package and other custom software written in Python. The 3′-end of the sequenced transcript was aligned to the reference genome using Bowtie ("bowtie-bio" preceeding ".sourceforge.net") and the RNAP profiles generated in MochiView ("johnsonlab.ucsf" preceeding ".edu/mochi.html").

Plasmid design and construction for CRISPRi in human cells

[00616] The sequence encoding mammalian codon optimized *Streptococcus pyogenes* Cas9 (DNA 2.0) was fused with three C-terminal SV40 nuclear localization sequences (NLS) or to tagBFP flanked by two NLS. Using standard ligation independent cloning we cloned these two fusion proteins into MSCV-Puro (Clontech). Guide sgRNAs were expressed using a lentiviral U6 based expression vector derived from pSico which co-expresses mCherry from a CMV promoter. The sgRNA expression plasmids were cloned by inserting annealed primers into the lentiviral U6 based expression vector that was digested by BstXI and XhoI.

Cell culture, DNA transfections and fluorescence measurements for CRISPRi in human cells

[00617] HEK293 cells were maintained in Dulbecco's modified eagle medium (DMEM) in 10%FBS, 2mM glutamine, 100units/mL streptomycin and 100µg/mL penicillin. HEK293 were infected with a GFP expressing MSCV retrovirus using standard protocols and sorted by flow cytometry using a BD FACS Aria2 for stable GFP expression. GFP expressing HEK293 cells were transiently transfected using TransIT-LT1 transfection reagent (Mirus) with the manufacturers recommended protocol in 24 well plates using 0.5 µg of the dCas9 expression plasmid and 0.5 µg of the RNA expression plasmid (with 0.25 µg of GFP reporter plasmid for Figure 45B). 72 hours following transfection, cells were trypsinized to a single cell suspension. The U6 vector contains a constitutive CMV promoter driving a mCherry gene. GFP expression was analyzed using a BD LSRII FACS machine by gating on the mCherry positive populations (>10-fold brighter mCherry over the negative control cells).

Designed RNAs

[00618] sgRNA designs used in the figures: only the 20 nucleotide matching region (DNA targeting segment) are listed (unless otherwise noted):

The mRFP-targeting sgRNAs used in Figure 40C (SEQ ID NOs:741-746);
The promoter-targeting sgRNAs used in Figure 40D (SEQ ID NOs:747-751);
Target promoter sequence in Figure 40D (SEQ ID NO:752);
The mRFP-targeting sgRNAs used in Figure 43B (SEQ ID NOs:753-760);

The sfGFP-targeting sgRNA (gfp) used in Figure 42B (SEQ ID NO:761);

The sfGFP-targeting sgRNAs used in Figure 43B (SEQ ID NOs:762-769);

The double-sgRNA targeting experiments in Figure 43F and Figure 51 (SEQ ID NOs:770-778);

The lac operon-targeting sgRNAs used in Figure 44B (SEQ ID NOs:779-787); and

The EGFP-targeting sgRNAs used in Figure 45 (SEQ ID NOs:788-794).

Table 3. Sequences used in the figures of Example 4 (listed above)

Sequence	SEQ ID NO:	Sequence	SEQ ID NO:	Sequence	SEQ ID NO:
T1	741	R2	771	crp	783
T2	742	R3	772	cya	784
T3	743	R4	773	A site	785
NT1	744	R5	774	O site	786
NT2	745	R6	775	P site	787
NT3	746	R7	776	eT1	788
P1	747	R8	777	eT2	789
P2	748	R9	778	eNT1	790
P3	749	lacZ	779	eNT2	791
P4	750	lacI	780	eNT3	792
P5	751	lacY	781	eNT4	793
R1	770	lacA	782	eNT5	794

RESULTS

[00619] The CRISPR (clustered regularly interspaced short palindromic repeats) system provides a new potential platform for targeted gene regulation. About 40% of bacteria and 90% of archaea possess CRISPR/CRISPR-associated (Cas) systems to confer resistance to foreign DNA elements. CRISPR systems use small base-pairing RNAs to target and cleave foreign DNA elements in a sequence-specific manner. There are diverse CRISPR systems in different organisms, and one of the simplest is the type II CRISPR system from Streptococcus pyogenes: only a single gene encoding the Cas9 protein and two RNAs, a mature CRISPR RNA (crRNA) and a partially complementary trans-acting RNA (tracrRNA), are necessary and sufficient for RNA-guided silencing of foreign DNAs (**Figure 46**). Maturation of crRNA requires tracrRNA and RNase III. However, this requirement can be bypassed by using an engineered small guide

RNA (sgRNA) containing a designed hairpin that mimics the tracrRNA-crRNA complex. Base pairing between the sgRNA and target DNA causes double-strand breaks (DSBs) due to the endonuclease activity of Cas9. Binding specificity is determined by both sgRNA-DNA basepairing and a short DNA motif (protospacer adjacent motif or PAM, sequence: NGG) juxtaposed to the DNA complementary region. Thus, the CRISPR system only requires a minimal set of two molecules-the Cas9 protein and the sgRNA, and therefore holds the potential to be used as a host-independent gene-targeting platform. It has been demonstrated that the Cas9/CRISPR can be harnessed for site-selective RNA-guided genome editing (Figure 39A).

[00620] **Figure 46** depicts the mechanism of the type II CRISPR system from S. pyogenes. The system consists of a set of CRISPR-associated (Cas) proteins and a CRISPR locus that contains an array of repeat spacer sequences. All repeats are the same and all spacers are different and complementary to the target DNA sequences. When the cell is infected by foreign DNA elements, the CRISPR locus will transcribe into a long precursor transcript, which will be cleaved into smaller fragments. The cleavage is mediated by a transacting antisense RNA (tracrRNA) and the host RNase III. After cleavage, one single protein, Cas9, recognizes and binds to the cleaved form of the crRNA. Cas9 guides crRNA to DNA and scans the DNA molecule. The complex is stabilized by basepairing between the crRNA and the DNA target. In this case, Cas9 causes double-stranded DNA breaks due to its nuclease activity. This usually removes cognate DNA molecules, and cells confer immunity to certain DNA populations.

[00621] **Figure 39** depicts the design of the CRISPR interference (CRISPRi) system. (**A**) The minimal interference system consists of a single protein and a designed sgRNA chimera. The sgRNA chimera consists of three domains (boxed region): a 20-nucleotide (nt) complementary region for specific DNA binding, a 42-nt hairpin for Cas9 binding (Cas9 handle), and a 40-nt transcription terminator derived from S. pyogenes. The wild-type Cas9 protein contains the nuclease activity. The dCas9 protein is defective in nuclease activity. (**B**) The wild-type Cas9 protein binds to the sgRNA and forms a protein-RNA complex. The complex binds to specific DNA targets by Watson-Crick basepairing between the sgRNA and the DNA target. In the case of wild-type Cas9, the DNA will be cleaved due to the nuclease activity of the Cas9 protein. In the case of nuclease defective Cas9, the complex disrupts appropriate transcription.

A minimal CRISPRi system consists of a single protein and RNA and can effectively silence transcription initiation and elongation

[00622] To implement such a CRISPRi platform in E. coli, the wild-type S. pyogenes Cas9 gene and an sgRNA were expressed from bacterial vectors to determine if the system could perturb gene expression at a targeted locus (Figure 40A). The S. pyogenes CRISPR system is orthogonal to the native E. coli system. The Cas9 protein is expressed from an anhydrotetracycline (aTc)-

inducible promoter on a plasmid containing a p15A replication origin, and the sgRNA is expressed from a minimal constitutive promoter on a plasmid containing a ColE1 replication origin. As an alternative strategy, a catalytically dead Cas9 mutant (dCas9), which is defective in DNA cleavage, was used and showed that this form of Cas9 still acts as a simple RNA-guided DNA binding complex.

[00623] Figure 40 demonstrates that CRISPRi effectively silences transcription elongation and initiation. (**A**) The CRISPRi system consists of an inducible Cas9 protein and a designed sgRNA chimera. The dCas9 contains mutations of the RuvC1 and HNH nuclease domains. The sgRNA chimera contains three functional domains as described in Fig. 1. (**B**) Sequence of designed sgRNA (NT1) and the DNA target. NT1 targets the non-template DNA strand of mRFP coding region. Only the region surrounding the base-pairing motif (20-nt) is shown. Base-pairing nucleotides are numbered and the dCas9-binding hairpin is overlined. The PAM sequence is underlined. (**C**) CRISPRi blocked transcription elongation in a strand-specific manner. A synthetic fluorescence-based reporter system containing an mRFP-coding gene was inserted into the *E. coli* MG1655 genome (the *nsfA* locus). Six sgRNAs that bind to either the template DNA strand or the non-template DNA strand were co-expressed with the dCas9 protein, with their effects on the target mRFP measured by *in vivo* fluorescence assay. Only sgRNAs that bind to the non-template DNA strand showed silencing (10~300-fold). The control shows fluorescence of the cells with dCas9 protein but without the sgRNA. (**D**) CRISPRi blocked transcription initiation. Five sgRNAs were designed to bind to different regions around an *E. coli* promoter (J23119). The transcription start site was labeled as +1. The dotted oval shows the initial RNAP complex that covers a 75-bp region from -55 to +20. Only sgRNAs targeting regions inside the initial RNAP complex showed repression (P1-P4). Unlike transcription elongation block, silencing was independent of the targeted DNA strand. (**E**) CRISPRi regulation was reversible. Both dCas9 and sgRNA (NT1) were under the control of an aTc-inducible promoter. Cell culture was maintained during exponential phase. At time T = 0, 1 μM of aTc was supplemented to cells with OD = 0.001. Repression of target mRFP started within 10 min. The fluorescence signal decayed in a way that is consistent with cell growth, suggesting the decay was due to cell division. In 240 min, the fluorescence reached the fully repressed level. At T = 370 min, aTc is washed away from the growth media, and cells were diluted back to OD = 0.001. Fluorescence started to increase after 50 min, and took about 300 min to rise to the same level as the positive control. Positive control: always without the inducer; negative control: always with 1 μM aTc inducer. Fluorescence results in 2C, 2D, and 2E represent average and SEM of at least three biological replicates. See also Figure 47 and Figure 48.

[00624] The sgRNA molecules co-expressed with Cas9 each consist of three segments: a 20-nucleotide (nt) target-specific complementary region, a 42-nt Cas9 binding hairpin (Cas9 handle) and a 40-nt transcription terminator derived from S. pyogenes (Figure 40B). A red fluorescent protein (mRFP)-based reporter system, was inserted it into the E. coli MG1655 genome.

[00625] Co-expression of the wild-type Cas9 protein and an sgRNA (NT1) targeted to the mRFP coding sequence dramatically decreased transformation efficiency, likely due to Cas9-induced double-stranded breaks on the genome (Figure 47A). Sequencing of a few survivor colonies showed that they all had sequence rearrangements around the target mRFP site on the genome, suggesting that there was strong selection against expression of wild-type Cas9 and an sgRNA targeted to a host sequence. The dCas9 mutant gene (non-cleaving), which contained two silencing mutations of the RuvC1 and HNH nuclease domains (D10A and H841A), alleviated this lethality, as verified by transformation efficiency and E. coli growth rates (Figure 47A&B).

[00626] **Figure 47** is related to Figure 40 and shows Growth curves of E. coli cell cultures co-transformed with dCas9 and sgRNA. (**A**) Transformation efficiency for transforming E. coli cells with two plasmids. One plasmid contains an sgRNA that targets to a genomic copy of mRFP and the other plasmid contains the wild-type Cas9 or dCas9. Co-transformation of wild-type Cas9 and sgRNA is highly toxic, which can be alleviated using dCas9. (B) The sgRNA (NT1) is designed to target the coding sequence of mRFP. Co-expression of dCas9 and sgRNA exhibits almost no effects on cellular growth rates, suggesting the dCas9-sgRNA interaction with DNA is strong enough to block RNA polymerase but not DNA polymerase or cell replication. The results represent average and SEM of at least three independent experiments.

[00627] To test whether the dCas9:sgRNA complex could yield highly efficient repression of gene expression, sgRNAs complementary to different regions of the mRFP coding sequence were designed, either binding to the template DNA strand or the non-template DNA strand. The results indicated that sgRNAs targeting the non-template DNA strand demonstrated effective gene silencing (10 to 300-fold of repression), while those targeting the template strand showed little effect (Figure 40C). The system exhibited similar repression effects for genes that were within the E. coli genome or on a high-copy plasmid (Figure 48). Furthermore, targeting to the promoter region also led to effective gene silencing (Figure 40D). Targeting of the sgRNA to the -35 box significantly knocked down gene expression (P1, ~100-fold of repression), whereas targeting to other adjacent regions showed a dampened effect (P2-P4). Targeting sequences about 100-bp upstream of the promoter showed no effects (P5). Unlike targeting the coding sequence, when targeting the promoter, the efficiency of silencing is independent of the DNA strand; targeting of template or non-template strands is equally effective (P2 and P3).

[00628] **Figure 48** is related to Figure 40C and shows that CRISPRi could silence expression of a reporter gene on a multiple-copy plasmid. The mRFP gene was cloned onto a p15A plasmid. Presence of the dCas9 and an mRFP-specific sgRNA (NT1) strongly represses mRFP (~300-fold). The repression effect is similar to that observed using the mRFP in the genome (Figure 40C). Silencing is only effective when the sgRNA acts on the nontemplate DNA strand but not the template DNA strand (T1). Also, silencing is highly specific, as a GFP-specific 3 sgRNA (gfp) shows no effect on mRFP expression. Fluorescence results represent average and SEM of at least three biological replicates.

CRISPRi gene knockdown is inducible and reversible

[00629] Unlike gene knockout methods, one advantage of using CRISPRi-based knock down of gene expression is the fact that this perturbation should be reversible. To test if CRISPRi regulation could be induced and subsequently reversed, both dCas9 and mRFP-specific sgRNA (NT1) were placed under the control of the aTc-inducible promoter, and time-course measurements of CRISPRi-mediated regulation of mRFP in response to inducers were performed (Figure 40E). At time zero, cell culture that grew to the early exponential phase without inducers was supplemented with 1 μM of aTc. The data indicated that the system could quickly respond to the presence of inducers – the fluorescent reporter protein signal started to decrease within 10 min of the addition of the inducer molecule. Because the mRFP protein is stable, the rate of fluorescence signal decrease is limited by protein dilution due to cell growth, as seen by a similar cell doubling time and loss of fluorescence half-time (both ~36 min). At 240 min, all cells were uniformly repressed to the same level as the negative control. At 420 min, the inducer was washed away from the growth media and cells were diluted back to a lower OD. After a delay of 50 min, mRFP fluorescence started to increase. It took a total 300 min for single-cell fluorescence to increase to the same level as the positive control. The 50 min delay is most likely determined by the dCas9/sgRNA turnover rate offset by dilution by cell growth and division. In summary, these results demonstrate that the silencing effects of dCas9-sgRNA can be induced and reversed.

Native elongating transcript sequencing (NET-Seq) confirms that CRISPRi functions by blocking transcription

[00630] dCas9 appeared to be functioning as an RNA-guided DNA binding complex that could block RNA polymerase (RNAP) binding during transcription elongation. Since the non-template DNA strand shares the same sequence identity as the transcribed mRNA and only sgRNAs that bind to the non-template DNA strand exhibited silencing, it remained a possibility that the dCas9:sgRNA complex interacts with mRNA and alters its translation or stability. To distinguish these possibilities, a recently described native elongating transcript sequencing (NET-seq)

approach was applied to E. coli, which could be used to globally profile the positions of elongating RNA polymerases and monitor the effect of the dCas9:sgRNA complex on transcription. In this NET-seq method, the CRISPRi system was transformed into an E. coli MG1655-derived strain that contained a FLAG-tagged RNAP. The CRISPRi contained an sgRNA (NT1) that binds to the mRFP coding region. In vitro immunopurification of the tagged RNAP followed by sequencing of the nascent transcripts associated with elongating RNAPs allowed for distinguishing the pause sites of the RNAP.

[00631] These experiments demonstrated that the sgRNA induced strong transcriptional pausing upstream of the sgRNA target locus (Figure 41A). The distance between the pause site and the target site is 19-bp, which is in perfect accordance with the previously reported ~18-bp distance between the nucleotide incorporation of RNAP and its front-edge. This finding is consistent with a mechanism of CRISPRi in which the transcription block is due to physical collision between the elongating RNAP and the dCas9:sgRNA complex (Figure 41B). Binding of the dCas9:sgRNA complex to the template strand had little repressive effect, suggesting that RNAP was able to read through the complex in this particular orientation. In this case, the sgRNA faces the RNAP, which might be unzipped by the helicase activity of RNAP. These experiments have demonstrated that CRISPRi utilizes RNAs to directly block transcription. This mechanism is distinct from that of RNAi, for which knockdown of gene expression requires the destruction of already transcribed messenger RNAs, prior to their translation.

[00632] **Figure 41** demonstrates that CRISPRi functions by blocking transcription elongation. (**A**) FLAG-tagged RNAP molecules were immunoprecipitated and the associated nascent mRNA transcripts were sequenced. The top panel shows sequencing results of the nascent mRFP transcript in cells without sgRNA, and the bottom panel shows results in cells with sgRNA. In the presence of sgRNA, a strong transcriptional pause was observed 19-bp upstream of the target site, after which the number of sequencing reads drops precipitously. (**B**) A proposed CRISPRi mechanism based on physical collision between RNAP and dCas9-sgRNA. The distance from the center of RNAP to its front edge is ~19 bp, which matches well with our measured distance between the transcription pause site and 3' of sgRNA basepairing region. The paused RNAP aborts transcription elongation upon encountering the dCas9-sgRNA roadblock.

CRISPRi sgRNA-guided gene silencing is highly specific

[00633] To evaluate the specificity of CRISPRi on a genome-wide scale, whole transcriptome shotgun sequencing (RNA-seq) of dCas9-transformed cells with and without sgRNA co-expression was performed (Figure 42A). In the presence of the sgRNA targeted to mRFP (NT1), the mRFP transcript was the sole gene exhibiting a decrease in abundance. No other genes showed significant change in expression upon addition of the sgRNA, within sequencing errors. We also

performed RNA-seq on cells with different sgRNAs that target different genes. None of these experiments showed significant changes of genes besides the targeted gene (Figure 49). Thus sgRNA-guided gene targeting and regulation is highly specific and does not have significant off-target effects.

[00634] **Figure 42** demonstrates the targeting specificity of the CRISPRi system. (**A**) Genome-scale mRNA sequencing (RNA-seq) confirmed that CRISPRi targeting has no off-target effects. The sgRNA NT1 that binds to the mRFP coding region was used. The dCas9, mRFP, and sfGFP genes are highlighted. (**B**) Multiple sgRNAs can independently silence two fluorescent protein reporters in the same cell. Each sgRNA specifically repressed its cognate gene but not the other gene. When both sgRNAs were present, both genes were silenced. Error bars represent SEM from at least three biological replicates. (**C**) Microscopic images for using two sgRNAs to control two fluorescent proteins. The top panel shows the bright-field images of the E. coli cells, the middle panel shows the RFP channel, and the bottom shows the GFP panel. Co-expression of one sgRNA and dCas9 only silences the cognate fluorescent protein but not the other. The knockdown effect was strong, as almost no fluorescence was observed from cells with certain fluorescent protein silenced. Scale bar, 10 μm. Control shows cells without any fluorescent protein reporters. Fluorescence results represent average and SEM of at least three biological replicates. See also Figure 49.

[00635] **Figure 49** is related to Figure 42A and depicts the RNA-seq data of cells with sgRNAs that target different genes. (**A**) (+/-) sgRNA that targets the promoter of the endogenous lacI gene in E. coli. The same lacI-targeting sgRNA was used as in Figure 44A. (**B**) (+/-) 1 mM IPTG for cells without auto-inhibited sgRNA (sgRNA repressed its own promoter). (**C**) (+/-) sgRNA that targets the endogenous lacZ gene in E. coli. The same lacZ-targeting sgRNA was used as in Figure 44A. 1 mM IPTG was also supplemented to cells with the lacZ-targeting sgRNA.

CRISPRi can be used to simultaneously regulate multiple genes

[00636] The CRISPRi system can allow control of multiple genes independently without crosstalk. A dual-color fluorescence-reporter system based on mRFP and sfGFP was devised. Two sgRNAs with distinct complementary regions to each gene were designed. Expression of each sgRNA only silenced the cognate gene and had no effect on the other. Co-expression of two sgRNAs knocked down both genes (Figure 42B&C). These results suggest that the sgRNA-guided targeting is specific, with the specificity dictated by its sequence identity, and not impacted by the presence of other sgRNAs. This behavior should enable multiplex control of multiple genes simultaneously by CRISPRi.

Factors that determine CRISPRi silencing efficiency

[00637] To find determinants of CRISPRi targeting efficiency, the role of length, sequence complementarity and position on silencing efficiency was investigated (Figure 43A). As suggested in Figure 40C, the location of the sgRNA target sequence along the gene was important for efficiency. sgRNAs were further designed to cover the full length of the coding regions for both mRFP and sfGFP (Supplemental Data for sgRNA sequences). In all cases, repression was inversely correlated with the target distance from the transcription start site (Figure 43B). A strong linear correlation was observed for mRFP. A similar, but slightly weaker correlation was observed when sfGFP was used as the target, perhaps indicating varying kinetics of the RNA polymerase during different points in elongation of this gene.

[00638] The sgRNA contains a 20-bp region complementary to the target. To identify the importance of this basepairing region, the length of sgRNA NT1 was altered (Figure 43C). While extension of the region from the 5' end did not affect silencing, truncation of the region severely decreased repression. The minimal length of the basepairing region needed for gene silencing was 12bp, with further truncation leading to complete loss of function. Single mutations were introduced into the basepairing region of sgRNA NT1 and the overall effect on silencing was tested. From the results, three sub-regions could be discerned, each with a distinct contribution to the overall binding and silencing (Figure 43D). Any single mutation of the first 7 nucleotides dramatically decreased repression, suggesting this sequence constitutes a "seed region" for binding, as noted previously for both the type I and type II CRISPR systems . Adjacent nucleotides were also mutated in pairs (Figure 43E and Figure 50). In most cases, the relative repression activity due to a double mutation was multiplicative, relative to the effects of the single mutants, suggesting an independent relationship between the mismatches. Furthermore, in agreement with previous results on the importance of the PAM sequence, an incorrect PAM totally abolished silencing even with a 20-bp perfect binding region (Figure 43E). Thus, the specificity of the CRISPRi system is determined jointly by the PAM (2-bp) and at least a 12-bp sgRNA-DNA stretch, the space of which is large enough to cover most bacterial genomes for unique target sites.

[00639] Two sgRNAs both targeting the same gene were tested(Figure 43F and Figure 51). Depending on the relative positioning of multiple sgRNAs, distinct combinatorial effects were observed. Combining two sgRNAs, each with about 300-fold repression, allowed for increased overall silencing up to a thousand-fold. Combining two weaker sgRNAs (~5-fold) showed multiplicative effects when used together. Suppressive combinatorial effects were observed when using two sgRNAs whose targets overlapped. This was probably due to competition of both sgRNAs for binding to the same region.

[00640] Figure 43 depicts the characterization of factors that affect silencing efficiency. (**A**) The silencing effects were measured of sgRNAs with different targeting loci on the same gene (distance from the translation start codon) and sgRNAs with different lengths of the basepairing region to the same target locus (based on NT1). (**B**) The silencing efficiency was inversely correlated with the target distance from the translation start codon (orange - mRFP & green - sfGFP). The relative repression activity was calculated by normalizing repression of each sgRNA to that of the sgRNA with the highest repression fold change. Error bars represent SEM from three biological replicates. (**C**) The length of the Watson-Crick basepairing region between the sgRNA and the target DNA affects repression efficiency. Extensions of the basepairing region all exhibited strong silencing effect, and truncations dramatically decreased repression. The minimal length of the basepairing region for detectable repression is 12-bp. Error bars represent SEM from three biological replicates. (**D**) Single mismatches were introduced into every nucleotide on sgRNA (NT1, Figure 40B) how these single mismatches affected repression efficiency was measured. Three sub-regions with distinct importance to the overall silencing can be discerned. They show a step function. The first 7-nuleotide region was critical for silencing, and likely constitutes a "seed" region for probing sgRNAs binding to the DNA target. The PAM sequence (NGG) was indispensible for silencing. Error bars represent SEM from three biological replicates. (**E**) Silencing effects of sgRNAs with adjacent double mismatches. The relative repression activity of single-mismatched sgRNAs is shown with the mismatch position labeled on the bottom. Experimentally measured activity of double-mismatched sgRNAs is shown. Calculated activity by multiplying the effects of two single-mismatched sgRNAs is shown in white and labeled with "Com." In most cases, the silencing activity of a double-mismatched sgRNA was simply a multiplication of the activities of single-mismatched sgRNAs (except Figure 50B), suggesting an independent relationship between single mismatches. Error bars represent SEM from three biological replicates. (**F**) Combinatorial silencing effects of using double sgRNAs to target a single mRFP gene. Using two sgRNAs that target the same gene, the overall knockdown effect can be improved to almost 1,000-fold. When two sgRNAs bind to non-overlapping sequences of the same gene, repression was augmented. When two sgRNAs target overlapping regions, repression was suppressed. Error bars represent SEM from three biological replicates.

[00641] Figure 50 is related to Figure 43E and depicts the silencing effects of sgRNAs with adjacent double mismatches. The relative repression activity of single-mismatched sgRNAs is shown with the mismatch position labeled on the bottom. Experimentally measured activity of double-mismatched sgRNAs is also shown. Activity calculated by multiplying the effects of two single-

mismatched sgRNAs is shown in white and labeled with "Com". Fluorescence results represent average and SEM of three biological replicates.

[00642] **Figure 51** is related to Figure 43F and depicts the combinatorial silencing effects of using two sgRNAs to regulate a single gene. In all cases, non-overlapping sgRNAs showed augmentative silencing effects, and overlapping sgRNAs showed suppressive effects. The combinatorial effect was independent of whether the sgRNA was targeting the template or non-template DNA strands. Fluorescence results represent average and SEM of three biological replicates.

Interrogating an endogenous regulatory network using CRISPRi gene knockdown

[00643] The CRISPRi system was next used as a gene knockdown platform to interrogate endogenous gene networks. Previous methods to interrogate microbial gene networks have mostly relied on laborious and costly genomic engineering and knockout procedures. By contrast, gene knockdown with CRISPRi requires only the design and synthesis of a small sgRNA bearing a 20-bp complementary region to the desired genes. To demonstrate this, CRISPRi was used to create E. coli knockdown strains by designing sgRNAs to systematically perturb genes that were part of the well-characterized E. coli lactose regulatory pathway (Figure 44A). β-galactosidase assays were performed to measure LacZ expression from the knockdown strains, with and without Isopropyl β-D-1-thiogalactopyranoside (IPTG), a chemical that inhibits the lac repressor (LacI). In wild-type cells, addition of IPTG induced LacZ expression. The results showed that a lacZ-specific sgRNA could strongly repress LacZ expression (Figure 44B). Conversely, an sgRNA targeting the lacI gene led to activation of LacZ expression, even in the absence of IPTG, as would be expected for silencing a direct repressor of LacZ expression.

[00644] It is known that cAMP-CRP is an essential activator of LacZ expression by binding to a cis regulatory site upstream of the promoter (A site). Consistently, the sgRNA that was targeted to the crp gene or to the A site in the LacZ promoter led to repression, demonstrating a means to link a regulator to its cis-regulatory sequence using CRISPRi experiments. Targeting the adenylate cylase gene (cya), which is necessary to produce the cAMP that makes CRP more effective at the LacZ promoter, only led to partial repression. Addition of 1mM cAMP to the growth media complemented the effects for cya knockdown but not for crp knockdown, suggesting that cya is an indirect regulator of LacZ. Furthermore, targeting the LacI cis-regulatory site (O site) with an sgRNA led to inhibition, presumably because Cas9 complex binding at this site sterically blocks RNA polymerase, mimicking the behavior of the LacI transcription repressor. Targeting the known RNAP binding site (P site) also blocked expression. In summary, these studies demonstrate that the CRISPRi-based gene knockdown method provides a rapid and effective approach for interrogating the regulatory functions (activating or repressing) of genes and cis elements in a complex regulatory network (Figure 44C).

[00645] **Figure 44** demonstrates functional profiling of a complex regulatory network using CRISPRi gene knockdown. (**A**) sgRNAs were designed and used to knock down genes (cya, crp, lacI, lacZ, lacY, lacA) in the lac regulatory pathway or block transcriptional operator sites (A/P/O). LacI is a repressor of the lacZYA operon by binding to a transcription operator site (O site). The lacZ gene encodes an enzyme that catalyzes lactose into glucose. A few trans-acting host genes such as cya and crp are involved in the activation of the lacZYA system. The cAMP-CRP complex binds to a transcription operator site (A site) and recruits RNA polymerase binding to the P site, which initiates transcription of lacZYA. IPTG, a chemical that inhibits LacI function, induces LacZ expression. (**B**) β-galactosidase assay of the knockdown strains without (white) and with (grey) IPTG. Control shows that the wild-type cells without CRISPRi perturbation could be induced by addition of IPTG. The sgRNA that targets LacZ strongly repressed LacZ expression, even in the presence of IPTG. When LacI was targeted, LacZ expression was high, even without IPTG. Targeting cya and crp genes led to decreased LacZ expression level in the presence of IPTG. Presence of 1mM cAMP rescued cya knockdown but not crp knockdown. Blocking the transcription operator sites resulted in LacZ repression, suggesting that these are important cis-acting regulatory sites for LacZ. Upon perturbation, decreased (down arrows) and increased (up arrows) expression of LacZ is indicated. Error bars represent SEM from three biological replicates. (**C**) The knockdown experiments allowed us to profile the roles of regulators in the lac regulatory circuit. The data is shown on a 2-D graph, with x-axis showing LacZ activity without IPTG and y-axis showing its activity with IPTG. The spreading of ovals along each axis shows the standard deviations. The β-galactosidase assay results represent average and SEM of three biological replicates. For RNA-seq data on LacI and LacZ targeting, see also Figure 49.

CRISPRi can knock down targeted gene expression in human cells

[00646] To test the generality of the CRISPRi approach for using the dCas9-sgRNA complex to repress transcription, the system was tested in HEK293 mammalian cells. The dCas9 protein was codon optimized, fused to three copies of a nuclear localization sequence (NLS), and expressed from a Murine Stem Cell Virus (MSCV) retroviral vector. The same sgRNA design shown in Figure 40B was used to express from the RNA polymerase III U6 promoter. A reporter HEK293 cell line expressing EGFP under the SV40 promoter was created by viral infection. Using an sgRNA (eNT2) that targeted the non-template DNA strand of the EGFP coding region, a moderate but reproducible knockdown of gene expression was observed (46% repression, Figure 45A). The repression was dependent on both the dCas9 protein and sgRNA, implying that repression was due to the dCas9-sgRNA complex and RNA-guided targeting. The same sgRNA exhibited better repression on the same gene when transiently expressed from a plasmid (63% repression, Figure

52). Consistent with the bacterial system, only sgRNAs targeted to the non-template strand exhibited repression. Factors such as the distance from the transcription start and the local chromatin state may be critical parameters determining repression efficiency (Figure 52). Optimization of dCas9 and sgRNA expression, stability, nuclear localization, and interaction will allow for further improvement of CRISPRi efficiency in mammalian cells.

[00647] Figure 45 demonstrates that CRISPRi can repress gene expression in human cells. (**A**) A CRISPRi system in HEK293 cells. The SV40-EGFP expression cassette was inserted into the genome via retroviral infection. The dCas9 protein was codon-optimized and fused with three copies of NLS sequence. The sgRNA was expressed from an RNA polymerase III U6 vector. Co-transfection of dCas9 and an sgRNA (eNT2) that targets the non-template strand of EGFP decreased fluorescence (~46%) while the expression of either dCas9 or sgRNA alone showe no effect. (**B**) The dCas9:sgRNA-mediated repression was dependent on the target loci. Seven sgRNAs were designed to target different regions of the EGFP coding sequence on the template or non-template strand. Only eNT2 and eNT5 showed moderate repression. Fluorescence results from 7A and 7B represent average and error of two biological replicates.

[00648] Figure 52 is related to Figure 45 and shows that sgRNA repression is dependent on the target loci and relatively distance from the transcription start. The same sgRNA was used to repress the same EGFP gene with different promoters. Cas9/sgRNA complexes repressed transcription from transiently transfected plasmid DNA. The level of transcriptional repression was slightly better (63%) than that observed for genomic genes, and the percentage of GFP-negative cells increased in the presence of sgRNA. The target locus has different distance from the transcription start. While SV40-EGFP showed repression, LTR-EGFP had no effect. Fluorescence results represent average and error of two biological replicates.

CRISPRi efficiently and selectively represses transcription of target genes

[00649] The CRISPRi system is a relatively simple platform for targeted gene regulation. CRISPRi does not rely on the presence of complex host factors, but instead only requires the dCas9 protein and guide RNAs, and thus is flexible and highly designable. The system can efficiently silence genes in bacteria. The silencing is very efficient, as no off-target effects were detected. Furthermore, the efficiency of the knockdown can be tuned by changing the target loci and the degree of basepairing between the sgRNA and the target gene. This will make it possible to create allelic series of hypomorphs -- a feature that is especially useful for the study of essential genes. The system functions by directly blocking transcription, in a manner that can be easily programmed by designing sgRNAs. Mechanistically, this is distinct from RNAi-based silencing, which requires the destruction of already transcribed mRNAs.

[00650] In addition, these dCas9:sgRNA complexes can also modulate transcription by targeting key cis-acting motifs within any promoter, sterically blocking the association of their cognate trans-acting transcription factors. Thus, in addition to its use as a gene knockdown tool, CRISPRi could be used for functional mapping of promoters and other genomic regulatory modules.

CRISPRi is amenable to genome-scale analysis and regulation

[00651] The CRISPRi method is based on the use of DNA-targeting RNAs, and only the DNA targeting segment needs to be designed for specific gene targets. With the advances of large-scale DNA oligonucleotide synthesis technology, generating large sets of oligonucleotides that contain unique 20-bp regions for genome targeting is fast and inexpensive. These oligonucleotide libraries could allow us to target large numbers of individual genes to infer gene function or to target gene pairs to map genetic interactions. Furthermore, CRISPRi could be used to simultaneously modulate the expression of large sets of genes, as the small size of sgRNAs allows one to concatenate multiple elements into the same expression vector.

CRISPRi provides new tools for manipulating microbial genomes

[00652] Because the CRISPRi platform is compact and self-contained, it can be adapted for different organisms. CRISPRi is a powerful tool for studying non-model organisms for which genetic engineering methods are not well developed, including pathogens or industrially useful organisms. Unlike most eukaryotes, most bacteria lack the RNAi machinery. As a consequence, regulation of endogenous genes using designed synthetic RNAs is currently limited. CRISPRi could provide an RNAi-like method for gene perturbation in microbes.

CRISPRi as a platform for engineering transcriptional regulatory networks

[00653] The CRISPRi can be utilized as a flexible framework for engineering transcriptional regulatory networks. The CRISPRi platform, because it is essentially a RNA-guided DNA-binding complex, also provides a flexible scaffold for directing diverse regulatory machinery to specific sites in the genome. Beyond simply blocking transcription of target genes, it is possible to couple the dCas9 protein with numerous regulatory domains to modulate different biological processes and to generate different functional outcomes (e.g transcriptional activation, chromatin modifications).

[00654] In the CRISPRi system, it is possible to link multiple sgRNAs into transcriptional circuits in which one upstream sgRNA controls the expression of a different downstream sgRNA. As RNA molecules in microorganisms tend to be short-lived, we suspect that the genetic programs regulated by sgRNAs might show rapid kinetics distinct from circuits that involve slow processes such as protein expression and degradation. In summary, the CRISPRi system is a general genetic programming platform suitable for a variety of biomedical research and clinical

applications including genome-scale functional profiling, microbial metabolic engineering, and cell reprogramming.

Example 5: A chimeric site-directed polypeptide can be used to modulate (activate or repress) transcription in human cells.

[00655] We have demonstrated that in human cells, a fusion protein comprisng a catalytically inactive Cas9 and an an activator domain or a repressor domain can increase or decrease transcription from a target DNA, respectively.

[00656] Figure 55. We fused the humanized catalytically inactive Cas9 with a transcription activator domain VP64. (**A**) To test the efficiency for gene activation using this system, we inserted a GAL4 UAS inducible promoter that controls a GFP into the HEK293 (human tissue culture cells) genome. (**B**) The GAL4 UAS promoter can be induced in the presence of yeast-derived protein GAL4. The dCas9-VP64 fusion can effectively activate GAL4 UAS by 20-fold in the presence of a cognate guide RNA that binds to the GAL4 UAS region. (**C**) Microscopic images for dCas9-VP64 activation. (**D**) Flow cytometry data for dCas9-VP64 activation.

[00657] Figure 56 We fused the humanized catalytically inactive Cas9 with a transcription repressor domain KRAB. (**Top**) We designed 10 guide RNAs that target a well-characterized promoter SV40 early promoter and one guide RNA that targets the EGFP coding region. (**Bottom**) Using a non-chimeric dCas9, we observed 2~3 fold repression for gRNAs of P9 and NT2. This efficiency was greatly improved using the dCas9-KRAB fusion. For example, with dCas9-KRAB fusion, P9 and NT2 showed 20-fold and 15-fold repression, respectively. In addition P1-P6 showed a significant reduction in expression when the fusion protein was used, but limited repression when a non-chimeric dCas9 was used.

Example 6: Cas9 can use artificial guide RNAs, not existing in nature, to perform target DNA cleavage

[00658] An artificial crRNA and an artificial tracrRNA were designed basesd on the protein-binding segment of naturally occurring transcripts of S. pyogenes crRNA and tracrRNAs, modified to mimic the asymmetric bulge within natural S. pyogenes crRNA:tracrRNA duplex (see the bulge in the protein-binding domain of both the artificial (top) and natural (bottom) RNA molecules depicted in **Figure 57A**). The artificial tracrRNA sequence shares less than 50% identity with the natural tracrRNA. The predicted secondary structure of the crRNA:tracrRNA protein-binding duplex is the same for both RNA pairs, but the predicted structure of the rest of the RNAs is much different.

[00659] **Figure 57** demonstrates that artificial sequences that share very little (roughly 50% identity) with naturally occurring a tracrRNAs and crRNAs can function with Cas9 to cleave target DNA as long as the structure of the protein-binding domain of the DNA-targeting RNA is conserved. (**A**) Co-folding of *S. pyogenes* tracrRNA and crRNA and artificial tracrRNA and crRNA. (**B**) Combinations of *S. pyogenes* Cas9 and tracrRNA:crRNA orthologs were used to perform plasmid DNA cleavage assays .Spy – *S. pyogenes*, Lin - *L. innocua*, Nme – N. *meningitidis*, Pmu – *P. multocida*. *S. pyogenes* Cas9 can be guided by some, but not all tracrRNA:crRNA orthologs naturally occuring in selected bacterial species. Notably, *S. pyogenes* Cas9 can be guided by the artificial tracrRNA:crRNA pair, which was designed based on the structure of the protein-binding segment of the naturally occurring DNA-targeting RNA using sequence completely unrelated to the CRISPR system.

[00660] The artificial "tracrRNA" (activator RNA) used was 5'-GUUUUCCCUUUUCAAAGAAAUCUCCUGGGCACCUAUCUUCUUAGGUGCCCUCCCUUGUUUAAACCUGACCAGUUAACCGGCUGGUUAGGUUUUU-3' (SEQ ID NO:1347). The artificial "crRNA" (targeter RNA) used was: 5'- GAGAUUUAUGAAAAGGGAAAAC-3' (SEQ ID NO:1348).

Example 7: Generation of non-human Transgenic Organisms

[00661] A transgenic mouse expressing Cas9 (either unmodified, modified to have reduced enzymatic activity, modified as a fusion protein for any of the purposes outline above) is generated using a convenient method known to one of ordinary skill in the art (e.g., (i) gene knock-in at a targeted locus (e.g., ROSA 26) of a mouse embryonic stem cell (ES cell) followed by blastocyst injection and the generation of chimeric mice; (ii) injection of a randomly integrating transgene into the pronucleus of a fertilized mouse oocyte followed by ovum implantation into a pseudopregnant female; etc.). The Cas9 protein is under the control of a promoter that expresses at least in embryonic stem cells, and may be additionally under temporal or tissue-specific control (e.g., drug inducible, controlled by a Cre/Lox based promoter system, etc.). Once a line of transgenic Cas9 expressing mice is generated, embryonic stem cells are isolated and cultured and in some cases ES cells are frozen for future use. Because the isolated ES cells express Cas9 (and in some cases the expression is under temporal control (e.g., drug inducible), new knock-out or knock-in cells (and therefore mice) are rapidly generated at any desired locus in the genome by introducing an appropriately designed DNA-targeting RNA that targets the Cas9 to a particular locus of choice. Such a system, and many variations thereof, is used to generate new genetically modified organisms at any locus of choice. When modified Cas9 is used to modulate transcription and/or modify DNA and/or modify polypeptides associated with DNA, the ES cells

themselves (or any differentiated cells derived from the ES cells (e.g., an entire mouse, a differentiated cell line, etc.) are used to study to properties of any gene of choice (or any expression product of choice, or any genomic locus of choice) simply by introducing an appropriate DNA-targeting RNA into a desired Cas9 expressing cell.

[00662] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

CLAIMS

1. A method of modifying a target DNA, the method comprising contacting the target DNA with a complex comprising:

(a) a Cas9 polypeptide and

(b) a single-molecule DNA-targeting RNA comprising:

(i) a DNA-targeting segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA, and

(ii) a protein-binding segment that interacts with said Cas9 polypeptide, wherein the protein-binding segment comprises two complementary stretches of nucleotides that hybridize to form a double stranded RNA (dsRNA) duplex, wherein said dsRNA duplex comprises complementary nucleotides of a tracrRNA and a CRISPR RNA (crRNA),

wherein said two complementary stretches of nucleotides are covalently linked by intervening nucleotides,

wherein said contacting is *in vitro* or in a cell *ex vivo*; and

wherein said modifying is cleavage of the target DNA.

2. The method of claim 1, wherein said dsRNA duplex has a length of from 8 base pairs (bp) to 30 bp.

3. The method of claim 1 or 2, wherein the percent complementarity between the nucleotides that hybridize to form the dsRNA duplex of the protein-binding segment is greater than 70%.

4. The method of claim 1, 2 or 3, wherein the target DNA is present in a bacterial cell, an archaeal cell, a single-cell eukaryotic organism, a plant cell, a cell from an invertebrate animal, or a cell from a vertebrate animal.

5. The method of claim 1, 2 or 3, wherein the target DNA is chromosomal DNA.

6. The method of any one of the preceding claims wherein the contacting comprises introducing into a cell (a) said Cas9 polypeptide, or a polynucleotide encoding said Cas9 polypeptide, and (b) said DNA-targeting RNA, or a DNA polynucleotide encoding said DNA-targeting RNA.

7. The method of claim 6 wherein the method further comprises introducing into the cell a donor polynucleotide.

8. The method according to any one of the preceding claims wherein a protein transduction domain is covalently linked to the amino terminus of the Cas9 polypeptide,

wherein said protein transduction domain facilitates the traversal of the Cas9 polypeptide from the cytosol to within an organelle of a cell.

9. The method according to any one of claims 1-7 wherein a protein transduction domain is covalently linked to the carboxyl terminus of the Cas9 polypeptide, wherein said protein transduction domain facilitates the traversal of the Cas9 polypeptide from the cytosol to within an organelle of a cell.

10. A composition comprising:

(a) a Cas9 polypeptide, or a polynucleotide encoding said Cas9 polypeptide, and

(b) a single-molecule DNA-targeting RNA, or a DNA polynucleotide encoding said DNA-targeting RNA, wherein said single-molecule DNA-targeting RNA comprises:

(i) a DNA-targeting segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA, and

(ii) a protein-binding segment that interacts with said Cas9 polypeptide, wherein the protein-binding segment comprises two complementary stretches of nucleotides that hybridize to form a double stranded RNA (dsRNA) duplex, wherein said dsRNA duplex comprises complementary nucleotides of a tracrRNA and a CRISPR RNA (crRNA), and

wherein said two complementary stretches of nucleotides are covalently linked by intervening nucleotides.

11. The composition of claim 10, wherein said dsRNA duplex has a length of from 8 base pairs (bp) to 30 bp.

12. The composition of claim 10 or 11, wherein the percent complementarity between the nucleotides that hybridize to form the dsRNA duplex of the protein-binding segment is greater than 70%.

13. The composition of claim 10, 11 or 12, wherein the target DNA is present in a bacterial cell, an archaeal cell, a single-cell eukaryotic organism, a plant cell, a cell from an invertebrate animal, or a cell from a vertebrate animal.

14. The composition of claim 10, 11 or 12, wherein the target DNA is chromosomal DNA.

15. The composition of any one of claims 10-14, wherein a protein transduction domain is covalently linked to the amino terminus of the Cas9 polypeptide, wherein said protein transduction domain facilitates the traversal of the Cas9 polypeptide from the cytosol to within an organelle of a cell.

16. The composition of any one of claims 10-14, wherein a protein transduction domain is covalently linked to the carboxyl terminus of the Cas9 polypeptide, wherein said protein transduction domain facilitates the traversal of the Cas9 polypeptide from the cytosol to within an organelle of a cell.

17. The composition of any one of claims 10-16, wherein the Cas9 polypeptide and the single-molecule DNA-targeting RNA are in a cell *in vitro* or *ex vivo*.

18. A single-molecule DNA-targeting RNA, or a DNA polynucleotide encoding said DNA-targeting RNA, wherein the single-molecule DNA-targeting RNA comprises:

(a) a DNA-targeting segment comprising a nucleotide sequence that is complementary to a target sequence in a target DNA, and

(b) a protein-binding segment that interacts with a Cas9 protein, wherein the protein-binding segment comprises two complementary stretches of nucleotides that hybridize to form a double stranded RNA (dsRNA) duplex, wherein said dsRNA duplex comprises complementary nucleotides of a tracrRNA and a CRISPR RNA (crRNA), and wherein said two complementary stretches of nucleotides are covalently linked by intervening nucleotides.

19. The single-molecule DNA-targeting RNA of claim 18 wherein said dsRNA duplex has a length of from 8 base pairs (bp) to 30 bp.

20. The single-molecule DNA-targeting RNA of claim 18 or 19, wherein the percent complementarity between the nucleotides that hybridize to form the dsRNA duplex of the protein-binding segment is greater than 70%.

21. The single-molecule DNA-targeting RNA of claim 18, 19 or 20, wherein the target DNA is present in a bacterial cell, an archaeal cell, a single-cell eukaryotic organism, a plant cell, a cell from an invertebrate animal, or a cell from a vertebrate animal.

22. The single-molecule DNA-targeting RNA of claim 18, 19 or 20, wherein the target DNA is chromosomal DNA.

23. One or more nucleic acids comprising:

(a) a first nucleotide sequence encoding a single-molecule DNA-targeting RNA comprising:

(i) a DNA-targeting segment comprising a nucleotide sequence that is complementary to a target sequence in a target DNA, and

(ii) a protein-binding segment that interacts with a Cas9 polypeptide, wherein the protein-binding segment comprises two complementary stretches of nucleotides that hybridize to form a double stranded RNA (dsRNA) duplex, wherein said dsRNA duplex comprises complementary nucleotides of a

tracrRNA and a CRISPR RNA (crRNA), and wherein said two complementary stretches of nucleotides are covalently linked by intervening nucleotides;

wherein the first nucleotide sequence encoding said DNA-targeting RNA is operably linked to a promoter; and, optionally,

(b) a second nucleotide sequence encoding a Cas9 polypeptide, wherein the nucleotide sequence encoding said Cas9 polypeptide is operably linked to a promoter.

24. The one or more nucleic acids of claim 23, wherein said nucleic acids are one or more recombinant expression vectors.

25. The one or more nucleic acids of claim 23 or 24, wherein said dsRNA duplex has a length of from 8 base pairs (bp) to 30 bp.

26. The one or more nucleic acids of claim 23, 24 or 25, wherein the percent complementarity between the nucleotides that hybridize to form the dsRNA duplex of the protein-binding segment is greater than 70%.

27. The one or more nucleic acids of any one of claims 23-26, wherein the target DNA is present in a bacterial cell, an archaeal cell, a single-cell eukaryotic organism, a plant cell, a cell from an invertebrate animal, or a cell from a vertebrate animal.

28. The one or more nucleic acids of any one of claims 23-26, wherein the target DNA is chromosomal DNA.

29. A kit comprising

(a) a Cas9 polypeptide, or a nucleic acid comprising a nucleotide sequence encoding said Cas9 polypeptide; and

(b) a single-molecule DNA-targeting RNA, or a nucleic acid comprising a nucleotide sequence encoding said single-molecule DNA-targeting RNA, wherein the single-molecule DNA-targeting RNA comprises:

(i) a DNA-targeting segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA, and

(ii) a protein-binding segment that interacts with said Cas9 polypeptide, wherein the protein-binding segment comprises two complementary stretches of nucleotides that hybridize to form a double stranded RNA (dsRNA) duplex, wherein said dsRNA duplex comprises complementary nucleotides of a tracrRNA and a CRISPR RNA (crRNA), and,

wherein said two complementary stretches of nucleotides are covalently linked by intervening nucleotides, and

wherein (a) and (b) are in the same or separate containers.

30. The kit of claim 29, wherein the target DNA is present in a bacterial cell, an archaeal cell, a single-cell eukaryotic organism, a plant cell, a cell from an invertebrate animal, or a cell from a vertebrate animal.

31. The kit of claim 29, wherein the target DNA is chromosomal DNA.

32. The composition of any one of claims 10-17, or the single-molecule DNA-targeting RNA of any one of claims 18-22, or the one or more nucleic acids of any one of claims 23-28, for use in a method of therapeutic treatment of a patient.