Abstract:
The invention is directed to a method of mutating one or more target nucleic acid sequences in a stem cell or a zygote comprising introducing into the stem cell or zygote (i) ribonucleic acid (RNA) sequences that comprise a portion that is comple-
mentary to a portion of each of the target nucleic acid sequences and comprise a binding site for a CRISPR associated (Cas) protein; and a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity. The stem cell or zygote is maintained under conditions in which the target nucleic acid sequences are mutated in the stem cell or zygote. The invention is also directed to methods of producing a non human mammal carrying mutations and methods of modulating the expression and/or activ-
ity target nucleic acid sequences and cells or zygo
tes.
METHODS OF MUTATING, MODIFYING OR MODULATING NUCLEIC ACID IN A CELL OR NONHUMAN MAMMAL

RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/812,720, filed on April 16, 2013; U.S. Provisional Application No. 61/824,920, filed on May 17, 2013; U.S. Provisional Application No. 61/858,437, filed on July 25, 2013; and U.S. Provisional Application No. 61/865,888, filed on August 14, 2013.

[0002] The entire teachings of the above applications are incorporated herein by reference.

GOVERNMENT SUPPORT

[0003] This invention was made with government support under HD 045022 and R37CA084198 from the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0004] Genetically modified mice represent a crucial tool for understanding gene function in development and disease. Mutant mice are conventionally generated by insertional mutagenesis (Copeland and Jenkins, 2010; Kool and Berns, 2009) or by gene targeting methods (Capecchi, 2005). In conventional gene targeting methods, mutations are introduced through homologous recombination in mouse embryonic stem (ES) cells. Targeted ES cells injected into wild-type blastocysts can contribute to the germline of chimeric animals, generating mice containing the targeted gene modification (Capecchi, 2005). It is costly and time-consuming to produce single gene knockout mice, and even more so to make double mutant mice. Moreover, in most other mammalian species no established ES cell lines are available that contribute efficiently to chimeric animals, which greatly limits the genetic studies in many species.
Alternative methods have been developed to accelerate the process of genome modification by directly injecting DNA or mRNA of site-specific nucleases into the one cell embryo to generate DNA double strand break (DSB) at a specified locus in various species (Bogdanove and Voytas, 2011; Carroll et al., 2008; Urnov et al., 2010). DSBs induced by these site-specific nucleases can then be repaired by either error-prone non-homologous end joining (NHEJ) resulting in mutant mice and rats carrying deletions or insertions at the cut site (Carbery et al., 2010; Geurts et al., 2009; Sung et al., 2013; Tesson et al., 2011). If a donor plasmid with homology to the ends flanking the DSB is co-injected, high-fidelity homologous recombination can produce animals with targeted integrations (Cui et al., 2011; Meyer et al., 2010). Because these methods require the complex designs of zinc finger nucleases (ZNFs) or Transcription activator-like effector nucleases (TALENs) for each target gene and because the efficiency of targeting may vary substantially, no multiplexed gene targeting has been reported to date.

Thus, improved methods for producing genetically modified non-human mammals, such as mice, are needed.

SUMMARY OF THE INVENTION

Described herein is the use of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated (Cas) proteins (CRISPR/Cas) system to drive both non-homologous end-joining (NHEJ) based gene disruption and homology directed repair (HDR) based precise gene editing to achieve highly efficient and simultaneous targeting of multiple nucleic acid sequences in cells and nonhuman mammals.

Accordingly, in one aspect, the invention is directed to a method of mutating one or more target nucleic acid sequences in a (one or more) stem cell or a zygote comprising introducing into the stem cell or zygote (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associated (Cas) protein; and a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity. The stem cell or zygote is maintained under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid
sequences, and the Cas protein cleaves each of the one or more target nucleic acid sequences upon hybridization of the one or more RNA sequences to the portion of the target nucleic acid sequence, thereby mutating one or more target nucleic acid sequences in the stem cell or zygote.

[0009] In some aspects, the invention is directed to a method of producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences comprising introducing into a zygote or an embryo (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associated (Cas) protein; and ii) a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity. The zygote or the embryo is maintained under conditions in which RNA hybridizes to the portion of each of the one or more target nucleic acid sequences, and the Cas protein cleaves each of the one or more target nucleic acid sequences upon hybridization of the RNA to the portion of the target nucleic acid sequence, thereby producing an embryo having one or more mutated nucleic acid sequences. The embryo having one or more mutated nucleic acid sequences may be transferred into a foster nonhuman mammalian mother. The foster nonhuman mammalian mother is maintained under conditions in which one or more offspring carrying the one or more mutated nucleic acid sequences are produced, thereby producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences.

[0010] In some aspects, the invention is directed to a method of modulating the expression and/or activity of one or more target nucleic acid sequences in one or more cells or zygotes comprising introducing into the cell or zygote (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associate (Cas) protein; (ii) a Cas nucleic acid sequence or a variant thereof that encodes the Cas protein that targets but does not cleave the target nucleic acid sequence; and (iii) an effector domain. The method further comprises maintaining the cell under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences, the Cas protein binds to each of the one or more RNA sequences and the effector domain modulates the expression and/or activity of the target nucleic acid, thereby
modulating the expression and/or activity of the one or more target nucleic acid sequences in the cell or zygote.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0012] Figures 1A-1E show multiplexed gene targeting in mESCs. Figure 1A shows a schematic of the Cas9/sgRNA targeting sites in Tet1, 2, and 3. The sgRNA targeting sequence is underlined, and the protospacer-adjacent motif (PAM) sequence is labeled in green. The restriction sites at the target regions are bold and capitalized. Restriction enzymes used for restriction fragment length polymorphism (RFLP) and Southern blot analysis are shown, and the Southern blot probes are shown as orange boxes (SEQ ID NOs: 27-29). Figure 1B shows a surveyor assay for Cas9-mediated cleavage at Tet1, 2, 3 loci in mESCs. Figure 1C shows the genotyping of triple targeted mESCs, clone #51, #52, and #53 are shown. The upper panels in Figure 1C is the RFLP analysis. Tet1 PCR products were digested with Sacl, Tet2 PCR products were digested with EcoRV, and Tet3 PCR products were digested with Xhol. Lower panel, Southern blot analysis. For the Tet1 locus, Sacl digested genomic DNA was hybridized with a 5' probe. Expected fragment size: WT (wild type) = 5.8 kb, TM (Targeted mutation) = 6.4 kb. For the Tet2 locus, Sacl and EcoRV double digested genomic DNA was hybridized with a 3' probe. Expected fragment size: WT = 4.3kb, TM = 5.6kb. For the Tet3 locus, BamHI and Xhol double digested genomic DNA was hybridized with a 5' probe. Expected fragment size: WT = 3.2kb, TM = 8.1kb. Figure 1D shows the sequence of six mutant alleles in triple targeted mESC clone #14 and #41 (SEQ ID NOs: 30-45). Protospacer-Adjacent Motif (PAM) sequence is labeled in red. Figure 1E shows the analysis of 5hmC levels in DNA isolated from triple targeted mES clones by dot blot assay using anti-5hmC antibody. A previously characterized DKO clone derived using traditional method is used as a control. See also Figures 5A-D, Tables 1 and 4.

[0013] Figures 2A-2F show single and double gene targeting in vivo by injection into fertilized eggs. (2A) Genotyping of Tet1 single targeted mice. Figure 2B shows genotyping of Tet2 single targeted mice. RFLP analysis is shown in the
upper panel, and Southern blot analysis is shown in the lower panel of Figure 2B. Figure 2C is the sequence of both alleles of targeted gene in Tetl biallelic mutant mouse #2 and Tet2 biallelic mutant mouse #4 (SEQ ID NOs: 30, 32-34, 46, 47).

Figure 2D is the genotyping of Tetl/Tet2 double mutant mice. Analysis of mice #1 to #12 is shown. RFLP analysis is shown in the upper panel, and Southern blot analysis is shown in the lower panel of Figure 2D. The Tetl locus is displayed in the left panel and the Tet2 locus in the right panel. Figure 2E is the sequence of four mutant alleles from double mutant mouse #9 and #10 (SEQ ID NOs: 30, 33, 42, 46, 48-53). PAM sequences are labeled in red. Figure 2F is a picture of three-week-old double mutant mice. All RFLP and Southern digestsions and probes are the same as those used in Figures 1A-E. See also Figures 6A-6F, Tables 2 and 3.

[0014] Figures 3A-3C show multiplexed HR-mediated genome editing in vivo. Figure 3A shows a schematic of the oligo targeting sites at Tetl and Tet2 loci (SEQ ID NOs: 54-57). The sgRNA targeting sequence is underlined, and the PAM sequence is labeled in green. Oligo targeting each gene is shown under the target site, with 2bp changes labeled in red. Restriction enzyme sites used for RFLP analysis are bold and capitalized. Figure 3B is RFLP analysis of double oligo injection mice with HDR-mediated targeting at the Tetl and Tet2 loci. Figure 3C shows the sequences of both alleles of Tetl and Tet2 in mouse #5 and #7 show simultaneously HDR-mediated targeting at one allele or two alleles of each gene, and NHEJ-mediated disruption at the other alleles (SEQ ID NOs: 30, 33, 40, 58-61). See also Figures 8A-8C.

[0015] Figures 4A-4B show multiplexed genome editing in mES cells and mouse. Figure 4A is a diagram representing multiple gene targeting in mES cells. Figure 4B shows one step generation of mice with multiple mutations. Upper panel, multiple targeted mutations with random indels introduced through NHEJ. Lower panel, multiple predefined mutations introduced through HDR-mediated repair.

[0016] Figures 5A-5D show single, triple, and quintuple gene targeting in mES cells. Figure 5A is RFLP analysis of clones from each single targeting experiment (#1 to #17 are shown). Figure 5B is RFLP analysis of triple gene targeted clones (#37 to #53 are shown). Tetl PCR products were digested with Sacl, Tet2 PCR products were digested with EcoRV, and Tet3 PCR products were digested with Xhol. WT control is shown in the last lane. Genotyping of clone #51, #52, and #53
are also shown in Figure 1C. Figure 5C is a schematic of the Cas9/sgRNA targeting sites in Sry and Uty (SEQ ID NOs: 62-64). The sgRNA targeting sequence is underlined, and the protospacer-adjacent motif (PAM) sequence is labeled in green. The restriction sites at the target regions are bold and capitalized. Restriction enzymes used for RFLP analysis are shown. Figure 5D is RFLP analysis of quintuple gene targeted clones (#1 to #10 are shown). Sry PCR products were digested with BsaJI, Uty PCR products were digested with AvrII. WT control is shown in the last lane. RFLP analysis of Tetl, 2, 3 loci are not shown. Figures 5A-5D are related to Figure 1A-1E, Tables 1 and 4.

[0017] Figures 6A-6F show one step generation of single gene mutant mice by zygote injection. Figure 6A is RFLP analysis of blastocysts injected with different concentration of Cas9 mRNA and Tetl sgRNA at 20¼/µl. Tetl PCR products were digested with SacI. Figure 6B shows commonly recovered Tetl and Tet2 alleles resulted from MMEJ (SEQ ID NOs: 30, 33, 34, 40, 46, 52). PAM sequence of each targeting sequence is labeled in green. Microhomology flanking the DSB is bold and underlined in WT sequence. Figure 6C is RFLP analysis of eight Tet3 targeted blastocysts demonstrated high targeting efficiency (embryo #3 and #5 failed to amplify). Tet3 PCR products were digested with Xhol. Figure 6D is a picture of how some Tet3 targeted mice show smaller size and all homozygous mutants died within one day after birth. Figure 6E is RFLP analysis of Tet3 single targeted new born mice. Mouse #8 and #14 survived after birth. Sample #2 and #6 failed to amplify. Figure 6F are sequences of both Tet3 alleles of surviving Tet3 targeted mouse #14. PAM sequences are labeled in red. Figures 6A-6F are related to Figures 2A-2F and Table 2.

[0018] Figures 7A-7B show off target analysis of double mutant mice. Figure 7A shows three potential off targets of Tetl sgRNA and four potential off targets of Tet2 sgRNA are shown (SEQ ID NOs: 66-74). The 12bp perfect matching seed sequence is labeled in blue, and NGG PAM sequence is labeled in red. Figure 7B shows a surveyor assay of all seven potential off target loci in seven double mutant mice derived with high concentration of Cas9 mRNA (100ng/µl) injection. WT control is included as the eighth sample. The weak cleavage activity at UbrI locus is not due to off target effect, since sequences of these PCR products show no mutations. Figures 7A-7B are related to Figures 2A-2F and Table 5.
Figures 8A-8C show multiplexed precise HDR-mediated genome editing in vivo. Figure 8A is RFLP analysis of single oligo injection embryos with HDR-mediated targeting at Tet1 and Tet2 locus. Figure 8B is RFLP analysis of double oligo injection embryos with multiplexed HDR-mediated targeting at both Tet1 and Tet2 loci. Figure 8C shows the sequences of both alleles of Tet1 and Tet2 in embryo #2 and show simultaneously HDR-mediated targeting at one allele of both genes, and NHEJ-mediated gene disruption at the other allele of each gene (SEQ ID NOs: 30, 33, 53, 58, 59, 75). Figures 8A-8C are related to Figures 3A-3C.

Figure 9 shows 20 bp sequences of Tet1, Tet2, Tet3, Sry and Uty and the full length sequences of the RNA sequences (SEQ ID NOs: 76-85).

Figures 1OA-IOC: dCas9ta guided by sgRNA targeting tet binding site activates TetO promoter in HeLa cell. (10A) Schematic of dCas9ta fusion protein generated by mutation of two amino acids of Cas9 protein and fusion to 3x VP16 minimal transactivation domain (10B) Schematic of a TetO::tdTomato reporter system to test dCas9ta fusion protein. (IOC) Phase contrast, fluorescent microscopy and fluorescent activated cell sorting (FACS) profile of HeLa/TetO::tdTomato;EFla::NLSM2rtTA cells transfected by pmaxGFP (i), under dox exposure for 2 days (ii), transfected with dCas9ta without sgRNA (iii), and transfected with dCas9ta with sgRNA (sgTetO) complementary to tet binding sites (iv).

Figures 11A-1C dCas9ta guided by sgRNAs targeting Nanog promoter activates a NanogGFP reporter and endogenous Nanog gene in NIH3T3 cells. (11A) Schematic of the experiment showing the target sites of the sgRNAs relative to Nanog locus and the NanogGFP reporter. (11B) qRT-PCR analysis of endogenous Nanog. The fold change of (i) NanogGFP-only, (ii) NanogGFP + dCas9ta, (iii) NanogGFP + dCas9ta + sgmNanog were expressed relative to NanogGFP-only control. (11C) Microscope pictures of cells transfected with (i) NanogGFP-only, (ii) NanogGFP + dCas9ta, (iii) NanogGFP + dCas9ta + sgmNanog.

Figures 12A-12D: dCas9ta guided by sgRNA targeting tet binding site activates TetO promoter in NIH3T3 cell. Phase contrast, fluorescent microscopy and fluorescent activated cell sorting (FACS) profile of NIH3T3/TetO::tdTomato;EF1a::NLSM2rtTA cells transfected by pmaxGFP (12A), under dox exposure for 2 days (12B), transfected with dCas9ta without sgRNA
(12C), and transfected with dCas9ta with sgRNA complementary to tet binding sites (12D).

[0024] Figures 13A-13D: Microscope pictures and FACS analysis of HeLa/TetO::tdTomato cells (13A) No transfection. (13B) Transfected with dCas9ta + sgTetO. (13C) Transfected with dCas9Cdk9 + dCas9CycT + sgTetO. (13D) Transfected with dCas9ta + dCas9Cdk9 + dCas9CycT + sgTetO.

[0025] Figure 14: the wild type (Wt) Cas9 (S. pyogenes) nucleotide sequence (SEQ ID NO: 485).

[0026] Figure 15A: Alignment of HMG box sequences of Sry proteins from different mammalian species (SEQ ID NOs: 86-91). Position 94 (shown in red) is highly conserved in different species (h: human; m: mouse; c: Chimpanzee; pc: Pygmy Chimpanzee; g: gorilla; py: Pongo; hi: Hylabates; b: Baboon and cj: Calitrix jaccus). (From Shahid et al. BMC Medical Genetics 2010 11:131 doi: 10.1186/1471-2350-11-131).

[0027] Figure 15B: Genetic modification of Sry using TALENs. Schematic of Sry TALEN pair 2 and its recognition sequence within high mobility group (HMG) domain of Sry (SEQ ID NOs: 86-96). TAL repeats are color-coded to represent each of four repeat variable di-residues (RVDs); each RVD recognizes one corresponding DNA base (NI = A, NG = T, HD = C, NN = G). Nucleotides bound by TALENs are capitalized. Shown below are clones (targeted mutation [TM] alleles 1-3) where Sry deletions induced by TALENs were indicated in dashed lines. Srytm4 (540-bp deletion) and Srytm5 (440-bp deletion) clones are not shown.

[0028] Figures 16A-16F: One step generation of the Sox2-V5 allele. (16A) Schematic of the Cas9/sgRNA/oligo targeting site at the Sox2 stop codon (SEQ ID NOs: 97 and 98). The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. The stop codon of Sox2 is labeled in orange. The oligo contained 60bp homologies flanking the DSB. In the oligo donor sequence, the V5 tag sequence is labeled as a green box. PCR primers (SF, V5F, and SR) used for PCR genotyping are shown as red arrowheads. (16B) Upper panel, PCR genotyping using primers V5F and SR produced bands with correct size in targeted ES samples T1 to T5, but not in WT sample. Lower panel, PCR genotyping using primers SF and SR produced slightly larger products, indicating the 42bp V5 tag sequence was integrated. T1 only contain
larger product, suggesting either both alleles were targeted, or one allele failed to amplify. (16C) PCR products using primers SF and SR were cloned into plasmid and sequenced. Sequence across the targeting region confirmed correct fusion of V5 tag to the last codon of Sox2. (16D) Western blot analysis identified Sox2-V5 protein using V5 antibody in ES cells containing Sox2-V5 allele. Beta-actin was shown as the loading control. (16E) Immunostaining of targeted blastocyst using V5 antibody showed signal in ICM. Scale bar, 50µm. (16F) Immunostaining of targeted ES cells using V5 antibody showed uniform Sox2 expression. Scale bar, 100µm.

Figures 17A-17F: One step generation of an endogenous reporter allele (SEQ ID NOs: 99 and 100). (17A) Schematic overview of strategy to generate a Nanog-mCherry knock-in allele. The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. The stop codon of Nanog is labeled in orange. The homologous arms of the donor vector are indicated as HA-L (2kb) and HA-R (3kb). The restriction enzyme used for Southern blot analysis is shown, and the Southern blot probes are shown as red boxes. (17B) Southern analysis of Nanog-mCherry targeted allele. Ncol-digested genomic DNA was hybridized with 3'external probe. Expected fragment size: WT (wild type) = 11.5 kb, T (Targeted) = 5.6 kb. The blot was then stripped and hybridized with mCherry internal probe. Expected fragment size: WT = N/A, T = 6.6 kb. (17C) Nanog-mCherry targeted blastocysts showed expression in ICM. Mouse ES cell lines derived from targeted blastocysts remain mCherry positive, and the mCherry expression disappear upon differentiation. Scale bar, 100µm. (17D) Schematic overview of strategy to generate an Oct4-eGFP knock-in allele (SEQ ID NO: 101). The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. The homologous arms of the donor vector are indicated as HA-L (4.5kb) and HA-R (2kb). The IRES-eGFP transgene is indicated as a green box, and the PGK-Neo cassette is indicated as a grey box. The restriction enzyme used for Southern blot analysis is shown, and the Southern blot probes are shown as red boxes. (17E) Oct4-eGFP targeted blastocysts showed expression in ICM. Scale bar, 50µm. Mouse ES cell lines derived from targeted blastocysts remain GFP positive. Scale bar, 100µm. (17F) Southern analysis of Oct4-eGFP targeted allele. Southern analysis of Oct4-eGFP targeted allele. HindIII-digested genomic DNA was hybridized with
3' external probe. Expected fragment size: WT = 9 kb, Targeted = 7.2 kb. The blot was then stripped and hybridized with eGFP internal probe. Expected fragment size: WT = N/A, Targeted = 7.2 kb.

[0030] Figures 18A-18E: One step generation of aMecp2 floxed allele. (18A) Schematic of the Cas9/sgRNA/oligo targeting sites in Mecp2 intron 2 and intron 3 (SEQ ID NOs: 102-106). The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. In the oligo donor sequence, theloxP site is indicated as an orange box, and the restriction site sequences are in bold and capitalized. Restriction enzymes used for RFLP and Southern blot analysis are shown, and the Southern blot probes are shown as red boxes. (18B) Southern analysis of targeted alleles. Data of five mice are shown. EcoRI/Nhel-digested genomic DNA was hybridized with the exon3 probe. Expected fragment size: WT = 5.2 kb, 21oxP = 0.7 kb, L2-loxP = 3.9kb, Rl-loxP = 2kb. The blot was then stripped and hybridized with the exon4 probe. Expected fragment size: WT = 5.2 kb, 21oxP = 3.2 kb, L2-loxP = 3.9kb, Rl-loxP = 3.2kb. The sequence of the floxed allele is shown in Fig. 26B. (18C) In vitro Cre-mediated recombination of the floxed Mecp2 allele. The genomic DNA of targeted mice #1 and #3 was incubated with Cre recombinase, and used as PCR template. Primers DF and DR flanking the floxed allele produce shorter products upon Cre-dependent excision. Primers CF and CR detect the circular molecule, which only form upon Cre-loxP recombination. The position of each primer is shown at the bottom cartoon. The deletion and circular PCR products were sequenced and the sequences are shown in Fig. 26C. (18D) Injection of Cas9 mRNA and both L2 and R1 sgRNA generated Mecp2 mutant allele with deletion of exon 3. PCR genotyping using primers DF and DR identified defined deletion events in mice #1, #6, and #8 (indicated by stars). (18E) Sequences of three mutant alleles with exon 3 deletions in three mice (SEQ ID NOs: 107-111). R2 and L1 sgRNA coding sequences were underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green.

[0031] Figures 19A-19C: Integration of loxP sites at Tet1 and Tet2 loci. (19A) Schematic of the Cas9/sgRNA/oligo targeting sites in Tet1 exon 4 and Tet2 exon 3 (SEQ ID NOs: 112-115). The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in
green. In oligo donor sequence, the loxP site is indicated by an orange box, and the restriction site sequence is in bold and capitalized. (19B) RFLP analysis of double sgRNA/oligo injection mice with HDR-mediated targeting at the Tetl and Tet2 loci. About 500bp regions around the targeting sites at Tetl and Tet2 were amplified from 16 embryos and digested with EcoRI. A corrected targeted allele is identified as a cleaved fragment. Samples containing targeted alleles are indicated by stars. (19C) The sequences of targeted alleles of Tetl and Tet2 in sample #2 and #9 confirmed precise integration of loxP sites at both loci.

Figures 20A-20C: Characterization of Nanog-mCherry alleles. (20A) ES clone with mosaic expression of mCherry. The mCherry negative colony is indicated by the arrow. (20B) Southern analysis of Nanog-mCherry targeted allele identified mosaic animal. Ncol-digested genomic DNA was hybridized with 3’external probe. Expected fragment size: WT = 11.5 kb, T = 5.6 kb. Mouse #6 is identified as mosaic, because the targeted band (indicated by arrow) is weaker than WT band. (20C) The blot was then stripped and hybridized with mCherry internal probe. Expected fragment size: WT = NA, Targeted = 6.6 kb. In addition to the targeted allele, one extra band (indicated by arrow) is present in mouse #3, indicating a random insertion of the donor vector.

Figures 21A-21B: Integration of loxP sites at Meclip2 intron 2 and 3. (21A) Schematic of the Cas9/sgRNA/oligo targeting sites (SEQ ID NOs: 116-124). The sgRNA coding sequence is underlined, capitalized, and labeled in red. The protospacer-adjacent motif (PAM) sequence is labeled in green. In oligo donor sequence, the loxP site is labeled as an orange box, and the restriction site sequence is in bold and capitalized. PCR primers used for RFLP analysis are shown as red arrows. For intron 2, two sgRNA coding sequences L1 and L2 are shown, and their corresponding oligos are named accordingly. For intron 3, R1, R2 and their targeting oligos are shown. PCR primers LF and LR are used to amplify the intron 2 region, while RF and RR are used to amplify the intron 3 region. (21B) RFLP analysis of single sgRNA/oligo injection mice with HDR-mediated targeting at Meclip2 intron 2 or intron 3. Cleavage of PCR product upon Nhel or EcoRI digestion indicates loxP integration at intron 2 or intron 3 respectively. LoxP integration efficiency at L1, L2, R1, and R2 sites are compared. Samples containing loxP site are labeled by stars. Three out of eight samples contained a loxP site at the L1 site, and four out of eight
contained a loxP site at the L2 site. Two out of six samples contained loxP site at R1 site, while none was detected at the R2 site. Primers used for each PCR are labeled.

[0034] Figures 22A-22C: Analysis of MeCP2 floxed allele. (22A) RFLP analysis detected loxP integration at intron 2 (MeCP2-L2) and intron 3 (MeCP2-R1) in mice derived from L2 and R1 double sgRNA/oligo injections. Primers LF and LR were used to amplify intron 2 region, and RF and RR were used to amplify intron 3 region. Mice containing loxP sites in both introns are marked by stars. (22B) Partial chromatograph from one single sequencing file crossing both loxP sites, exon 3, and flanking intron sequences. (22C) Partial chromatograph from sequences of Cre-mediated recombination PCR products (deletion and circular products from Fig. 18C).

[0035] Figures 23A-23D: CRISPR-on activates exogenous transgenes. (23A) Schematic of the dCas9VP48 mediated transgene activation in HeLa cells. dCas9VP48 was generated by fusing dCas9 (indicated by black circle) to VP48 domain (indicated by green diamond). sgRNA complementary to rtTA binding site is indicated by small hairpin labeled sgTetO. (23B) dCas9VP48 activates TetO::tdTomato transgene in HeLa cells. Upper (top) panel, phase contrast picture of transfected cells; middle panel, tdTomato signal using fluorescent microscopy; bottom panel, FACS analysis of transfected cells. Column i, cells transfected with GFP plasmid; Column ii, cells treated with doxycycline; Column iii, cells transfected with dCas9VP48 only; Column iv, cells transfected with dCas9VP48 and sgTetO. Cells were transfected with the indicated plasmids and 48 hr later were analyzed by flow cytometry for tdTomato expression. (23C) Schematic of the dCas9VP48 mediated reporter activation in early mouse embryos. dCas9VP48, Nanog::EGFP vector, and 7 sgRNAs targeted on Nanog promoter were co-injected into mouse zygotes and cultured into blastocyst stage. (23D) dCas9VP48/sgRNA can activate gene in vivo. Left panel, embryos injected with dCas9VP48 and Nanog::EGFP vector; right panel, embryos injected with dCas9VP48, Nanog::EGFP vector and sgRNAs targeting Nanog promoter. Embryos two, three, four days post-injection were shown.

start site (TSS) and start codon (ATG) are indicated. Short lines with number indicate targeting sites of the sgRNAs. (24C) Activation of human *IL1RN* expression in HEK293T cells. Cells transfected with dCas9VP160 and indicated sgRNAs were analyzed by qRT-PCR 2 days later. sgTetO-mut, negative control sgRNA. Error bars show standard deviation (SD) among triplicates. (24D) Schematic of the human *SOX2* promoter region. Locations of TSS and start codon (ATG) are indicated. Short lines with number indicate locations of sgRNAs. (24E) Activation of *SOX2*. Cells transfected with dCas9VP160 and indicated sgRNAs were analyzed by qRT-PCR 2 days later. sgTetO-mut, negative control sgRNA. Error bars show SD among triplicates. (24F) Schematic of the human *OCT4* promoter region. Locations of transcription start site (TSS) and start codon (ATG) are indicated. Short lines with number indicate locations of sgRNAs. (24G) Activation of *OCT4*. Cells transfected with dCas9VP160 and indicated sgRNAs were analyzed by qRT-PCR 2 days later. sgTetO-mut, negative control sgRNA. Error bars show SD among triplicates.

[0037] Figures 25A-25B: Multiple exogenous and endogenous genes were simultaneously activated by CRISPR-on. (25A) One exogenous and two endogenous genes were simultaneously activated by CRISPR-on. Cells transfected with dCas9VP160 and indicated sgRNAs were analyzed by qRT-PCR 2 days later. sgTetO-mut, negative control sgRNA. Error bars show SD among triplicates. (25B) Three endogenous genes *SOX2*, *IL1RN*, and *OCT4*, can be simultaneously activated by dCas9VP160/sgRNAs. Cells were transfected with dCas9VP160 and indicated sgRNAs and were analyzed by qRT-PCR 2 days later. sgTetO-mut, negative control sgRNA. The last three sets of bars represent triple activation experiments using sgSOX2, sgOCT4 and sgIL1RN with three different ratios of sgSOX2:sgIL1RN, keeping the amount of sgOCT4 constant, as indicated by numbers above line. Error bars show SD among triplicates.

[0038] Figures 26A-26D: CRISPR-on is specific. (26A) The histogram showing distribution of Log₂ fold changes of gene expression in sample transfected with dCas9VP160/sgTetO over dCas9VP160/sgTetO-mut control. (26B) A histogram showing distribution of Log₂ fold changes of gene expression in sample transfected with dCas9VP160/sgIL1RN 1–3 over dCas9VP160/sgTetO-mut control. The vertical line marks the fold change of the target gene IL1RN. (26C) Column graph showing the log₂ fold changes of genes up-regulated by at least two fold in cells
transfected with dCas9VP160/sgTetO over dCas9VP160/sgTetO-mut. The dotted line indicates the 2 fold cut-off. (26D) Column graph showing the log$_2$ fold changes of genes up-regulated by more than two fold in cells transfected with dCas9VP160/sgIL1RNl~3 over dCas9VP160/sgTetO-mut.

[0039] Figure 27: The persistence of CRISPR-on mediated transgene expression. Cells were transfected with dCas9VP48/sgTetO in HeLa/TetO::tdTomato;EFla::rtTA-M2 cells and mean fluorescence measured by FACS was shown as fold relative to non-transfected control for the indicated samples 2 days, 12 days and 18 days after transfection.

[0040] Figures 28A-28B: CRISPR-on activates transgene in mouse cells. dCas9VP48 guided by sgRNA targeting rtTA binding site activates TetO promoter inNIH3T3/TetO::tdTomato;EFla::rtTA-M2 cells. (28A) Schematic of the dCas9VP48 mediated transgene activation in NIH3T3 cells. dCas9VP48 was generated by fusion of dCas9 to VP48 and then co-transfected with sgRNA complementary to tet binding site inNIH3T3/TetO::tdTomato;EFla::rtTA-M2 cells. (28B) dCas9VP48 depends on sgRNA to bind to the target tetO promoter to activate TetO::tdTomato transgene in NIH3T3 cells. Cells were transfected with the indicated plasmids or sgRNAs and were analyzed by flow cytometry for tdTomato expression 48 hours later.

[0041] Figure 29: CRISPR-on activated a single-copy transgene in ESCs. Cells were transfected with the indicated plasmids into a Tet-inducible MSI1 over-expression mouse embryonic stem cell (mESC) line and were analyzed by western blot for MSI1 expression 48 hours later.

[0042] Figures 30A-30B: Tunable gene activation can be achieved by titration of sgRNA. (30A) Schematic of CRISPR-on-mediated transgene activation with titration of sgRNA in Hela cells. (30B) Fold changes of mean fluorescence of tdTomato under the control of TetO promoter. Cells were transfected with dCas9 activator and indicated amount of sgTetO and mean fluorescence measured by FACS were analyzed 2 days later. NT = not transfected; C = negative control sgRNA.

[0043] Figures 31A-31B: dCas9VP48 with 6 sgRNAs failed to activate the IL1RN gene. (31A) Schematic of the human IL1RN promoter region. Locations of transcription start site (TSS) and start codon (ATG) are indicated. Short lines with
number indicate locations of sgRNAs. (3 IB) Activation of human IL1RN expression in HEK293T cells by dCas9VP48/sgRNAs. Cells were transfected with dCas9VP48 and six sgRNAs and 2 days later were analyzed by qRT-PCR. sgTetO-mut, negative control sgRNA. Error bars show SD among triplicates.

Figures 32A-32C: Nucleotide sequences of dCas9VP64 on pmax expression vector (SEQ ID NO: 486), dCas9Vp96 on pmax expression vector (SEQ ID NO: 487), and dCas9Vp60 on pmax expression vector (SEQ ID NO: 488).

DETAILED DESCRIPTION OF THE INVENTION

A description of example embodiments of the invention follows.

Mice carrying mutations in multiple genes are traditionally generated by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with single mutants. Described herein is the development of an efficient technology for the generation of animals carrying multiple mutated genes. Specifically, the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated genes (Cas genes), referred to herein as the CRISPR/Cas system, has been adapted as an efficient gene targeting technology, e.g., for multiplexed genome editing. Demonstrated herein is that CRISPR/Cas mediated gene editing allows the simultaneous disruption of five genes (Tet1, Tet2, Tet3, Sry, Uty - 8 alleles) in mouse embryonic stem cells (mESCs) with high efficiency. Co-injection of Cas9 mRNA and single guide RNA (sgRNA) targeting Tet1 and Tet2 into zygotes generated mice with biallelic mutations in both genes with an efficiency of 80%. In addition, co-injection of Cas9 mRNA/sgRNAs with mutant oligos generated precise point mutations in target genes. Thus, shown herein is that the CRISPR/Cas system allows the one step generation of animals carrying mutations in multiple genes, an approach that will greatly accelerate the in vivo study of, for example, functionally redundant genes and of epistatic gene interactions. In certain embodiments a method described herein generates non-human mammals, e.g., mice, with biallelic mutations in 1, 2, 3, 4, 5, or more genes with an efficiency of between 20% and 95%, or even more, e.g., at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, or more, e.g., up to 96%, 97%, 98%, 99%, or more. For example, in certain embodiments a method described herein generates non-human mammals, e.g., mice, with biallelic mutations in 2, 3, 4, 5, or more genes with an efficiency of
at least 70%, 80%, 85%, 90%, 95%, or more, e.g., between 70% and 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more.

[0047] Accordingly, in one aspect, the invention is directed to a method of mutating or modulating one or more target nucleic acid sequences in a (one or more) stem cell or a zygote comprising introducing into the stem cell or zygote (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associated (Cas) protein; and a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity. The stem cell or zygote is maintained under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences, and the Cas protein cleaves each of the one or more target nucleic acid sequences upon hybridization of the one or more RNA sequences to the portion of the target nucleic acid sequence, thereby mutating one or more target nucleic acid sequences in the stem cell or zygote. In a particular aspect, the stem cell or zygote into which the one or more RNA sequences and Cas nucleic acid sequence are introduced is an isolated stem cell or isolated zygote. The method can also further comprise introducing the stem cell or zygote into a nonhuman mammal.

[0048] The methods described herein can be used to mutate or modulate one or more nucleic acid sequences in a variety of stem cells which include totipotent, pluripotent, multipotent, oligipotent and unipotent stem cells. Specific examples of stem cells include embryonic stem cells, fetal stem cells, adult stem cells, and induced pluripotent stem cells (iPSCs) (e.g., see U.S. Published Application Nos. 20100144031, 20110076678, 20110088107, 20120028821 all of which are incorporated herein by reference).

[0049] In some embodiments a stem cell is a pluripotent cell. A "pluripotent" cell has the ability to self-renew and to differentiate into cells of all three embryonic germ layers (endoderm, mesoderm and ectoderm) and, typically, has the potential to divide in vitro for a long period of time, e.g., at least 20, at least 25, or at least 30 passages, or more (e.g., up to 80 passages, or up to 1 year, or more), without losing its self-renewal and differentiation properties. A pluripotent cell is said to exhibit or be in a "pluripotent state". A pluripotent cell line or cell culture is often characterized in that the cells can differentiate into a wide variety of cell types in
vitro and in vivo. Cells that are able to form teratomas containing cells having characteristics of endoderm, mesoderm, and ectoderm when injected into SCID mice are considered pluripotent. Cells that possess ability to participate in formation of chimeras (upon injection into a blastocyst of the same species that is transferred to a suitable foster mother of the same species) that survive to term are pluripotent. If the germ line of the chimeric animal contains cells derived from the introduced cell, the cell is considered germline-competent in addition to being pluripotent.

ES cells are examples of pluripotent cells. ES cells have been derived from mice, primates (including humans), and some other species. ES cells are often derived from cells obtained from the inner cell mass (ICM) of a vertebrate blastocyst but can also be derived from single blastomeres (e.g., removed from a morula).

Pluripotent cells can also be obtained using somatic cell nuclear transfer in at least some species, e.g., mice and various non-human primates. Pluripotent cells can also be obtained using parthenogenesis, e.g., from germ cells, e.g., oocytes. Other pluripotent cells include embryonic carcinoma (EC) and embryonic germ (EG) cells.


"Reprogramming", as used herein, refers to a process that alters the differentiation state or identity of a cell. Induced pluripotent stem (iPS) cells are pluripotent, ES-like cells derived from somatic cells (e.g., fibroblasts, keratinocytes, hematopoietic cells, neural precursor cells) by reprogramming. Reprogramming can be performed using a variety of different methods. As used herein, "reprogramming protocol" refers to any treatment or combination of treatments that causes at least some cells to become reprogrammed. In some embodiments "reprogramming protocol" refers to a set of manipulations (e.g., introduction of nucleic acid(s), e.g., vector(s), carrying particular genes) and/or culture conditions (e.g., culture in medium containing particular compounds) that generates pluripotent cells from somatic cells, e.g., in vitro. As used herein, the term "reprogramming factor" encompasses genes, RNAs, or proteins that promote or contribute to cell reprogramming, e.g., in vitro. Many useful reprogramming factors are transcription factors. In some aspects the terms "reprogramming", "reprogramming to a pluripotent state", "reprogramming to pluripotency", refer to in vitro reprogramming methods that do not require and typically do not include nuclear or cytoplasmic transfer or cell fusion, e.g., with oocytes, embryos, germ cells, or
pluripotent cells. Any embodiment or claim may specifically exclude compositions or methods relating to or involving nuclear or cytoplasmic transfer or cell fusion, e.g., fusion of a somatic cell with oocytes, embryos, germ cells, or pluripotent cells or transfer of a somatic cell nucleus to oocytes, embryos, germ cells, or pluripotent cells.

Differentiated cells can be reprogrammed to a pluripotent state by overexpress of the four transcription factors Oct4, Sox2, Klf4, and c-Myc (Takahashi, K. & Yamanaka, S. Cell 126, 663-676, 2006). Fully reprogrammed induced pluripotent stem cells (iPSCs) can contribute to the three germ layers and give rise to fertile mice by tetraploid complementation (Wernig, M., et al. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448, 318-324); Hanna, J., et al. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature 462, 595-601). The reprogramming process is characterized by widespread epigenetic changes that generate iPSCs that are functionally and molecularly similar to embryonic stem (ES) cells (Carey, B. W. et al. Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. Cell Stem Cell 9, 588-598, (2011)).

Reprogramming somatic cells to a pluripotent state can be achieved by infecting cells with retroviruses that encode the transcription factors Oct4, Sox2, Klf4, and c-Myc (termed "OSKM factors") under control of a viral LTR. Oct4, Sox2 and Klf4 ("OSK factors") are also sufficient to reprogram mammalian, e.g., rodent or human, somatic cells to pluripotency. Other sets of reprogramming factors, e.g., Oct4, Sox2, Nanog, and Lin28 (OSNL factors) can be used to reprogram mammalian cells, e.g., rodent or human cells, with Lin28 being dispensable. The ectopically expressed factors induce expression of endogenous pluripotency genes such as Oct4 and Nanog. Since the retroviral vectors in iPS cells derived by this approach are silenced, maintenance of pluripotency relies on expression of such endogenous genes and establishment of an appropriate transcriptional network in the reprogrammed cells. Furthermore, reprogramming factors that are members of the same gene family may be used in place of one another in certain embodiments. For example, Klf2 and Klf5 can substitute for Klf4, Sox1 for Sox2 and N-Myc for c-Myc. It has recently been discovered that reprogramming can be achieved using
Sall4, Nanog, Esrrb, and Lin28 as reprogramming factors (SNEL factors) or using Sal4, Lin28, Essrb, and Dppa2 (SLED factors) (Buganim Y, et al, Cell. 2012 Sep 14;150(6):1209-22). Thus, examples of reprogramming factors of interest for reprogramming somatic cells to pluripotency in vitro include Oct4, Sall4, Nanog, Esrrb, Lin28, Klf4, c-Myc, Dppa2, and any gene/RNAprotein that can substitute for one or more of these in a method of reprogramming somatic cells in vitro.

Exogenous reprogramming factors may be introduced into somatic cells in any form that is capable of maintaining exogenous reprogramming factors for a period of time and at levels sufficient to activate endogenous pluripotency genes and for reprogramming of at least some of the somatic cells into which the exogenous reprogramming factors are introduced to occur. As used herein, "exogenous" refers to a substance present in a cell or organism other than its native source. For example, the terms "exogenous nucleic acid" or "exogenous protein" refer to a nucleic acid or protein that has been introduced by a process involving the hand of man into a biological system such as a cell or organism in which it is not normally found or in which it is found in lower amounts. A substance will be considered exogenous if it is introduced into a cell or an ancestor of the cell that inherits the substance. In contrast, the term "endogenous" refers to a substance that is native to the biological system.

Somatic cells of use in aspects of the invention may be primary cells (non-immortalized cells), such as those freshly isolated from an animal, or may be derived from a cell line capable of prolonged proliferation in culture (e.g., for longer than 3 months) or indefinite proliferation (immortalized cells). Adult somatic cells may be obtained from individuals, e.g., human subjects, and cultured according to standard cell culture protocols available to those of ordinary skill in the art. Cells may be maintained in cell culture following their isolation from a subject. In certain embodiments, the cells are passaged once or more following their isolation from the individual (e.g., between 2-5, 5-10, 10-20, 20-50, 50-100 times, or more) prior to their use in a method of the invention. In some embodiments, cells may be frozen and subsequently thawed prior to use. In some embodiments, cells will have been passaged no more than 1, 2, 5, 10, 20, or 50 times following their isolation from an individual prior to their use in a method of the invention. Somatic cells of use in aspects of the invention include mammalian cells, such as, for example, human cells,
non-human primate cells, or rodent (e.g., mouse, rat) cells. They may be obtained by well-known methods from various organs, e.g., skin, lung, pancreas, liver, stomach, intestine, heart, breast, reproductive organs, muscle, blood, bladder, kidney, urethra and other urinary organs, etc., generally from any organ or tissue containing live somatic cells. Mammalian somatic cells useful in various embodiments include, for example, fibroblasts, Sertoli cells, granulosa cells, neurons, pancreatic cells, epidermal cells, epithelial cells, endothelial cells, hepatocytes, hair follicle cells, keratinocytes, epithelial cells, endothelial cells, hepatocytes, hair follicle cells, keratinocytes, hematopoietic cells, melanocytes, chondrocytes, lymphocytes (B and T lymphocytes), macrophages, monocytes, mononuclear cells, cardiac muscle cells, skeletal muscle cells, etc. In some embodiments a somatic cell is a terminally differentiated somatic cell. In some embodiments a somatic cell is a progenitor (precursor) cell, which has not terminally differentiated.

[0056] In some embodiments, reprogramming factors are introduced into somatic cells in the form of one or more nucleic acid sequences encoding the reprogramming factors. In some embodiments, reprogramming factors are introduced into somatic cells in the form of one or more nucleic acid sequences encoding the reprogramming factors. In some embodiments, the one or more nucleic acid sequences comprise DNA. In some embodiments, the one or more nucleic acid sequences comprise RNA. In some embodiments, the one or more nucleic acid sequences comprise a nucleic acid construct. In some embodiments, the one or more nucleic acid sequences comprise a vector for delivery of the reprogramming factors into a target cell (e.g., a mammalian somatic cell, e.g., a human or mouse fibroblast cell). Any suitable vector may be used. Examples of suitable vectors are described by Stadtfeld and Hochedlinger (Genes Dev. 24:2239-2263, 2010, incorporated herein by reference in its entirety). Other suitable vectors are apparent to those skilled in the art.

[0057] In some embodiments, a vector comprises an inducible vector. In some embodiments, the inducible vector is a doxycycline inducible vector (i.e., a vector activates expression of said reprogramming factors in the presence of doxycycline in a culture medium). "Expression" refers to the cellular processes involved in producing RNA and proteins as applicable, for example, transcription, translation, folding, modification and processing. "Expression products" include RNA transcribed from a gene and polypeptides obtained by translation of mRNA
transcribed from a gene. In some embodiments, the inducible vector is a tamoxifen inducible vector or encodes a tamoxifen-inducible protein. In some embodiments, a vector is an integrating vector that integrates into a genome of a host cell (e.g., a mammalian somatic cell). In some embodiments, a vector comprises a viral vector, e.g., a retroviral vector, e.g., a lentiviral vector. In some embodiments, a vector comprises an excisable vector. In some embodiments, the excisable vector comprises a transposon, wherein said excisable vector is excisable from said genome by transient expression of a transposase. In certain embodiments, the transposon comprises a piggyback transposon (See, e.g., Woltjen et al. Nature 458:766-770, 2009; Yusa et al. Nat Methods 6:363-369, 2009, incorporated herein by reference in its entirety). In some embodiments, the excisable vector comprises one or more loxP site incorporated into said vector, wherein said vector can be excised from said genome by transient expression of a Cre recombinase (See, e.g., Kaji et al. Nature 458:771-775, 2009; Soldner et al. Cell 136:964-977, 2009, each of which is incorporated herein by reference in its entirety). In some embodiments, the excisable vector comprises a floxed lentiviral vector.

[0058] In some embodiments, the vector does not integrate into the genome of said somatic cell. In some embodiments, the vector comprises an adenoviral vector (See, e.g., Zhou and Freed. Stem Cells 27:2667-2674, 2009, the teachings of which are incorporated herein by reference). In some embodiments, the vector comprises a Sendai viral vector (See, e.g., Fusaki et al. Proc Jpn Acad 85:348-362, 2009, the teachings of which are incorporated herein by reference). In some embodiments, the vector comprises a plasmid. In some embodiments, the vector comprises an episome (Yu et al. Science 324(5928):797-801, 2009, the teachings of which are incorporated herein by reference).

[0059] In some embodiments, to minimize the number of independent proviral integrations required for reprogramming, a nucleic acid construct comprises a polycistronic vector that can transduce any combination of reprogramming factors with a goal of reducing the number of proviral integrations. Such polycistronic nucleic acid constructs, expression cassettes, and vectors that employ internal ribosomal entry sites and self-cleaving peptides and are capable of transducing any combination of reprogramming factors are described in PCT Application Publication No. WO 2009/152529, incorporated herein by reference in its entirety.
In certain embodiments reprogramming factors are provided by polycistronic nucleic acid constructs (e.g., expression cassettes, and vectors comprising such constructs). In certain embodiments the polycistronic nucleic acid constructs comprise a portion that encodes a self-cleaving peptide. In certain embodiments a polycistronic nucleic acid construct comprises at least two, three, or four, coding regions, wherein the coding regions are linked to each by a nucleic acid that encodes a self-cleaving peptide so as to form a single open reading frame, and wherein the coding regions encode at least first and second reprogramming factors capable, either alone or in combination with one or more additional reprogramming factors, of reprogramming a mammalian somatic cell to pluripotency. In some embodiments of the invention the construct comprises two coding regions separated by a self-cleaving peptide. In some embodiments constructs encode a polyprotein that comprises 2, 3, or 4 reprogramming factors, separated by self-cleaving peptides. In some embodiments the construct comprises expression control element(s), e.g., a promoter, suitable to direct expression in mammalian cells, wherein the portion of the construct that encodes the polypeptide is operably linked to the expression control element(s). The promoter drives transcription of a polycistronic message that encodes the reprogramming factors, each reprogramming factor being linked to at least one other reprogramming factor by a self-cleaving peptide. The promoter can be a viral promoter (e.g., a CMV promoter) or a mammalian promoter (e.g., a PGK promoter). The expression cassette or construct can comprise other genetic elements, e.g., to enhance expression or stability of a transcript. In some embodiments of the invention any of the foregoing constructs or expression cassettes may further include a coding region that does not encode a reprogramming factor, wherein the coding region is separated from adjacent coding region(s) by a self-cleaving peptide. In some embodiments the additional coding region encodes a selectable marker. In some embodiments, the self-cleaving peptide is a viral 2A peptide. In some embodiments, the self-cleaving peptide is an aphthovirus 2A peptide.

In some embodiments a construct comprises sites for a recombinase that is functional in mammalian cells, wherein the sites flank at least the portion of the construct that comprises the coding regions for the factors (i.e., one site is positioned 5’ and a second site is positioned 3’ to the portion of the construct that encodes the
polyprotein), so that the sequence encoding the factors can be excised from the genome after reprogramming. The recombinase can be, e.g., Cre or Flp, where the corresponding recombinase sites are LoxP sites and Frt sites. In some embodiments the recombinase is a transposase. It will be understood that the recombinase sites need not be directly adjacent to the region encoding the polyprotein but will be positioned such that a region whose eventual removal from the genome is desired is located between the sites. In some embodiments the recombinase sites are on the 5’ and 3’ ends of an expression cassette. Excision may result in a residual copy of the recombinase site remaining in the genome, which in some embodiments is the only genetic change resulting from the reprogramming process.

In some embodiments, one or more nucleic acids for introducing reprogramming factors comprise mRNA that is translatable in a mammalian somatic cell. In some embodiments, the mRNA can be introduced in vitro into somatic cells to be reprogrammed and translated by endogenous enzymes into proteins that can activate one or more endogenous pluripotency genes in the cell. As used herein, "pluripotency gene", refers to a gene whose expression under normal conditions (e.g., in the absence of genetic engineering or other manipulation designed to alter gene expression) occurs in and is typically restricted to pluripotent stem cells, and is crucial for their functional identity as such. It will be appreciated that the polypeptide encoded by a pluripotency gene may be present as a maternal factor in the oocyte. The gene may be expressed by at least some cells of the embryo, e.g., throughout at least a portion of the preimplantation period and/or in germ cell precursors of the adult. The gene may be expressed in ES cells and/or in embryonic carcinoma cells. The pluripotency gene is typically substantially not expressed in somatic cell types that constitute the body of an adult animal under normal conditions (with the exception of germ cells or precursors thereof, or possibly in certain disease states such as cancer). For example, the pluripotency gene may be one whose average expression level (based on RNA or protein) in ES cells is at least 50-fold or 100-fold greater than its average level in those terminally differentiated cell types present in the body of an adult mammal. In some embodiments, the pluripotency gene is one that encodes multiple splice variants or isoforms of a protein, wherein one or more such variants or isoforms is expressed in at least some adult somatic cell types, while one or more other variants or isoforms is not
substantially expressed in adult somatic cells under normal conditions. In some embodiments, expression of the pluripotency gene is essential to maintain the viability or pluripotent state of iPSCs. Thus if the gene is knocked out or its expression is inhibited (i.e., its expression is eliminated or substantially reduced, e.g., such that the average steady state level of RNA transcript and/or protein encoded by the gene is decreased by at least 50%, 60%, 70%, 80%, 90%, 95%, or more), the iPSCs are not formed, die or, in some embodiments, differentiate or cease to be pluripotent. In some embodiments the pluripotency gene is characterized in that its expression in an ES cell or iPS cell decreases (resulting in, e.g., a reduction in the average steady state level of RNA transcript and/or protein encoded by the gene by at least 50%, 60%, 70%, 80%, 90%, 95%, or more) when the cell differentiates into a terminally differentiated cell. Oct4 and Nanog are exemplary pluripotency genes. In some embodiments, the mRNA is in vitro transcribed mRNA. Non-limiting examples of producing in vitro transcribed mRNA are described by Warren et al. (Cell Stem Cell 7(5):618-30, 2010, Mandal PK, Rossi DJ. Nat Protoc. 2013 Mar;8(3):568-82, and/or PCT/US2011/032679 (WO/2011/130624) the teachings of each of which are incorporated herein by reference). The protocols described may be adapted to produce one or more mRNAs of interest in the present invention. In some embodiments, mRNA, e.g., in vitro transcribed mRNA, comprises a sequence encoding SV40 large T (LT). In some embodiments, mRNA, e.g., in vitro transcribed mRNA, comprises one or more modifications that increase stability or translatability of said mRNA. In some embodiments, mRNA, e.g., in vitro transcribed mRNA comprises a 5' cap. The cap may be wild-type or modified. Examples of suitable caps and methods of synthesizing mRNA containing such caps are apparent to those skilled in the art.

[0063] In some embodiments, mRNA, e.g., in vitro transcribed mRNA, comprises an open reading frame flanked by a 5' untranslated region and a 3' untranslated region that enhance translation of said open reading frame, e.g., a 5' untranslated region that comprises a strong Kozak translation initiation signal, and/or a 3' untranslated region comprises an alpha-globin 3' untranslated region.

[0064] In some embodiments, mRNA, e.g., in vitro transcribed mRNA comprises a polyA tail. Methods of adding a polyA tail to mRNA are known in the art, e.g., enzymatic addition via polyA polymerase or ligation with a suitable ligase.
The methods provided herein can also be used to mutate or modulate one or more nucleic acids in stem cells that are present in cell compositions such as embryos, zygotes, fetuses, and post-natal mammals. In some embodiments, a stem cell (e.g., an ES or iPS cell), zygote, embryo, or post-natal mammal is already genetically modified (already harbors one or more genetic modifications) prior to being subjected to the methods described herein. For example, the stem cell (e.g., an ES or iPS cell), zygote, embryo, or post-natal mammal may be one into which an exogenous nucleic acid has been introduced by a process involving the hand of man (or may be descended at least in part from a cell or organism into which an exogenous nucleic acid has been introduced by a process involving the hand of man). The nucleic acid may for example contain a sequence that is exogenous to the cell, it may contain native sequences (i.e., sequences naturally found in the cells) but in a non-naturally occurring arrangement (e.g., a coding region linked to a promoter from a different gene), or altered versions of native sequences, etc. In some embodiments, a stem cell (e.g., an ES or iPS cell), zygote, embryo, or post-natal mammal is not already genetically modified (does not already harbor one or more genetic modifications) prior to being subjected to the methods described herein.

The stem cell, zygote, embryo, or post-natal mammal can be of vertebrate (e.g., mammalian) origin. In some aspects, the vertebrates are mammals or avians. Particular examples include primate (e.g., human), rodent (e.g., mouse, rat), canine, feline, bovine, equine, caprine, porcine, or avian (e.g., chickens, ducks, geese, turkeys) stem cells, zygotes, embryos, or post-natal mammals. In some embodiments, the stem cell, zygote, embryo, or post-natal mammal is isolated (e.g., an isolated stem cell; an isolated zygote; an isolated embryo). In some embodiments, a mouse stem cell, mouse zygote, mouse embryo, or mouse post-natal mammal is used. In some embodiments, a rat stem cell, rat zygote, rat embryo, or rat post-natal mammal is used. In some embodiments, a human stem cell, human zygote or human embryo is used.

In some aspects, the invention is directed to a method of producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences comprising introducing into a zygote or an embryo (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a
CRISPR associated (Cas) protein; and ii) a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity. The zygote or the embryo is maintained under conditions in which RNA hybridizes to the portion of each of the one or more target nucleic acid sequences, and the Cas protein cleaves each of the one or more target nucleic acid sequences upon hybridization of the RNA to the portion of the target nucleic acid sequence, thereby producing an embryo having one or more mutated nucleic acid sequences. The embryo having one or more mutated nucleic acid sequences may be transferred into a foster nonhuman mammalian mother. The foster nonhuman mammalian mother is maintained under conditions in which one or more offspring carrying the one or more mutated nucleic acid sequences are produced, thereby producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences.

[0068] As will be apparent to those of skill in the art, the nonhuman mammals can also be produced using methods described herein and/or with conventional methods, see for example, U.S. Published Application No. 201 10302665. A method of producing a non-human mammalian embryo can comprise injecting non-human mammalian ES cells (e.g., iPSCs) genetically modified according to an inventive method of the present invention into non-human tetraploid blastocysts and maintaining said resulting tetraploid blastocysts under conditions that result in formation of embryos, thereby producing a non-human mammalian embryo. In some embodiments, said non-human mammalian cells are mouse cells and said non-human mammalian embryo is a mouse. In some embodiments, said mouse cells are mutant mouse cells and are injected into said non-human tetraploid blastocysts by microinjection. In some embodiments laser-assisted micromanipulation or piezo injection is used. In some embodiments, a non-human mammalian embryo comprises a mouse embryo.

[0069] Another example of such conventional techniques is two step cloning which involves introducing embryonic stem (ES) and/or induced pluripotent stem (iPS) cells comprising the one or more mutations into a blastocyst (e.g., a tetraploid blastocyst) and maintaining the blastocyst under conditions that result in development of an embryo. The embryo is then transferred (impregnated) into an appropriate foster mother, such as a pseudopregnant female (e.g., of the same
species as the embryo). The foster mother is then maintained under conditions that result in development of live offspring that harbor the one or more mutations.

Another example is the use of the tetraploid complementation assay in which cells of two mammalian embryos are combined to form a new embryo (Tarn and Rossant, Develop, 750:6156-6163 (2003)). The assay involves producing a tetraploid cell in which every chromosome exists fourfold. This is done by taking an embryo at the two-cell stage and fusing the two cells by applying an electrical current. The resulting tetraploid cell continues to divide, and all daughter cells will also be tetraploid. Such a tetraploid embryo develops normally to the blastocyst stage and will implant in the wall of the uterus. In the tetraploid complementation assay, a tetraploid embryo (either at the morula or blastocyst stage) is combined with normal diploid embryonic stem cells (ES) from a different organism. The embryo develops normally; the fetus is exclusively derived from the ES cell, while the extra-embryonic tissues are exclusively derived from the tetraploid cells.

Another conventional method used to produce nonhuman mammals includes pronuclear microinjection. DNA is introduced directly into the male pronucleus of a nonhuman mammal egg just after fertilization. Similar to the two-step cloning described above, the egg is implanted into a pseudopregnant female. Offspring are screened for the integrated transgene. Heterozygous offspring can be subsequently mated to generate homozygous animals.

A variety of nonhuman mammals can be used in the methods described herein. For example, the nonhuman mammal can be a rodent (e.g., mouse, rat, guinea pig, hamster), a nonhuman primate, a canine, a feline, a bovine, an equine, a porcine or a caprine.

In some aspects, various mouse strains and mouse models of human disease are used in conjunction with the methods of producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences described herein. One of ordinary skill in the art appreciates the thousands of commercially and non-commercially available strains of laboratory mice for modeling human disease. Mice models exist for diseases such as cancer, cardiovascular disease, autoimmune diseases and disorders, inflammatory diseases, diabetes (type 1 and 2), neurological diseases, and other diseases. Examples of commercially available research strains include, and is not limited to, 11BHSD2 Mouse, GSK3B Mouse, 129-E Mouse
HSD1 1B1 Mouse, AKR Mouse Immortomouse®, Athymic Nude Mouse, LCAT Mouse, B6 Albino Mouse, Lox-1 Mouse, B6C3F1 Mouse, Ly5 Mouse, B6D2F1 (BDF1) Mouse, MMP9 Mouse, BALB/c Mouse, NIH-III Nude Mouse, BALB/c Nude Mouse, NOD Mouse, NOD SCID Mouse, Black Swiss Mouse, NSE-p25 Mouse, C3H Mouse, NU/NU Nude Mouse, C57BL/6-E Mouse, PCSK9 Mouse, C57BL/6N Mouse, PGP Mouse (P-glycoprotein Deficient), CB6F1 Mouse, repTOP™ ERE-Luc Mouse, CD-I® Mouse, repTOP™ mitoIRE Mouse, CD-I® Nude Mouse, repTOP™ PPRE-Luc Mouse, CD1-E Mouse, Rip-HAT Mouse, CD2F1 (CDF1) Mouse, SCID Hairless Congenic (SHC™) Mouse, CF-1™ Mouse, SCID Hairless Outbred (SHO™) Mouse, DBA/2 Mouse, SJL-E Mouse, Fox Chase CB17™ Mouse, SKH1-E Mouse, Fox Chase SCID® Beige Mouse, Swiss Webster (CFW®) Mouse, Fox Chase SCID® Mouse, TARGATT™ Mouse, FVB Mouse, THE POUND MOUSE™, and GLUT 4 Mouse. Other mouse strains include BALB/c, C57BL/6, C57BL/10, C3H, ICR, CBA, A/J, NOD, DBA/1, DBA/2, MOLD, 129, HRS, MRL, NZB, NIH, AKR, SJL, NZW, CAST, KK, SENCAR, C57L, SAMR1, SAMP1, C57BR, andNZO.

[0074] In some aspects, the method of producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences further comprises mating one or more commercially and/or non-commercially available nonhuman mammal with the nonhuman mammal carrying mutations in one or more target nucleic acid sequences produced by the methods described herein. The invention is also directed to nonhuman mammals produced by the methods described herein.

[0075] In the methods provided herein, one or more ribonucleic acid (RNA) sequences comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associated (Cas) protein is introduced into the stem cell, zygote and/or embryo, etc. In some embodiments, the RNA sequence is referred to as guide RNA (gRNA) or single guide RNA (sgRNA).

[0076] In some aspects, a single RNA sequence can be complementary to one or more (e.g., all) of the target nucleic acid sequences that are being modulated or mutated. In one aspect, a single RNA is complementary to a single target nucleic acid sequence. In a particular aspect in which two or more target nucleic acid sequences are to be modulated or mutated, multiple (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or
more) RNA sequences are introduced wherein each RNA sequence is
complementary to (specific for) one target nucleic acid sequence. In some aspects,
two or more, three or more, four or more, five or more, or six or more RNA
sequences are complementary to (specific for) different parts of the same target
sequence. In one aspect, two or more RNA sequences bind to different sequences of
the same region (e.g. promoter) of DNA (see e.g., Figure 30A). In some aspects, a
single RNA sequence is complementary to at least two target or more (all) of the
target nucleic acid sequences. It will also be apparent to those of skill in the art that
the portion of the RNA sequence that is complementary to one or more of the target
nucleic acid sequences and the portion of the RNA sequence that binds to Cas
protein can be introduced as a single sequence or as 2 (or more) separate sequences
into a cell, zygote, embryo or nonhuman animal. In some embodiments the sequence
that binds to Cas protein comprises a stem-loop.

[0077] In some embodiments, the RNA sequence used to modify gene
expression in a nonhuman mammal is a naturally occurring RNA sequence, a
modified RNA sequence (e.g., a RNA sequence comprising one or more modified
bases), a synthetic RNA sequence, or a combination thereof. As used herein a
"modified RNA" is an RNA comprising one or more modifications (e.g., RNA
comprising one or more non-standard and/or non-naturally occurring bases) to the
RNA sequence (e.g. modifications to the backbone and or sugar). Methods of
modifying bases of RNA are well known in the art. Examples of such modified
bases include those contained in the nucleosides 5-methylcytidine (5mC),
pseudouridine (Ψ), 5-methyluridiné, 2′0-methyluridine, 2-thiouridine, N-6
methyladenosine, hypoxanthine, dihydouridine (D), inosine (I), and 7-
methylguanosine (m7G). It should be noted that any number of bases in a RNA
sequence can be substituted in various embodiments. It should further be understood
that combinations of different modifications may be used.

[0078] In some aspects, the RNA sequence is a morpholino. Morpholinos are
typically synthetic molecules, of about 25 bases in length and bind to
complementary sequences of RNA by standard nucleic acid base-pairing.
Morpholinos have standard nucleic acid bases, but those bases are bound to
morpholine rings instead of deoxyribose rings and are linked through
phosphorodiamidate groups instead of phosphates. Morpholinos do not degrade
their target RNA molecules, unlike many antisense structural types (e.g., phosphorothioates, siRNA). Instead, morpholinos act by steric blocking and bind to a target sequence within a RNA and block molecules that might otherwise interact with the RNA.

Each RNA sequence can vary in length from about 8 base pairs (bp) to about 200 bp. In some embodiments, the RNA sequence can be about 9 to about 190 bp; about 10 to about 150 bp; about 15 to about 120 bp; about 20 to about 100 bp; about 30 to about 90 bp; about 40 to about 80 bp; about 50 to about 70 bp in length.

The portion of each target nucleic acid sequence to which each RNA sequence is complementary can also vary in size. In particular aspects, the portion of each target nucleic acid sequence to which the RNA is complementary can be about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 100 nucleotides (contiguous nucleotides) in length. In some embodiments, each RNA sequence can be at least about 70%, 75%, 80%, 85%, 90%, 95%, 100%, etc. identical or similar to the portion of each target nucleic acid sequence. In some embodiments, each RNA sequence is completely or partially identical or similar to each target nucleic acid sequence. For example, each RNA sequence can differ from perfect complementarity to the portion of the target sequence by about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etc. nucleotides. In some embodiments, one or more RNA sequences are perfectly complementary (100%) across at least about 10 to about 25 (e.g., about 20) nucleotides of the target nucleic acid.

As will be apparent to those of ordinary skill in the art, the one or more RNA sequences can further comprise one or more expression control elements. For example, in some embodiments the RNA sequences comprises a promoter, suitable to direct expression in cells, wherein the portion of the RNA sequence is operably linked to the expression control element(s). The promoter can be a viral promoter (e.g., a CMV promoter) or a mammalian promoter (e.g., a PGK promoter). The RNA sequence can comprise other genetic elements, e.g., to enhance expression or
stability of a transcript. In some embodiments the additional coding region encodes a selectable marker (e.g., a reporter gene such as green fluorescent protein (GFP)).

As described herein, the one or more RNA sequences also comprise a (one or more) binding site for a (one or more) CRISPR associated (Cas) protein, and, upon hybridization of the one or more RNA sequences to the one or more target sequences, a (one or more) Cas protein or variant thereof cleaves or nicks each of the target nucleic acid sequences. In a particular aspect, upon hybridization of the one or more RNA sequences to the one or more target nucleic acid sequences, the Cas protein or variants thereof binds to the one or more RNA sequences and cleaves the one or more target nucleic acids sequences. Bacteria and Archaea have evolved an RNA-based adaptive immune system that uses CRISPR (clustered regularly interspaced short palindromic repeat) and Cas (CRISPR-associated) proteins to detect and destroy invading viruses and plasmids (Horvath and Barrangou, Science, 327(5962): 167-170 (2010); Wiedenheft et al, Nature, 482(7385):33 1-338 (2012)). Cas proteins, CRISPR RNAs (crRNAs) and trans-activating crRNA (tracrRNA) form ribonucleoprotein complexes, which target and degrade foreign nucleic acids, guided by crRNAs (Gasiunas et al, Proc. Natl. Acad. Sci, 109(39):E2579-86 (2012); Jinek et al, Science, 337:816-821 (2012)).

In one aspect, the method further comprises introducing one or more Cas nucleic acid or variant thereof into the cell, embryo, zygote, or non-human mammal. In some aspects, a Cas protein or variant thereof is introduced into the cell, embryo, zygote, or non-human mammal. In some aspects, a cell, e.g., stem cell (ES or iPS cell), zygote, embryo, or animal may already harbor a nucleic acid that encodes Cas (may be constitutive or inducible) and/or may already contain Cas protein. For example, in some embodiments a cell, e.g., stem cell (ES or iPS cell), zygote, embryo, or animal, may be descended from a cell or organism into which a nucleic acid encoding a Cas protein has been introduced by a process involving the hand of man.

A variety of CRISPR associated (Cas) genes or proteins which are known in the art can be used in the methods of the invention and the choice of Cas protein will depend upon the particular conditions of the method (e.g., www.ncbi.nlm.nih.gov/gene/?term=cas9). Specific examples of Cas proteins include Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 and Cas10. In a particular
aspect, the Cas nucleic acid or protein used in the methods is Cas9. In some embodiments a Cas protein, e.g., a Cas9 protein, may be from any of a variety of prokaryotic species. In some embodiments a particular Cas protein, e.g., a particular Cas9 protein, may be selected to recognize a particular protospacer-adjacent motif (PAM) sequence. In certain embodiments a Cas protein, e.g., a Cas9 protein, may be obtained from a bacteria or archaea or synthesized using known methods. In certain embodiments, a Cas protein may be from a gram positive bacteria or a gram negative bacteria. In certain embodiments, a Cas protein may be from a Streptococcus, (e.g., a S. pyogenes, a S. thermophilus) a Crptococcus, a Corynebacterium, a Haemophilus, a Eubacterium, a Pasteurella, a Prevotella, a VeilUonella, or a Marinobacter. In some embodiments nucleic acids encoding two or more different Cas proteins, or two or more Cas proteins, may be introduced into a cell, zygote, embryo, or animal, e.g., to allow for recognition and modification of sites comprising the same, similar or different PAM motifs.

[0085] The Cas protein can cleave one strand or both strands (e.g., of a double stranded target nucleic acid), or alternatively, nick one strand or both strands (e.g., of a double stranded target nucleic acid). In some embodiments a Cas9 nickase may be generated by inactivating one or more of the Cas9 nuclease domains. In some embodiments, an amino acid substitution at residue 10 in the RuvC I domain of Cas9 converts the nuclease into a DNA nickase. For example, the aspartate at amino acid residue 10 can be substituted for alanine (Cong et al, Science, 339:819-823). Other amino acids mutations that create a catalytically inactive Cas9 protein includes mutating at residue 10 and/or residue 840. Mutations at both residue 10 and residue 840 can create a catalytically inactive Cas9 protein, sometimes referred herein as dCas9. For example, a D10A and a H840A Cas9 mutant is catalytically inactive.

[0086] As shown herein, fusions of a catalytically inactive (D10A; H840A) Cas9 protein (dCas9) tethered with all or a portion of (e.g., biologically active portion of) an (one or more) effector domain create chimeric proteins that can be guided to specific DNA sites by one or more RNA sequences (sgRNA) to modulate activity and/or expression of one or more target nucleic acids sequences (e.g., exert certain effects on transcription or chromatin organization, or bring specific kind of molecules into specific DNA loci, or act as sensor of local histone or DNA state).
As used herein, a "biologically active portion of an effector domain" is a portion that maintains the function (e.g., completely, partially, minimally) of an effector domain (e.g., a "minimal" or "core" domain). Specifically, shown herein is that fusion of the Cas9 (e.g., dCas9) with all or a portion of one or more effector domains (e.g., transcriptional activation domains) created a chimeric protein. In one aspect, fusion of a dCas9 with one or more effector domains created a chimeric protein dCas9TA. In some aspects, the one or more effector domains are the same (e.g., VP16 transcriptional activation domains). In other aspects, the one or more effector (e.g., transcriptional activation) domains are different. In some aspects, dCas9TA is guided to specific nucleic acid sites by one or more RNA (e.g. sgRNA). In some aspects, dCas9TA is guided to specific nucleic acid sites by RNA (e.g. sgRNA) to modulate gene expression. In some aspects, all or a portion of one or more VP16 effector domains are fused with Cas9 (e.g., dCas9). In other aspects, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more VP16 effector domains (all or a biologically active portion) are fused with dCas9. In some aspects, a chimeric protein comprising a fusion of a catalytically inactive Cas to all or a portion of one or more effector domains is referred to herein as "CRISPRzyme" or "CRISPR-on".

[0087] In one aspect, fusion of Cas9 with all or a portion of one or more effector domains comprise one or more linkers. As used herein, a "linker" is something that connects or fuses two or more effector domains (e.g. see Hermanson, Bioconjugate Techniques, 2nd Edition, which is hereby incorporated by reference in its entirety). As will be appreciated by one of ordinary skill in the art, a variety of linkers can be used. In one aspect, a linker comprises one or more amino acids. In some aspects, a linker comprises 2 or more amino acids. In one aspect, a linker comprises the amino acid sequence GS. In some aspects, fusion of Cas9 (e.g., dCas9) with two or more effector domains (e.g., VP16 core domain such as DALDDFDLDML) comprises one or more interspersed linkers (e.g., GS linkers) between the domains. In some aspects, dCas9 is fused with 3 VP16 core domains with interspersed linkers, referred to herein as dCas9VP48. In other aspects, dCas9 is fused with 4 VP16 core domains with interspersed GS linkers between the core domains, referred herein as dCas9VP48 (SEQ ID NO: 14). In other aspects, dCas9 is fused with 6 VP16 core domains with interspersed GS linkers between the core domains, referred herein as dCas9VP96 (SEQ ID NO: 15). In other aspects, fusion of dCas9 with 10 VP16 core
domains with interspersed GS linkers between the core domains, referred herein as dCas9VP160 (SEQ ID NO: 16).

Accordingly, in some aspects, the invention is directed to a method of modulating the expression and/or activity of one or more target nucleic acid sequences in a cell or zygote comprising introducing into the cell or zygote (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associate (Cas) protein; (ii) a Cas nucleic acid sequence or a variant thereof that encodes the Cas protein that targets but does not cleave the target nucleic acid sequence; and (iii) an (one or more) effector domain. The method further comprises maintaining the cell or zygote under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences. The Cas protein binds to each of the one or more RNA sequences and the effector domain modulates the expression and/or activity of the target nucleic acid, thereby modulating the expression and/or activity of a target nucleic acid sequence. As with some aspects of the invention, one or more RNA sequences, Cas nucleic acid sequences and effector domains can be introduced into a cell, zygote, embryo or non-human mammal.

In some aspects, the method of modulating the expression and/or activation of one or more target nucleic acids in a cell is used to reprogram a cell's potency. Cells can be reprogrammed, e.g., by the methods described herein. In one aspect, the invention is directed to a method of modulating the expression and/or activity of one or more target nucleic acid sequences in a cell wherein the cell or cell's potency (e.g., totipotency, pluripotency, multipotency, oligopotency and unipotency) is reprogrammed (e.g., a differentiated cell; a non-differentiated cell). In one aspect, the method results in differentiation of a cell (e.g., a totipotent or pluripotent cell differentiates into a unipotent cell or differentiated cell). In another aspect, the methods results in dedifferentiation of a cell (e.g. a differentiated cell reverts to an earlier developmental stage). For example, the invention is directed to reprogramming a differentiated cell to a totipotent, pluripotent, or multipotent state. In other aspects the method results in transdifferentiation of the cell (e.g. a fibroblast is reprogrammed to a fat cell or a fat cell is reprogrammed to a fibroblast). In one aspect, the one or more target nucleic acid sequences in a cell are overexpressed
causing the cell to be reprogrammed. In another aspect, one or more transcription factors are modulated altering cell potency or dedifferentiation. In another aspect, one or transcription factors such as Oct4, Sox2, Klf4, and c-Myc are modulated (e.g. overexpressed) in a cell. (Takahashi, K. & Yamanaka, S. Cell 126, 663-676, 2006).

In some aspects, the invention is directed to a method of modulating one or more target nucleic acid sequences comprising simultaneous activation of the one or more target nucleic acid sequences. In another aspect, the method of modulating one or more target nucleic acid sequences comprises adjusting the level of modulation of one or more target nucleic acid sequences by adjusting the amount (e.g. grams, milligrams, micrograms, nanograms, moles, millimoles, micromoles, nanomoles, stoichiometric amount, molar ratio) of the one or more ribonucleic acid sequences introduced into the cell or zygote (Figure 30B). In some aspects, the level of modulation of one target nucleic acid sequence is the same or different compared to the level of modulation of another target nucleic acid sequence in the same cell or zygote (Figure 25B). In one aspect, multiple target nucleic acid sequences are modulated (e.g. multiplexed activation).

In some aspects the invention is directed to (e.g., a composition comprising, consisting essentially of, consisting of) a nucleic acid sequence that encodes a fusion protein (chimeric protein) comprising all or a portion of a Cas protein fused to all or a portion of an effector domain. In some aspects, the invention is directed to (e.g., a composition comprising, consisting essentially of, consisting of) a fusion protein comprising all or a portion of a cas protein fused to all or a portion of an effector domain. In some aspects, all or a portion of the cas protein has endonuclease activity (e.g., can cleave and/or nick a target nucleic acid sequence) and/or targeting activity. In some aspects all or a portion of the Cas protein targets but does not cleave a nucleic acid sequence. In some aspects, the Cas protein can be fused to the N-terminus or C-terminus of the effector domain. In some aspects, the portion of the effector domain modulates the expression and/or activation of a target nucleic acid sequence (e.g., gene).

In some aspects, the nucleic acid sequence encoding the fusion protein and/or the fusion protein are isolated. An "isolated," "substantially pure," or "substantially pure and isolated" nucleic acid sequence, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleotide sequence (as
in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (e.g., as in an RNA or cDNA library). For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. An "isolated," "substantially pure," or "substantially pure and isolated" protein (e.g., chimeric protein; fusion protein), as used herein, is one that is separated from or substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system, or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example, as determined by agarose gel electrophoresis or column chromatography such as HPLC. Preferably, an isolated nucleic acid molecule comprises at least about 50%, 80%, 90%, 95%, 98% or 99% (on a molar basis) of all macromolecular species present.

[0093] "Modulate" is used consistently with its use in the art, i.e., meaning to cause or facilitate a qualitative or quantitative change, alteration, or modification in a process, pathway, or phenomenon of interest. Without limitation, such change may be an increase, decrease, or change in relative strength or activity of different components or branches of the process, pathway, or phenomenon. A "modulator" is an agent that causes or facilitates a qualitative or quantitative change, alteration, or modification in a process, pathway, or phenomenon of interest.

[0094] In some aspects, "modulating" ("modulates"; "modulation") the expression and/or activity of a target nucleic acid sequence refers to any of a variety of alterations to the expression and/or activation of the one or more target nucleic acid sequences. For example, the method of modulating the expression and/or activity of the one or more target nucleic acid sequences includes activating, increasing, decreasing, coactivating, regulating, repressing, organizing, remodeling, modifying, and/or fusing the expression and/or activity of one or more target nucleic acid sequences.
Thus, the one or more RNA sequences can be complementary to any of a variety of all or a portion of a target nucleic acid sequence that is to be modulated. In some aspects of the invention, the method of modulating one or more target nucleic acid sequences comprises introducing one or more RNA sequences that are complementary to all or a portion of a (one or more) regulatory region, an open reading frame (ORF; a splicing factor), an intronic sequence, a chromosomal region (e.g., telomere, centromere) of the one or more target nucleic acid sequences into a cell. In some aspects, the target nucleic acid sequence is all or a portion of a plasmid or linear double stranded DNA (dsDNA). In some aspects, the regulatory region targeted by the one or more target nucleic acid sequences is a promoter, enhancer, and/or operator region. In some aspects, all or a portion of the regulatory region is targeted by the one or more target nucleic acid sequences. In some aspects, the regulatory region targeted by the one or more target nucleic acid sequences is exactly or within about 25 bases, 50 bases, 100 bases, 200 bases, 300 bases, 400 bases, 500 bases, 600 bases, 700 bases, 800 bases, 900 bases, 1000 bases, 1500 bases, 2000 bases, or more upstream to the one or more genes (e.g., endogenous genes; exogenous genes) or a (one or more) transcription start site (TSS). In some aspects, the one or more target nucleic acid sequences is exactly or within about 25 bases, 50 bases, 100 bases, 200 bases, 300 bases, 400 bases, 500 bases, 600 bases, 700 bases, 800 bases, 900 bases, 1000 bases, 1500 bases, 2000 bases, or more downstream to the one or more genes (e.g., endogenous genes; exogenous genes) or a TSS. As will be appreciated by one of ordinary skill in the art, the regulatory region targeted by one or more target nucleic acid sequences can be entirely or partially found at or about the 5' end of the gene (e.g., endogenous or exogenous) or a TSS. The 5' end of a gene can include untranscribed (flanking) regions (e.g., all or a portion of a promoter) and a portion of the transcribed region.

As used herein, a "regulatory region" is any segment of a nucleic acid sequence capable of modulating (e.g. increasing, decreasing) expression and/or activity of one or more target nucleic acid sequences (e.g. genes). Examples of regulatory regions include a promoter, enhancer, telomere, locus control region, insulator, centromere, repeat sequence, transposable element, synthetic sequence, and operator. Specific examples of regulatory regions include CAAT box, CCAAT
box, Pribnow box, TATA box, SECIS element, Polyadenylation signals, A-box, Z-box, C-box, E-box, and/or G-box.

[0097] The method of modulating one or more target nucleic acid sequences comprises introducing a Cas nucleic acid sequence or a variant thereof that encodes the Cas protein that targets but does not cleave the target nucleic acid sequence into the cell. In some aspects, a Cas protein or variant thereof is introduced into the cell. In some aspects, the Cas nucleic acid sequence encodes a Cas protein that does not have endonuclease activity. In some aspects, the Cas nucleic acid sequence encodes a Cas protein that does not have nickase activity. In some aspects, the Cas nucleic acid sequence encodes a Cas protein that does not have endonuclease and nickase activity. In some aspects, the Cas nucleic acid sequence encodes a Cas protein that does not have enzymatic activity or is catalytically inactive.

[0098] In some aspects of the invention, the method of modulating one or more target nucleic acid sequences comprises introducing a Cas nucleic acid sequence or a variant thereof that encodes a Cas9 protein. In some aspects, the Cas nucleic acid sequence encodes a Cas9 protein that comprises one or more mutations. In some aspects, the Cas nucleic acid sequence encodes a Cas9 protein that comprises a mutation at amino acid position 10, 840, or a combination thereof. In some aspects, the Cas nucleic acid sequence encodes a Cas9 protein wherein the amino acid at position 10 is mutated from aspartate (D) to alanine (A) and the amino acid at position 840 is mutated from histidine (H) to alanine (A).

[0099] The method of modulating one or more target nucleic acid sequences also comprises introducing one or more effector domains. As used herein an "effector domain" is a molecule (e.g., protein) that modulates the expression and/or activation of a target nucleic acid sequence (e.g., gene). In some aspects, the effector domain targets one or both alleles of a gene. The effector domain can be introduced as a nucleic acid sequence and/or as a protein. In some aspects, the effector domain can be a constitutive or an inducible effector domain. In some aspects, a Cas nucleic acid sequence or variant thereof and an effector domain nucleic acid sequence are introduced into the cell as a chimeric sequence. In some aspects, the effector domain is fused to a molecule that associates with (e.g., binds to) Cas protein (e.g., the effector molecule is fused to an antibody or antigen binding fragment thereof that binds to Cas protein). In some aspects, a Cas protein or variant
thereof and an effector domain are fused or tethered creating a chimeric protein and are introduced into the cell as the chimeric protein. In some aspects, the Cas protein and effector domain bind as a protein-protein interaction. In some aspects, the Cas protein and effector domain are covalently linked. In some aspects, the effector domain associates non-covalently with the Cas protein. In some aspects, a Cas nucleic acid sequence and an effector domain nucleic acid sequence are introduced as separate sequences and/or proteins. In some aspects, the Cas protein and effector domain are not fused or tethered.

Examples of effector domains include a transcriptional activating domain (e.g., VP16, VP48, VP64, VP96 and VP160), a coactivator domain, a transcription factor, a transcriptional pause release factor domain, a negative regulator of transcriptional elongation domain, a transcriptional repressor domain, a chromatin organizer domain, a remodeler domain, a histone modifier domain, a DNA modification domain, a RNA binding domain, a protein interaction input devices domain (Grunberg and Serrano, *Nucleic Acids Research*, 38(8):2663-2675 (2010)), and a protein interaction output device domain (Grunberg and Serrano, *Nucleic Acids Research*, 38(8):2663-2675 (2010)). As used herein a "protein interaction input device" and a "protein interaction output device" refers to a protein-protein interaction (PPI). In some embodiments the PPI is regulatable, e.g., by a small molecule or by light. In some aspect, binding partners are targeted to different sites in the genome using the inactive Cas protein. The binding partners interact, thereby bringing the targeted loci into proximity. A protein interaction output device is a system for detecting/monitoring occurrence of a PPI, generally by producing a detectable signal when the PPI occurs (e.g., by reconstituting a fluorescent protein) or to trigger specific cellular responses (e.g., by reconstituting a caspase protein to induce apoptosis). The idea in this context is to target different sites in the genome with the components of the "output device". If the interaction occurs, the "output device" generates a signal. This can be used to determine or monitor the proximity of the targeted loci. In some aspects, cells are treated with an agent and the effect of the agent on the cell is determined. Other examples of effector domains include histone marks readers/interactors (http://www.cell.com/abstract/S0092-8674(10)00951-7) and DNA modification readers/interactors.
In some aspects, the effector domain is a VP16 effector domain. In some aspects, the effector domain is a VP48 effector domain. In some aspects, the effector domain is a VP64 effector domain. In some aspects, the effector domain is a VP96 effector domain. In some aspects, the effector domain is a VP160 effector domain.

In one aspect of the invention, fusion of the Cas9 to an effector domain can be to that of a single copy or multiple/tandem copies of full-length or partial-length effectors. Other fusions can be with split (functionally complementary) versions of the effector domains. Effectors domains for use in the methods include any one of the following classes of proteins: proteins that mediate drug inducible looping of DNA and/or contacts of genomic loci, proteins that aid in the three-dimensional proximity of genomic loci bound by dCas9 with different sgRNA.

Specific examples of transcription activators or coactivators include VP16, tandem copies comprising all or a biologically active portion of the activation peptide from VP16 (e.g. minimal transactivation domain), such as ADALDDFDLDMLP (SEQ ID NO: 125) and DALDDFDLDML (SEQ ID NO: 126), VP48 (e.g., 3 copies of VP16 minimal transactivation domain), VP64 (e.g., 4 copies of VP16 minimal TA), VP96 (e.g., 6 copies of VP16 minimal TA), VP160 (e.g., 10 copies of VP16 minimal TA), Brd4, and p65.

A specific example of a transcription factor is MYC.

Specific examples of transcriptional pause release factors include proteins in the PTEFb complex, such as Cyclin T1, Cyclin T2, Cyclin T3, Cdk9.

Specific examples of negative regulators of transcriptional elongation include negative elongation factor (NELF) components.

Specific examples of transcriptional repressors include engrailed (EnR), KRAB, Sin3-interaction domain (SID) and EMSY.

Specific examples of chromatin organizers and remodelers include insulator proteins, such as CTCF (transcriptional repressor CTCF or CCCTC-binding factor) to disrupt interactions between enhancers and promoters, cohesin complex and mediator complex Medl to activate gene expression, switch/sucrose nonfermentable (SWI/SNF) complex - INI1, BAF155b, BAF170, BRG1, hBRM to open up chromatin, and polycomb repressive complex to induce repressive domains on chromatin.
Specific examples of histone modifiers include histone acetyltransferases such as p300/EP300 (p300HAT), CBP/CREBBP (CBPHAT), MGEA5, CDYL, CLOCK, ELP3, GTF3C4, KAT2A, KAT2B, KAT5, MYST2, MYST3, MYST4, HAT1, NAT10, NCOA1, NCOA3, MYST1, CDY1B, CDY1; histone methyltransferases such as SET7, PRMT1, PRMT2, PRMT5, PRMT6, PRMT7, PRMT8, G9a, CARM1, MLL, Set2/SET1A, Ash2, Wdr5, Rbbp5, EZH1, EZH2, MLL2, MLL3, MLL4, MLL5, WHSC1L1, PRDM9, SETD1A, SETD1B, SETD2, SETD7, SETD8, SETDB1, SETDB2, SETMAR, SUV39H1, SUV39H2, SUV420H1, SUV420H2, NSD1, DOT1L, EHMT2, EHMT1, SMYD2, PRDM2, ASH1L, WHSC1, SMYD3; histone Deiminases such as PADI4; histone biotinases such as HLCS; histone ribosylases such as PARP1; histone ubiquitinases such as RNF20, RNF40, DTX3L, HUWE1, RBX1, RING1, RNF2, RNF168, RNF8, UBR2, UHRF1, RAG1; histone kinases such as CDK17, CDK3, CDK5, DAPK3, PRKDC, GSK3B, CHUK, LIMK2, MASTL, MAP3K8, MLT, BUB1, PRKCB, PRKCD, RPS6KA4, RPS6KA5, ATM, STK10, AURKB, STK4, ATR, GSG2, PKN1, NEK6, NEK9, PAK2, TLK1, BAZ1B, JAK2; histone demethylases such as Jarid1l, Rbr-2, JMJD6, PHF8, KDM2A, KDM2B, KDM3A, KDM3B, KDM4A, KDM4B, KDM4C, KDM4D, KDM5A, KDM5B, KDM5C, KDM5D, KDM6A, KDM6B, JHDM1D, JMJD5, C14orf189, KDM1A, KDM1B; histone deribosylases such as PARG; histone deubiquitinases such as MSYM1, BRCC3, USP16, USP22, USP3; histone phosphatases such as DUSP1, EYA1, EYA2, EYA3, PPM1D, PNP2CA, PPP2CB,PPP4C, PPP5C, PPIICC; histone deacetylases such as HDAC1, HDACIO, HDAC11, HDAC2, HDAC3, HDAC4, HDAC5, HDAC6, HDAC7, HDAC8, HDAC9, SIRT1, SIRT2, SIRT3, SIRT6.

Specific examples of DNA modifiers include 5hmC conversion from 5mC such as Tete (Tet1CD); DNA demethylation by Tet1, ACIDA, MBD4, Apobec1, Apobec2, Apobec3, Tdg, Gadd45a, Gadd45b, ROS1; DNA methylation by Dnmt1, Dnmt3a, Dnmt3b, CpG Methyltransferase M.SssI, and/or M.EcoHK31I.

Specific examples of RNA binding domains to bring RNA molecules to specific genomic loci include Rbfox2, CUG-BP, MBNL1, MBNL2, MBNL3, MS2 coat protein (MS2 hairpin), and engineered Pumilio.

Specific examples of protein interaction input devices (Grunberg and Serrano, Nucleic Acids Research, 38(8): 2663-2675 (2010)) to mediate drug
inducible looping of DNA and/or contact of genomic loci include rapamycin induced FKBP:FRB interaction, Jun:Fos, engineered variants of constitutive leuzine zipper interaction, and light-inducible PIF3:PhyB interaction.

Specific examples of protein interaction output devices (Grunberg and Serrano, *Nucleic Acids Research*, 38(8):2663-2675 (2010)) to report three-dimensional proximity of genomic loci bound by dCas9 with different sgRNA targeting different genomic loci includes split green fluorescent protein (GFP), Fluorescent Resonance Energy Transfer (FRET), split lactamase (antibiotic resistance-based selection) and split capase. These proteins can also be extended to a screening platform for proximal domains in chromatin with a library of sgRNA expression constructs.

Specific examples of histone marks readers/interactors include Sgf29, BPTF, C17orf49/BAP18, GATAD1, TRRAP, PHF8, N-PAC, MSH-6, and NSD1, NSD2, CBX1, CBX3, CBX5, CDYL, and CDYL2.

Specific examples of DNA modification readers/interactors include MeCP2, MBD1, MBD2, MBD3 MBD4, ZBTB4, ZBTB33, ZBTB38, UHRF1, and UHRF2.

In some aspects of the invention, the method of modulating one or more target nucleic acid sequences in a cell can further comprise introducing an effector molecule. As used herein, an "effector molecule" is a molecule (e.g., nucleic acid sequence; protein; organic molecule; inorganic molecule, small molecule) or physical trigger that associates with (e.g., binds to; specifically binds to) the effector domain to modulate the expression and/or activity of a target nucleic acid sequence (e.g., an inducer molecule; a trigger molecule). In some aspects, the effector molecule is a physical signal such as light (e.g., at one or more specific wavelengths; temperature (e.g., temperature-sensitivity); magnetism; stressor and the like. The effector molecule can be contacted with the cell and/or introduced into the cell (e.g., as a nucleic acid sequence or as protein sequence). In some embodiments, the effector molecule is endogenous. In other embodiments, the effector molecule is exogenous. For example, an exogenous effector molecule can be introduced to the cell. In some aspects, the effector molecule binds to the effector domain. In some aspects, the effector molecule is a nucleic acid, protein, drug, small organic molecule and derivatives/variants thereof. In some aspects of the invention, the
effector molecule is an antibiotic or derivatives/variants thereof. For example, the antibiotic is doxycycline. One of ordinary skill in the art can appreciate other types of antibiotics used, including but not limited to, tetracycline, ampicillin, puromycin, and neomycin. In some aspects, the effector molecule is rapamycin, tamoxifen and/or derivative/variants thereof (e.g., (Z)-4-hydroxytamoxifen).

[00117] As will be appreciated by those of skill in the art, the effector molecule can also associate with one or more domains (e.g., binding domains) that are fused to or associated with the effector domain. For example, the effector domain can be fused to or associated with a receptor domain and/or an antigen binding domain, and the effector molecule (e.g., a ligand specific to the receptor domain; an antibody specific to the antigen binding domain) can bind to the receptor domain and/or antigen binding domain which activates the effector domain, thereby modulating the expression and/or activity of the one or more target nucleic acid sequences.

[00118] As will be apparent to those of skill in the art, the method can further comprise introducing other molecules or factors into the cell to facilitate modulation of the activation and/or expression of the target nucleic acid sequence. Examples of such molecules include coactivators, chromatin remodelers, histone acetylases, deacetylases, kinases, and methylases. The methods described herein can also be used to silence expression of a nucleic acid sequence (e.g., a gene) by guiding a repressor to a target nucleic acid sequence.

[00119] A variety of target nucleic acid sequences can be mutated or modulated using the methods described herein and will depend upon the desired results. In one aspect, the target nucleic acid sequence is a gene sequence. In particular aspects, the methods described herein can be used to genetically modify two or more different genes in the same gene family, two or more genes that have a redundant function (e.g., redundant may mean that one needs to inactivate at least two of the genes to produce a particular phenotype, e.g., a detectable phenotype), two or more genes of which at least one gene does not or is believed not to produce detectable phenotype when inactivated (e.g., in the strain background used), two or more genes at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more identical, two or more copies of the same gene, two or more genes in same biological pathway (e.g., signaling pathway, metabolic pathway), two or more genes that share at least one biological activity and/or act on at least one common substrate and/or are part of the same
protein or protein-nucleic acid complex (e.g., a heteroligomeric protein, spliceosome, proteasome, RISC, transcription complex, replication complex, kinetochore, channel, transporter).

[00120] In some aspects, the target nucleic acid sequence is associated with a disease or condition (e.g., see van der Weyden et al, *Genome Biol*, 12;224 (2011)). Specific examples of genetic modifications of interest include modifying sequence(s), (e.g., gene(s)) to match sequence in different species (e.g., change mouse sequence to human sequence for any gene(s) of interest), alter sites of potential or known post-translational modification of proteins (e.g., phosphorylation, glycosylation, lipidation, acylation, acetylation), alter sites of potential or known epigenetic modification, alter sites of potential or known protein-protein or protein-nucleic acid interaction, inserting tag, e.g., epitope tag, and/or inserting or deleting splice sites. Other examples, include mutating a cell or nonhuman mammal to insert an epitope tag or transgene at an endogenous locus, make a reporter mouse, introduce loxP sites or FlpRT sites flanking certain genomic regions, and/or insert a cassette (e.g., a loxP-stop-loxP or FRT-stop-FRT cassette) in front of a gene to produce conditional alleles (e.g., see Frese and Tuveson, *Nature Rev*, 7:645-658 (2007); Nern et al., *PNAS*, 108(34): 14198-14203 (2011); Freidal et al., *Meth Molec Biol*, 693:205-231 (2011)).

[00121] In some aspects, one copy of the one or more target nucleic acid sequences is mutated. In some aspects, both copies of one or more of the target nucleic acid sequences in the stem cell or zygote are mutated. In some aspects, the one or more target nucleic acid sequences that are mutated are endogenous to the stem cell or zygote.

[00122] In particular aspects, at least two of the target nucleic acid sequences are endogenous nucleic acid sequences. In some aspects, at least two of the target nucleic acid sequences are exogenous nucleic acid sequences. In some aspects where there are at least two target nucleic acid sequences, at least one of the target nucleic acid sequences is an endogenous nucleic acid sequence and at least one of the target nucleic acid sequences is an exogenous nucleic acid sequence. In some aspects, at least two of the target nucleic acid sequences are endogenous genes. In some aspects, at least two of the target nucleic acid sequences are exogenous genes. In some aspects where there are at least two target nucleic acid sequences, at least
one of the target nucleic acid sequences is an endogenous gene and at least one of
the target nucleic acid sequences is an exogenous gene. In some aspects, at least
two of the target nucleic acid sequences are at least 1 kB apart. In some aspects, at
least two of the target nucleic acid sequences are on different chromosomes.

[00123] As used herein "mutate", "mutated" or "mutation" and the like refers to
alteration of a sequence (a target sequence). For example, in some aspects, a target
sequence that has been mutated refers to the replacement, introduction, and/or
deletion of one or more nucleotides in the target sequence. In some aspects, a target
sequence has been mutated to replace one or more nucleotides in the sequence with
one or more nucleotides that occur in one or more natural states of the sequence
(e.g., target sequence that is mutated with respect to a wild type sequence has been
mutated to replace one or more nucleotides in the sequence with one or more
nucleotides that occur in a wild type sequence). In some aspects, a target sequence
has been mutated to replace one or more nucleotides that occurs in one or more
natural states of the sequence (wild type) with one or more other nucleotides.

[00124] In particular aspects, at least one mutation comprises an insertion of a tag
(e.g., an epitope tag such as a V5 tag; a fluorescent tag), a transgene (e.g, a reporter
gene such as p2A-mCherry, GFP), a translation initiation site (e.g., IRES sequence),
a transcription initiation site (e.g., TATA box) and/or an insertion of a site
recognized by a recombinase (e.g., Cre). In some aspects, at least one mutation
renders expression of an endogenous gene conditional. In yet some aspects, at least
one mutation renders expression of an endogenous gene inducible, repressible, or
tissue-specific. In still some aspects, the mutations comprise inserting recombination
sites (e.g., loxP sites or FRT sites) flanking a selected genomic region, wherein the
selected genomic region is optionally within a gene. The mutations can also
comprise inserting a recombination-site-STOP-recombination site cassette (e.g., a
loxP-STOP-loxP or FRT-STOP-FRT cassette) in a gene, between a promoter and a
coding region of a gene, or in a regulatory region of a gene. In this aspect, the
recombination-site-STOP-recombination site cassette is positioned so as to disrupt
expression of the gene and wherein excision of the cassette by a recombinase
renders the gene expressible.

[00125] The methods provided herein provide for multiplexed genome editing in
cells, embryos, zygotes and nonhuman mammals. As shown herein, cells, embryos,
zygotes and non-human mammals carrying mutations in multiple genes can be generated in a single step. In some aspects, the methods described herein allow for the mutation of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, etc. nucleic acid sequences (e.g., genes) in a (single) cell, zygote, embryo or nonhuman mammal using the methods described herein. In a particular aspect, 1 nucleic acid sequence is mutated in a (single) cell, zygote, embryo or nonhuman mammal. In some aspects, 2 nucleic acid sequences are mutated in a (single) cell, zygote, embryo or nonhuman mammal. In some aspects, 3 nucleic acid sequences are mutated in a (single) cell, zygote, embryo or nonhuman mammal. In some aspects, 4 nucleic acid sequences are mutated in a (single) cell, zygote, embryo or nonhuman mammal. In some aspects, 5 nucleic acid sequences are mutated in a (single) cell, zygote, embryo or nonhuman mammal, etc.

[00126] The methods described herein can further comprising introducing one or more additional nucleic acid sequences that are complementary to a portion of the one or more target nucleic acid sequences cleaved by the Cas protein. A variety of nucleic acid sequences can be introduced, and include a single stranded oligonucleotide, a double stranded oligonucleotide, a plasmid, a cDNA, a gene block (e.g., gBlocks™ Gene Fragments (IDT)), a PCR product and the like. Thus, the size of the nucleic acid sequences can vary and will depend upon the reason for introducing the nucleic acid sequence. For example, the one or more nucleic acid sequences can be used to replace one or more nucleotides, introduce one or more additional nucleotides, delete one or more nucleotides or a combination thereof in the one or more target nucleic acid sequences. In a particular aspect, the one or more nucleic acid sequences introduce a point mutation in one or more of the target sequences. In some aspects, the one or more nucleic acid sequences replace one or more mutant nucleotides with one or more wild type nucleotides in one or more of the target sequences. In some aspects, the one or more nucleic acid sequences replace one or more wild type nucleotides with one or more (mutant) nucleotides in one or more of the target sequences. In some aspects, the one or more nucleic acids introduce a tag (e.g., a fluorescent protein such as green fluorescent protein), label and/or cleavage site. Thus, the nucleic acid sequence can be from about 10 nucleotides to about 5000 nucleotides, about 20 to 4500 nucleotides, about 30 to 4000 nucleotides, about 50 to 3500 nucleotides, about 60 to about 3000 nucleotides,
about 70 to about 2500 nucleotides, about 80 to about 2000 nucleotides, about 90 to about 1500 nucleotides, about 100 to about 1000 nucleotides, etc. In a particular aspect, the nucleic acid sequence is about 10 to about 500 nucleotides.

[00127] In a particular aspect, the nucleic acid sequence (e.g., oligonucleotide) is used to further modify (alter, edit, mutate) the cleaved target nucleic acid sequence (e.g., such oligo-mediated repair allows for precise genome editing,). Thus, this aspect allows for genome editing, however as shown herein the other allele is often mutated through nonhomologous end joining (NHEJ, see Fig 3B, 3C, and 8C. Shown herein is that using lower Cas mRNA concentration generates more mice with heterozygous mutations. Therefore, the methods provided herein allow for optimization of the system for more efficient generation of nonhuman mammals with only one oligo-modified allele. In some embodiments, employment of Cas nickase is desirable, since it mainly induces DNA single strand breaks, which is typically repaired through HDR (Cong et al, Science 339:8 19-823 (2013); Mali et al, Science, 339:823-826 (2013)).

[00128] As will be apparent to those of skill in the art, a variety of methods can be used to introduce nucleic acid and/or protein into a stem cell, zygote, embryo, and/or mammal. Suitable methods include calcium phosphate or lipid-mediated transfection, electroporation, injection, and transduction or infection using a vector (e.g., a viral vector such as an adenoviral vector). In some aspects, the nucleic acid and/or protein is complexed with a vehicle, e.g., a cationic vehicle, that facilitates uptake of the nucleic acid and/or protein, e.g., via endocytosis.

[00129] The method described herein can further comprise isolating the stem cell or zygote produced by the methods. Thus, in some aspects, the invention is directed to a stem cell or zygote (an isolated stem cell or zygote) produced by the methods described herein. In some aspects, the disclosure provides a clonal population of cells harboring the mutation(s), replicating cultures comprising cells harboring the mutation(s) and cells isolated from the generated animals.

[00130] The methods described herein can further comprise crossing the generated animals with other animals harboring genetic modifications (optionally in same strain background) and/or having one or more phenotypes of interest (e.g., disease susceptibility - such as NOD mice). In addition, the methods may comprise modifying a stem cell, zygote, and/or animal from a strain that harbors one or more
genetic modifications and/or has one or more phenotypes of interest (e.g., disease susceptibility).

[00131] In some aspects, various mouse strains and mouse models of human disease are used. One of ordinary skill in the art appreciates the thousands of commercially and non-commercially available strains of laboratory mice for modeling human disease. Mice models exist for diseases such as cancer, cardiovascular disease, autoimmune, inflammatory, diabetes (type 1 and 2), neurobiology, and other diseases. Examples of commercially available research strains include, and is not limited to, 11BHSD2 Mouse, GSK3B Mouse, 129-E Mouse HSD11B1 Mouse, AKR Mouse Immortomouse®, Athymic Nude Mouse, LCAT Mouse, B6 Albino Mouse, Lox-1 Mouse, B6C3F1 Mouse, Ly5 Mouse, B6D2F1 (BDF1) Mouse, MMP9 Mouse, BALB/c Mouse, NIH-III Nude Mouse, BALB/c Nude Mouse, NOD Mouse, NOD SCID Mouse, Black Swiss Mouse, NSE-p25 Mouse, C3H Mouse, NU/NUNude Mouse, C57BL/6-E Mouse, PCSK9 Mouse, C57BL/6N Mouse, PGP Mouse (P-glycoprotein Deficient), CB6F1 Mouse, repTOP™ERE-Luc Mouse, CD-I® Mouse, repTOP™mitoIRE Mouse, CD-I®Nude Mouse, repTOP™PPRE-Luc Mouse, CD1-E Mouse, Rip-HAT Mouse, CD2F1 (CDF1) Mouse, SCID Hairless Congenic (SHC™) Mouse, CF-1™ Mouse, SCID Hairless Outbred (SHO™) Mouse, DBA/2 Mouse, SJL-E Mouse, Fox Chase CB17™ Mouse, SKH1-E Mouse, Fox Chase SCID® Beige Mouse, Swiss Webster (CFW®) Mouse, Fox Chase SCID® Mouse, TARGATT™ Mouse, FVB Mouse, THE POUND MOUSE™, and GLUT 4 Mouse. Other mouse strains include BALB/c, C57BL/6, C57BL/10, C3H, ICR, CBA, A/J, NOD, DBA/1, DBA/2, MOLD, 129, HRS, MRL, NZB, NIH, AKR, SJL, NZW, CAST, KK, SENCAR, C57L, SAMR1, SAMP1, C57BR, and NZO.

[00132] The methods described herein can further comprise assessing whether the one or more target nucleic acids have been mutated and/or modulated using a variety of known methods.

[00133] In some embodiments methods described herein are used to produce multiple genetic modifications in a stem cell, zygote, embryo, or animal, wherein at least one of the genetic modifications knocks out (functionally inactivates completely or partially) a gene whose knockout does not produce a detectable phenotype, and at least one of the genetic modifications is in a different gene or
genomic location. The resulting stem cell, zygote, embryo, or animal, or a cell, zygote, embryo, or animal generated therefrom, is analyzed for the presence of one or more detectable phenotypes. Such methods may be used to identify genes or genomic locations that have synthetic effects (e.g., effects that are greater in degree or different in kind from the sum of the effects caused by either mutation alone). In some embodiments an effect is synthetic lethality. In some embodiments at least one of the genetic modifications may be conditional (e.g., the effect of the modification, such as gene knockout, only becomes manifest under certain conditions, which are typically under control of the artisan). In some embodiments animals are permitted to develop at least to post-natal stage, e.g., to adult stage. The appropriate conditions for the modification to produce an effect (sometimes termed "inducing conditions") are imposed, and the phenotype of the animal is subsequently analyzed. A phenotype may be compared to that of an unmodified animal or to the phenotype prior to the imposition of the inducing conditions.

[00134] In any aspect or embodiment herein, analysis may comprise any type of phenotypic analysis known in the art, e.g., examination of the structure, size, development, weight, or function, of any tissue, organ, or organ system (or the entire organism), analysis of behavior, activity of any biological pathway or process, level of any particular substance or gene product, etc. In some embodiments analysis comprises gene expression analysis, e.g., at the level of mRNA or protein. In some embodiments such analysis may comprise, e.g., use of microarrays (e.g., oligonucleotide microarrays, sometimes termed "chips"), high throughput sequencing (e.g., RNASeq), ChIP on Chip analysis, ChIPSeq analysis, etc. In some embodiments high content screening may be used, in which elements of high throughput screening may applied to the analysis of individual cells through the use of automated microscopy and image analysis (see, e.g., Zanella et al, 2010). High content screening: seeing is believing. Trends Biotechnol. 28:237-245). In some embodiments analysis comprises quantitative analyses of components of cells such as spatio-temporal distributions of individual proteins, cytoskeletal structures, vesicles, and organelles, e.g., when contacted with test agents, e.g., chemical compounds. In some embodiments activation or inhibition of individual proteins and protein-protein interactions and/or changes in biological processes and cell functions may be assessed. A range of fluorescent probes for biological processes,
functions, and cell components are available and may be used, e.g., with fluorescence microscopy. In some embodiments cells or animals generated according to methods herein may comprise a reporter, e.g., a fluorescent reporter or enzyme (e.g., a luciferase such as Gaussia, Renilla, or firefly luciferase) that, for example, reports on the expression or activity of particular genes. Such reporter may be fused to a protein, so that the protein or its activity is rendered detectable, optionally using a non-invasive detection means, e.g., an imaging or detection means such as PET imaging, MRI, fluorescence detection. Multiplexed genome editing according to the invention may allow installation of reporters for detection of multiple proteins, e.g., 2 - 20 different proteins, e.g., in a cell, tissue, organ, or animal, e.g., in a living animal.

[00135] Multiplexed genome editing according to the present invention may be useful to determine or examine the biological role(s) and/or roles in disease of genes of unknown function (e.g., genes whose complete knockout does not produce a detectable phenotype). For example, discovery of synthetic effects caused by mutations in first and second genes may pinpoint a genetic or biochemical pathway in which such gene(s) or encoded gene product(s) is involved. In some embodiments mutations may be generated in stem cells or zygotes from any existing knockout or deletion strain or animals produced according to methods described herein may be crossed with animals from such strain. In some embodiments one or more gain-of-function and/or loss-of-function alleles are generated.

[00136] In some embodiments it is contemplated to use, in methods described herein, cells or zygotes generated in or derived from animals produced in projects such as the International Knockout Mouse Consortium (IKMC), the website of which is http://www.knockoutmouse.org). In some embodiments it is contemplated to cross animals generated as described herein with animals generated by or available through the IKMC. For example, in some embodiments a mouse gene to be modified according to methods described herein is any gene from the Mouse Genome Informatics (MGI) database for which sequences and genome coordinates are available, e.g., any gene predicted by the NCBI, Ensembl, and Vega (Vertebrate Genome Annotation) pipelines for mouse Genome Build 37 (NCBI) or Genome Reference Consortium GRCm38.
In some embodiments a gene or genomic location to be modified is included in genome of a species for which a fully sequenced genome exists. Genome sequences may be obtained, e.g., from the UCSC Genome Browser (http://genome.ucsc.edu/index.html). For example, in some embodiments a human gene or sequence to be modified according to methods described herein may be found in Human Genome Build hg19 (Genome Reference Consortium). In some embodiments a gene is any gene for which a Gene ID has been assigned in the Gene Database of the NCBI (http://www.ncbi.nlm.nih.gov/gene). In some embodiments a gene is any gene for which a genomic, cDNA, mRNA, or encoded gene product (e.g., protein) sequence is available in a database such as any of those available at the National Center for Biotechnology Information (www.ncbi.nih.gov) or Universal Protein Resource (www.uniprot.org). Databases include, e.g., GenBank, RefSeq, Gene, UniProtKB/SwissProt, UniProtKB/Trembl, and the like.

In some embodiments a gene encodes a polypeptide. In some embodiments a gene may not encode a polypeptide. A gene may, for example, comprise a template for transcription of a functional RNA, i.e., an RNA that has at least one function other than providing a messenger RNA (mRNA) to be translated into protein. Examples, include, e.g., long non-coding RNA (e.g., greater than 200 bases in length, e.g., 200 - 5,000 bases), small RNA (e.g., small nuclear RNA), transfer RNA, ribosomal RNA, microRNA precursor, Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs). In some embodiments a small RNA is 25 bases or less, 50 bases or less, 100 bases or less, 200 bases or less in length. Sequences of functional RNAs are available, e.g., from databases such as miRBase (website is http://www.mirbase.org/) (Kozomara A, et al., miRBase: integrating microRNA annotation and deep-sequencing data. NAR 2011 39(Database Issue):D152-D157), or the Long Non-Coding RNA Database, also called IncRNAdb (website is http://www.lncrnadb.org/), (Amaral PP, et al. (2011) IncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res 39: D146-151). In some embodiments a genomic sequence may be suspected of potentially comprising a template for transcription of a functional RNA. A genetic modification may be made in the sequence to determine whether such genetic modification alters the phenotype of a cell or animal or affects production of an RNA or protein or alters susceptibility to a disease.
In some embodiments it is of interest to genetically modify a known or suspected regulatory region, e.g., a known or suspected enhancer region or a known or suspected promoter region. The effect on expression of one or more genes in (e.g., within up to about 1, 2, 5, 10, 20, 50, 100, 500 kB or within about 1, 2, 5, or 10 MB from the modification) may be assessed. A genetic modification may be made in the sequence to determine whether such genetic modification alters the phenotype of a cell or animal or affects production of an RNA or protein or alters susceptibility to a disease.

In some embodiments any method described herein may comprise isolating one or more cells, samples, or substances from an animal generated according to methods described herein, e.g., any genetically modified animal generated as described herein. In some embodiments a method may further comprise analyzing the one or more cells, samples, or substances. Such analysis may, for example assess the effect of a genetic medication(s) introduced according to the methods.

In some embodiments animals generated according to methods described herein may be useful in the identification of candidate agents for treatment of disease and/or for testing agents for potential toxicity or side effects. In some embodiments any method described herein may comprise contacting an animal generated according to methods described herein, e.g., any genetically modified animal generated as described herein, with a test agent (e.g., a small molecule, nucleic acid, polypeptide, lipid, etc.). In some embodiments contacting comprises administering the test agent. Administration may be by any route (e.g., oral, intravenous, intraperitoneal, gavage, topical, transdermal, intramuscular, enteral, subcutaneous), may be systemic or local, may include any dose (e.g., from about 0.01 mg/kg to about 500 mg/kg), may involve a single dose or multiple doses. In some embodiments a method may further comprise analyzing the animal. Such analysis may, for example assess the effect of the test agent in an animal having a genetic medication(s) introduced according to the methods. In some embodiments a test agent that reduces or enhances an effect of one or more genetic modification(s) may be identified. In some embodiments if a test agent reduces or inhibits development of a disease associated with or produced by the genetic modification(s), (or reduces or inhibits one or more symptoms or signs of such a
disease) the test agent may be identified as a candidate agent for treatment of a
disease associated with or produced by the genetic modification(s) or associated
with or produced by naturally occurring mutations in a gene or genomic location
harboring the genetic modification.

[00142] In some embodiments a cell (e.g., a somatic cell to be used to generate an
iPS cell) may be a diseased cell or may originate from a subject suffering from a
disease, e.g., a disease affecting the cell or organ from which the cell was obtained.
In some embodiments a mutation is introduced into a genomic region of the iPS cell
that is associated with a disease (e.g., any disease of interest, such as diseases
mentioned herein). For example, in some embodiments it is of interest to knock out
or otherwise modify a gene or genomic location that is known or suspected to be
involved in disease pathogenesis and/or known or suspected to be associated with
increased or decreased risk of developing a disease or particular manifestation(s) of
a disease. In some embodiments it is of interest to knock out or otherwise modify a
gene or genomic location and determine whether such knockout or modification
alters the risk of developing a disease or one or more manifestations of a disease,
alters progression of the disease, or alters the response of a subject to therapy or
candidate therapy for a disease. In some embodiments it is of interest to modify an
abnormal or disease-associated nucleotide or sequence to one that is normal or not
associated with disease. In some embodiments this may allow production of
genetically matched cells or cell lines (e.g., iPS cells or cell lines) that differ only at
one or more selected sites of genetic modification. Multiplexed genome editing as
described herein may allow for production of cells or cell lines that are isogenic
except with regard to, e.g., between 2 and 20 selected sites or genetic alterations.
This may allow for the study of the combined effect of multiple mutations that are
suspected of or known to play a role in disease risk, development or progression.

[00143] The methods of modulating the expression and/or activity of one or more
target nucleic acid sequences in a cell have a variety of uses (e.g., therapeutic,
pharmaceutical and/or academic uses). For example, CRISPRzymes can be designed
to target specific chromatin loci to exert modification (e.g., methylation or
demethylation) on causative genes of diseases due to aberrant chromatin state to
correct the chromatin states. In addition, CRISPRzymes can be used to detect/sense
certain sequence variation or chromatin states at defined loci guided by sgRNA, or
interactions between genomic loci guided by pairs or set of sgRNAs and to exert specific therapeutic outcomes dependent on chromatin state or the interaction of genomic loci.

[00144] For example, split fragments of Caspase can be fused to dCas9 and only reconstitute apoptosis-inducing activity when two genomic loci targeted by specific sgRNAs are proximal due to looping under certain disease conditions or cell types, e.g., cancer stem cells. [http://www.ncbi.nlm.nih.gov/pubmed/22070901]. CRISPRzymes can be coupled with biosensors to kill cells on detecting specific histone or DNA modifications at specific loci, e.g., DNA methylation (http://www.ncbi.nlm.nih.gov/pubmed/21797230). A pair of fusions: CRISPR-CaspaseA, MBD1-CaspaseB fusion. MBD1-CaspaseB binds to mCpG, CRISPR-CaspaseA binds to a genomic loci (e.g., hypermethylated genes in cancer) guided by an sgRNA. Only at that defined loci and when the loci is methylated is the Caspase reconstituted, and triggering the killing of cancer cells but not in normal cells. CRISPRzymes can be used to detect chromosomal translocation events resulting in fusion of DNA fragments. dCas9 can be fused to split fragments of fluorescent marker, or luciferase gene and sgRNA targeting the fused genes are used and only when the two specific gene fragments are fused is the reporter reconstituted. This strategy can be used to screen for/detection of subtypes of cancer cells in patient samples/biopsy, at single cell resolution. Similarly fusion with split caspase will allow specific killing/depletion of aberrant cells characteristic of specific chromosomal translocation events. Conversely, CRISPRzymes can be used to restore DNA looping in patients with deficient DNA looping, e.g., Cornelia de Lange patients (defeats in cohesin complex.)

[00145] CRISPRzymes can also be used in pharmaceutical and/or academic research. For example, a screen can be used by a library of sgRNA sequences in combination with a CRISPRzyme or a set of CRISPRzymes. The screen can be in the format of library, where each samples (cells, embryos, or tissues) are treated with known and predefined sgRNA or a set of sgRNA. Alternatively, the screen can be pooled whereby vectors expressing different sgRNAs are mixed and introduced to the target (cells, embryos, tissues, etc.) and cells with appropriate phenotype are selected or enriched and the sgRNA harboring the specific phenotype identified by sequencing. CRISPRzymes can be used to elicit chromatin state changes, or
transcription activation of specific gene or specific sets of genes in somatic cell,
adult stem cells or embryonic stem cells to induce them to reprogram into
pluripotent states, to differentiate or transdifferentiate.

[00146] In some aspects, methods described herein may be used to produce non-
human mammals that have a mutation in the SRY (sex determining region Y) gene. The SRY gene is an intronless gene located on the Y chromosome in therian mammals that encodes a transcription factor that is a member of the SOX (SRY-like box) gene family of DNA-binding proteins. Since a functional Sry protein is required for male development, a mammal that has an X and Y chromosome, wherein the Y chromosome harbors a loss-of-function mutation in SRY, is an anatomic female. An anatomic female may be recognized, e.g., by the presence of a uterus and ovaries and the absence of testes.

[00147] As described herein, the CRISPR/Cas system may be used to generate mutations in SRY, e.g., in a stem cell, zygote, or embryo. Thus in some embodiments, a target nucleic acid sequence mutated according to methods described herein is the SRY gene or a portion thereof. In some embodiments the mutation is a loss-of-function mutation. In some embodiments the loss-of-function mutation is a deletion of part or all of the SRY gene. In some embodiments the mutation, e.g., deletion, is in a portion of the gene that is essential for its function. In some embodiments a mutation is in the portion of the SRY gene that encodes the high mobility group (HMG) DNA binding domain of Sry, termed the HMG box. The HMG box (Nasrin, Nature, 354, 317-320 (1991)) is the characteristic domain of the SOX (SRY-type HMG box) family of transcription factors. It is a 79 amino acid domain that is highly conserved among SRY proteins (at least 50% identical to the human Sry HMG box). In humans, the HMG box extends from amino acid 58 to amino acid 137 of Sry. The corresponding sequences in other species are immediately evident upon aligning the Sry protein sequences with the human sequence (see, e.g., Fig. 15A). For example, in mouse the Sry HMG box extends from amino acid 3 to amino acid 82. The HMG domain is essential for the function of SRY proteins.

[00148] In some aspects, the present disclosure relates to the recognition that targeted mutations in SRY cause anatomic sex reversal, resulting in non-human mammals that have X and Y chromosomes but are anatomic females. For example,
Applicants have generated XY mice having a variety of deletions or insertion in the SRY gene (Wang H, et al., TALEN-mediated editing of the mouse Y chromosome. Nat Biotechnol. 2013; May 12, doi: 10.1038/nbt.2595. ePub ahead of print, incorporated herein by reference). The mice were generated using transcription activator-like effector nuclease (TALEN) technology to mutate the Sry gene in mouse ES cells. Two pairs of TALENs were generated to target the high mobility group (HMG) DNA binding domain of Sry and were transfected into mouse embryonic stem (ES) cells to generate deletions. TALEN pairs 1 and 2 showed gene modification efficiencies of 15% and 20%+, respectively, based on a Surveyor assay. The deletions ranged in size from 11 to 540 bp (Wang, H., supra). Three of the generated deletions are depicted schematically in Figure 15B. The TALEN cleavage site is in the middle between the binding of the TALEN2 pair as depicted in Figure 15B. The mutated ES cells were used to produce living mice by tetraploid complementation using standard methods. The resulting mice were found to be anatomic females. In addition, the insertion of a sequence encoding GFP at the same site lead to sex reversal. Adult Sry-targeted mice (anatomic females) showed reduced fertility, but they were fertile and transmitted the Sry-mutated Y chromosome to offspring.

<table>
<thead>
<tr>
<th>TALEN</th>
<th>TALEN recognition sequences (bold and underlined) &amp; Amino acid sequences of the RVDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sry pair 1</td>
<td>5’ TGGCCAGCAGAATCCAGATGCAAATACTACAGAGATCAAGCAG (SEQ ID NO: 127) 3’ ACCGGGTGCTTCAGTTGTCATACGTTTTATGTCCTCTAGTCGTTCG (SEQ ID NO: 128)</td>
</tr>
<tr>
<td></td>
<td>NG NN NG HD HD HD NI NN HD NI NN NI NG (Left TALEN) (SEQ ID NO: 129)</td>
</tr>
<tr>
<td></td>
<td>NN HD NG NGN HD NG NN NI NG HD NG HD NG NG NG (Right TALEN) (SEQ ID NO: 130)</td>
</tr>
<tr>
<td>Sry pair 2</td>
<td>5’ GAAGCATATCATGTGCTCCCAGGAGGCACAAATGGCTTGCCCAGG (SEQ ID NO: 131) 3’ CTTCAGTAAATACCCACACAGGGCCACACACTCTCCGTTGCTTCAACCGGGTCG (SEQ ID NO: 132)</td>
</tr>
<tr>
<td></td>
<td>NN NI NI NG NN HD NI NG NG NI NG NN NN NG NN NG (SEQ ID NO: 133)</td>
</tr>
<tr>
<td></td>
<td>NN HD NG NN NN NN HD HD NI NI HD NG NG NG NN HD HD NG (SEQ ID NO: 134)</td>
</tr>
</tbody>
</table>
The distributions of genotypes and anatomic sexual phenotypes in progeny from six litters.

<table>
<thead>
<tr>
<th>Parents</th>
<th>$XY^{Sry(tml)}$;Tg(Sry)2Ei</th>
<th>$XY^{Sry(tml)}$;Tg(Sry)2Ei</th>
<th>XX; Tg(Sry)2Ei</th>
<th>XX; Tg(Sry)2Ei</th>
<th>XX; Tg(Sry)2Ei</th>
<th>XX; Tg(Sry)2Ei</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progeny</td>
<td>Genotype</td>
<td>Genotype</td>
<td>Male</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
<td>$XY^{Sry(dllRlb)}$; Tg(Sry)2Ei</td>
<td>$XY^{Sry(tml)}$; Tg(Sry)2Ei</td>
<td>Male</td>
<td>Female</td>
<td>1a</td>
<td>2ab</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>Number</td>
<td>(total=22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the age of ~2 months, each of seven $XY^{Sry(tml)}$ females was housed with a single $XY^{Sry(dllRlb)}$; Tg(Sry)2Ei male for 5-7 months. The result was that three $XY^{Sry(tml)}$ females gave birth to a total of eight litters (two eaten at birth). It has been reported that, in XY female meiosis, the X and Y chromosomes do not pair efficiently and segregate randomly, leading to sex chromosome aneuploidy in the offspring of XY females 1, 2. aThese mice may carry either one or two X chromosomes. bThese mice may also carry $YSry(dllRlb)$.

[00149] In some embodiments the portion of the SRY gene that is targeted is within or overlaps with the portion of the gene that encodes the HMG box. In some embodiments the mutation removes at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 40, 50, 100, or more nucleotides from the gene, e.g., at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 40, 50, 100, or more nucleotides from the portion of the gene that encodes the HMG box. In some embodiments the mutation is in a portion of the gene upstream (5') of the region that encodes the HMG box, e.g., k encoding a portion of the Sry protein that lies N-terminal to the HMG box. In some embodiments a mutation is an insertion upstream of or within the sequence that encodes the HMG box, wherein the insertion results in a frameshift or stop codon.
For example, insertion of 1 or 2 amino acids or a longer sequence not divisible by 3 would result in a frameshift. Insertion of a stop codon in the region located 5' of the sequence encoding the HMG box would result in a truncated and nonfunctional Sry protein. In some embodiments a mutation may be located in a portion of the SRY gene that encodes a portion of Sry that is C-terminal to the HMG box. In some embodiments a mutation may be in a regulatory region, e.g., a promoter. In some embodiments a mutation may be upstream of the start codon, e.g., in a promoter.

[00150] In some embodiments the SRY gene is mutated in a zygote, and the zygote is transferred to the uterus of a foster mother (e.g., a pseudopregnant female) to develop to birth. It will be understood that the zygote may be maintained in culture after mutation of the SRY gene, e.g., to an early embryonic stage (e.g., a blastocyst) and then transferred to the uterus of a foster mother. In some embodiments, the invention provides a zygote having an X and Y chromosome, wherein the Y chromosome has an engineered mutation in the SRY gene, wherein the zygote is capable of developing to an anatomic female. The mammal may be any non-human mammal. In some embodiments a method comprises generating a non-human mammal that has an X chromosome and a Y chromosome (i.e., somatic cells of which contain an X and a Y chromosome).

[00151] Methods of creating anatomic females may be useful in any context in which it is desired to reduce the number or proportion of male offspring and/or increase the number of proportion of anatomically female offspring. In some embodiments, methods of generating anatomic females are useful in animal husbandry, which generally refers to the breeding and raising of non-human animals for any of a variety of purposes, e.g., for meat, as sources of animal products (e.g., milk, wool, hair, leather, skin, horn, eggs, or meat), for performing work, or providing companionship, e.g., as pets. In some embodiments it may be of interest to generate anatomic females which may be capable of producing offspring or serving as foster mothers for offspring of that species or producing a product of interest. In some embodiments the non-human mammal is allowed to develop at least until adulthood. In some embodiments the adult non-human mammal gives rise to offspring, which inherit the mutation. In some embodiments a useful product, e.g., milk, wool, hair, leather, skin, horn, or meat, is obtained from the anatomi
cally female non-human mammal.
In the context of dairy farming there is considerable interest in reducing the number of male offspring, as they are not useful for producing milk. In some embodiments, a non-human mammal useful in dairy farming is a cow, goat, sheep, or camel, or other non-human animal useful for the production of milk. In some embodiments a cow is of any of the following breeds: a Holstein (also referred to as Holstein-Friesian), Brown Swiss, Canadienne, Dutch Belted, Guernsey, Ayrshire, Jersey, Kerry, Milking Shorthorn, Milking Devon, or Norwegian Red.

In some embodiments methods of creating anatomic females may be useful in the context of managing species at risk of extinction, e.g., in programs that attempt to maintain or increase the number of individuals of a particular species. In some embodiments a species at risk of extinction may be any species recognized as near threatened, threatened (vulnerable, endangered, or critically endangered), or extinct in the wild by the International Union of Conservation (IUCN). Such species are listed, e.g., on the IUCN Red List of Threatened Species (also known as the IUCN Red List or Red Data List), e.g., the 2012 version (available at the IUCN website at http://www.iucnredlist.org/). In some embodiments the population of a species at risk of extinction may be declining. In some embodiments a species, e.g., a species at risk of extinction, may be, e.g., a bear, canine, caprine, elephant, feline, non-human primate, ovine, rodent, or ungulate species. In some embodiments a species, e.g., a species at risk of extinction, may be a marsupial, e.g., a Tasmanian Devil.

In some embodiments, methods of generating non-human mammals may comprise mutating one or more genes whose mutation results in a phenotype of interest. In some embodiments both copies of the gene are mutated. A phenotype of interest may be any phenotype, e.g., any property of interest. In some embodiments the non-human mammal is a source of food (e.g., milk or meat) or other products useful for humans. In some embodiments at least some humans may be allergic to a component, e.g., a protein, found in the food. A phenotype of interest may comprise reduced or absent production of an allergenic component, or alteration in an allergenic component so as to reduce its allergenicity. For example, in some embodiments the gene encoding a whey protein, e.g., the whey protein beta-lactoglobulin (BLG), a component found in the milk of cows, sheep, and a variety of other species (but not humans) that constitutes a major milk allergen, is mutated. In
some embodiments a gene is mutated so as to remove an allergenic epitope or alter it to a non-allergenic form, e.g., by changing or deleting one or more amino acids. The protein may still be produced and able to fulfill its normal function but is no longer allergenic or has reduced allergenicity to humans. In some embodiments a gene is mutated so as to reduce or eliminate production of the protein. In some embodiments a mutation is insertion of a stop codon or deletion or alteration of a start codon or at least a portion of a promoter.

[00155] In some embodiments a phenotype of interest may comprise any alteration that qualitatively or quantitatively alters one or more characteristics of a product that is obtained from the non-human mammal, e.g., in a way that makes the product more useful, easier to manipulate, less allergenic, or improved in any way. In some embodiments a characteristic may be color, texture, flavor, consistency, viscosity, thickness, roughness, toughness, tenderness, stringiness, fat content, protein content, sugar content, etc. In some embodiments a phenotype of interest may comprise any alteration that increases the yield of a product (e.g., on a per animal basis, per month or year basis); increases the growth rate; reduces the amount of food, resources, or care consumed or required by the animal; renders the animal more resistant to disease; renders the animal more tolerant of high or low temperature, or reduces the environmental impact of the animal (e.g., reduces methane production). In some embodiments, a phenotype may comprise increased milk production.

[00156] In some embodiments a polymorphism, e.g., a single nucleotide polymorphism, may be identified as being associated with a phenotype of interest using methods known in the art (e.g., genetic association studies). Methods described herein may be used to generate non-human mammals having a polymorphism that is associated with the phenotype. The animal may be compared with an otherwise isogenic animal that has not been genetically modified. The effect specifically due to variation at the polymorphic position may be determined. If a mutation or polymorphism confers a phenotype of interest, the non-human mammal may be used as a source of additional animals having the mutation or polymorphism and/or additional mammals having the mutation or polymorphism may be produced using methods described herein.
In some embodiments, methods of generating anatomically female non-human mammals may comprise mutating one or more additional nucleic acids in addition to the SRY gene. For example, any gene the mutation of which results in a phenotype of interest (e.g., reduced allergen content), may be mutated.

The terms "disease", "disorder" or "condition" are used interchangeably and may refer to any alteration from a state of health and/or normal functioning of an organism, e.g., an abnormality of the body or mind that causes pain, discomfort, dysfunction, distress, degeneration, or death to the individual afflicted. Diseases include any disease known to those of ordinary skill in the art. In some embodiments a disease is a chronic disease, e.g., it typically lasts or has lasted for at least 3-6 months, or more, e.g., 1, 2, 3, 5, 10 or more years, or indefinitely. Disease may have a characteristic set of symptoms and/or signs that occur commonly in individuals suffering from the disease. Diseases and methods of diagnosis and treatment thereof are described in standard medical textbooks such as Longo, D., et al. (eds.), Harrison's Principles of Internal Medicine, 18th Edition; McGraw-Hill Professional, 2011 and/or Goldman's Cecil Medicine, Saunders; 24 edition (August 5, 2011). In certain embodiments a disease is a multigenic disorder (also referred to as complex, multifactorial, or polygenic disorder). Such diseases may be associated with the effects of multiple genes, sometimes in combination with environmental factors (e.g., exposure to particular physical or chemical agents or biological agents such as viruses, lifestyle factors such as diet, smoking, etc.). A multigenic disorder may be any disease for which it is known or suspected that multiple genes (e.g., particular alleles of such genes, particular polymorphisms in such genes) may contribute to risk of developing the disease and/or may contribute to the way the disease manifests (e.g., its severity, age of onset, rate of progression, etc.) In some embodiments a multigenic disease is a disease that has a genetic component as shown by familial aggregation (occurs more commonly in certain families than in the general population) but does not follow Mendelian laws of inheritance, e.g., the disease does not clearly follow a dominant, recessive, X-linked, or Y-linked inheritance pattern. In some embodiments a multigenic disease is one that is not typically controlled by variants of large effect in a single gene (as is the case with Mendelian disorders). In some embodiments a multigenic disease may occur in familial form and sporadically. Examples include, e.g., Parkinson's disease,
Alzheimer's disease, and various types of cancer. Examples of multigenic diseases include many common diseases such as hypertension, diabetes mellitus (e.g., type II diabetes mellitus), cardiovascular disease, cancer, and stroke (ischemic, hemorrhagic). In some embodiments a disease, e.g., a multigenic disease is a psychiatric, neurological, neurodevelopmental disease, neurodegenerative disease, cardiovascular disease, autoimmune disease, cancer, metabolic disease, or respiratory disease. In some embodiments at least one gene is implicated in a familial form of a multigenic disease.

[00159] In some embodiments a disease is cancer, which term is generally used interchangeably to refer to a disease characterized by one or more tumors, e.g., one or more malignant or potentially malignant tumors. The term "tumor" as used herein encompasses abnormal growths comprising aberrantly proliferating cells. As known in the art, tumors are typically characterized by excessive cell proliferation that is not appropriately regulated (e.g., that does not respond normally to physiological influences and signals that would ordinarily constrain proliferation) and may exhibit one or more of the following properties: dysplasia (e.g., lack of normal cell differentiation, resulting in an increased number or proportion of immature cells); anaplasia (e.g., greater loss of differentiation, more loss of structural organization, cellular pleomorphism, abnormalities such as large, hyperchromatic nuclei, high nuclearcytoplasmic ratio, atypical mitoses, etc.); invasion of adjacent tissues (e.g., breaching a basement membrane); and/or metastasis. Malignant tumors have a tendency for sustained growth and an ability to spread, e.g., to invade locally and/or metastasize regionally and/or to distant locations, whereas benign tumors often remain localized at the site of origin and are often self-limiting in terms of growth. The term "tumor" includes malignant solid tumors, e.g., carcinomas (cancers arising from epithelial cells), sarcomas (cancers arising from cells of mesenchymal origin), and malignant growths in which there may be no detectable solid tumor mass (e.g., certain hematologic malignancies). Cancer includes, but is not limited to: breast cancer; biliary tract cancer; bladder cancer; brain cancer (e.g., glioblastomas, medulloblastomas); cervical cancer; choriocarcinoma; colon cancer; endometrial cancer; esophageal cancer; gastric cancer; hematological neoplasms including acute lymphocytic leukemia and acute myelogenous leukemia; T-cell acute lymphoblastic leukemia/lymphoma; hairy cell
leukemia; chronic lymphocytic leukemia, chronic myelogenous leukemia, multiple myeloma; adult T-cell leukemia/lymphoma; intraepithelial neoplasms including Bowen's disease and Paget's disease; liver cancer; lung cancer; lymphomas including Hodgkin's disease and lymphocytic lymphomas; neuroblastoma; melanoma, oral cancer including squamous cell carcinoma; ovarian cancer including ovarian cancer arising from epithelial cells, stromal cells, germ cells and mesenchymal cells; neuroblastoma, pancreatic cancer; prostate cancer; rectal cancer; sarcomas including angiosarcoma, gastrointestinal stromal tumors, leiomyosarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma, and osteosarcoma; renal cancer including renal cell carcinoma and Wilms tumor; skin cancer including basal cell carcinoma and squamous cell cancer; testicular cancer including germinal tumors such as seminoma, non-seminoma (teratomas, choriocarcinomas), stromal tumors, and germ cell tumors; thyroid cancer including thyroid adenocarcinoma and medullary carcinoma. It will be appreciated that a variety of different tumor types can arise in certain organs, which may differ with regard to, e.g., clinical and/or pathological features and/or molecular markers.

Tumors arising in a variety of different organs are discussed, e.g., the WHO Classification of Tumours series, 4th ed, or 3rd ed (Pathology and Genetics of Tumours series), by the International Agency for Research on Cancer (IARC), WHO Press, Geneva, Switzerland, all volumes of which are incorporated herein by reference. In some embodiments a cancer is one for which mutation or overexpression of particular genes is known or suspected to play a role in development, progression, recurrence, etc., of a cancer. In some embodiments such genes are targets for genetic modification according to methods described herein. In some embodiments a gene is an oncogene, proto-oncogene, or tumor suppressor gene. The term "oncogene" encompasses nucleic acids that, when expressed, can increase the likelihood of or contribute to cancer initiation or progression. Normal cellular sequences ("proto-oncogenes") can be activated to become oncogenes (sometimes termed "activated oncogenes") by mutation and/or aberrant expression.

In various embodiments an oncogene can comprise a complete coding sequence for a gene product or a portion that maintains at least in part the oncogenic potential of the complete sequence or a sequence that encodes a fusion protein. Oncogenic mutations can result, e.g., in altered (e.g., increased) protein activity, loss of proper
regulation, or an alteration (e.g., an increase) in RNA or protein level. Aberrant expression may occur, e.g., due to chromosomal rearrangement resulting in juxtaposition to regulatory elements such as enhancers, epigenetic mechanisms, or due to amplification, and may result in an increased amount of proto-oncogene product or production in an inappropriate cell type. Proto-oncogenes often encode proteins that control or participate in cell proliferation, differentiation, and/or apoptosis. These proteins include, e.g., various transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers, and apoptosis regulators. A TSG may be any gene wherein a loss or reduction in function of an expression product of the gene can increase the likelihood of or contribute to cancer initiation or progression. Loss or reduction in function can occur, e.g., due to mutation or epigenetic mechanisms. Many TSGs encode proteins that normally function to restrain or negatively regulate cell proliferation and/or to promote apoptosis. Exemplary oncogenes include, e.g., MYC, SRC, FOS, JUN, MYB, RAS, RAF, ABL, ALK, AKT, TRK, BCL2, WNT, HER2/NEU, EGFR, MAPK, ERK, MDM2, CDK4, GLI1, GLI2, IGF2, TP53, etc. Exemplary TSGs include, e.g., RB, TP53, APC, NF1, BRCA1, BRCA2, PTEN, CDK inhibitory proteins (e.g., pl6, p21), PTCH, WT1, etc. It will be understood that a number of these oncogene and TSG names encompass multiple family members and that many other TSGs are known. In some embodiments any such gene may be genetically modified, e.g., to generate a cancer model, which may be used, e.g., to determine effect of particular alterations on development of cancer, to determine effect of particular alterations on efficacy of or resistance to treatment, to identify or characterize existing or potential candidate therapeutic agents, etc. Similar methods are envisioned for genes associated with other diseases.

[00160] In some embodiments a disease is a cardiovascular disease, e.g., atherosclerotic heart disease or vessel disease, congestive heart failure, myocardial infarction, cerebrovascular disease, peripheral artery disease, cardiomyopathy.

[00161] In some embodiments a disease is a psychiatric, neurological, or neurodevelopmental disease, e.g., schizophrenia, depression, bipolar disorder, epilepsy, autism, addiction. Neurodegenerative diseases include, e.g., Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, frontotemporal dementia.
[00162] In some embodiments a disease is an autoimmune diseases e.g., acute disseminated encephalomyelitis, alopecia areata, antiphospholipid syndrome, autoimmune hepatitis, autoimmune myocarditis, autoimmune pancreatitis, autoimmune polyendocrine syndromes autoimmune uveitis, inflammatory bowel disease (Crohn's disease, ulcerative colitis), type I diabetes mellitus (e.g., juvenile onset diabetes), multiple sclerosis, scleroderma, ankylosing spondylitis, sarcoid, pemphigus vulgaris, pemphigoid, psoriasis, myasthenia gravis, systemic lupus erythematosus, rheumatoid arthritis, juvenile arthritis, psoriatic arthritis, Behcet's syndrome, Reiter's disease, Berger's disease, dermatomyositis, polymyositis, antineutrophil cytoplasmic antibody-associated vasculitides (e.g., granulomatosis with polyangiitis (also known as Wegener's granulomatosis), microscopic polyangiitis, and Churg-Strauss syndrome), scleroderma, Sjogren's syndrome, anti-glomerular basement membrane disease (including Goodpasture's syndrome), dilated cardiomyopathy, primary biliary cirrhosis, thyroiditis (e.g., Hashimoto's thyroiditis, Graves' disease), transverse myelitis, and Guillane-Barre syndrome.

[00163] In some embodiments a disease is a respiratory disease, e.g., allergy affecting the respiratory system, asthma, chronic obstructive pulmonary disease, pulmonary hypertension, pulmonary fibrosis, and sarcoidosis.

[00164] In some embodiments a disease is a renal disease, e.g., polycystic kidney disease, lupus, nephropathy (nephrosis or nephritis) or glomerulonephritis (of any kind).

[00165] In some embodiments a disease is vision loss or hearing loss, e.g., associated with advanced age.

[00166] In some embodiments a disease is an infectious disease, e.g., any disease caused by a virus, bacteria, fungus, or parasite. In some embodiments it is of interest to modify genes that may be involved in susceptibility to the disease.

[00167] It will be understood that classification of diseases herein is not intended to be limiting. One of ordinary skill in the art will appreciate that various diseases may be appropriately classified in multiple different groups.

[00168] In some embodiments a disease is one for which at least one genome-wide association (GWA) study (GWAS) has been performed. In some embodiments a GWAS types multiple "cases" (subjects having a disease of interest or particular manifestations thereof) and "controls" (subjects not having the disease or
manifestations) for several thousand to millions, e.g., 1 million or more, e.g., 1-5 million or more, alleles (e.g., single nucleotide polymorphisms) positioned throughout the genome or a substantial portion thereof (e.g., at least 80%, 90%, 95%, or more of the genome). It will be understood that control data may be obtained from historical data. Genotyping may be performed using microarrays or other methods. Alleles associated (e.g., in a statistically significant manner) with increased (or decreased) risk of a disease (or particular manifestations) may thereby be identified. It will be appreciated that statistical results may be corrected for multiple hypothesis testing, e.g., using methods known in the art. In some embodiments a p value of less than about $10^{-7}$, $10^{-8}$, or $10^{-9}$ is considered evidence of association. In some embodiments a gene or allele or polymorphism has been identified as contributing to disease risk or severity in at least one GWAS. See, e.g., http://www.genome.gov/gwastudies for examples of GWAS studies and genetic variants (alleles, polymorphisms) associated with various diseases. In some embodiments a gene (or any sequence) is one for which an allele or polymorphism is associated with an increased or decreased risk of developing a disease of at least 1.1, 1.2, 1.5, 2, 3, 4, 5, 7.5, 10, or more, relative to individuals not having the allele or polymorphism. In some embodiments an allele or polymorphism is associated with an increased or decreased risk of developing a disease of at least 1.1, 1.2, 1.5, 2, 3, 4, 5, 7.5, 10, or more, relative to individuals not having the allele or polymorphism. Genes, alleles, polymorphisms, or genetic loci that may contribute to any phenotypic trait of interest such as longevity, weight, resistance to infection, response or lack thereof to various therapeutic agents, resistance or susceptibility to potentially harmful substances such as toxins or infectious agents (e.g., viruses, bacteria, fungi, parasites), are of interest. A phenotypic trait may be a physical sign (such as blood pressure), a biochemical marker, which in some embodiments may be detectable in a body fluid such as blood, saliva, urine, tears, etc., such as level of a metabolite, LDL, etc., wherein an abnormally low or high level of the marker may correlate with having or not having the disease or with susceptibility to or protection from a disease.

[00169] In some embodiments a sequence to be inserted into a genome encodes a tag. The sequence may be inserted into a gene in an appropriate position such that a fusion protein comprising the tag is produced. The term "tag" is used in a broad
sense to encompass any of a wide variety of polypeptides. In some embodiments, a tag comprises a sequence useful for purifying, expressing, solubilizing, and/or detecting a polypeptide. In some embodiments a tag may serve multiple functions. In some embodiments a tag is a relatively small polypeptide, e.g., ranging from a few amino acids up to about 100 amino acids long. In some embodiments a tag is more than 100 amino acids long, e.g., up to about 500 amino acids long, or more.

In some embodiments, a tag comprises an HA, TAP, Myc, 6XHis, Flag, V5, or GST tag, to name few examples. A tag (e.g., any of the afore-mentioned tags) that comprises an epitope against which an antibody, e.g., a monoclonal antibody, is available (e.g., commercially available) or known in the art may be referred to as an "epitope tag". In some embodiments a tag comprises a solubility-enhancing tag (e.g., a SUMO tag, NUS A tag, SNUT tag, a Strep tag, or a monomeric mutant of the Ocr protein of bacteriophage T7). See, e.g., Esposito D and Chatterjee DK. Curr Opin Biotechnol; 17(4):353-8 (2006). In some embodiments, a tag is cleavable, so that at least a portion of it can be removed, e.g., by a protease. In some embodiments, this is achieved by including a protease cleavage site in the tag, e.g., adjacent or linked to a functional portion of the tag. Exemplary proteases include, e.g., thrombin, TEV protease, Factor Xa, PreScission protease, etc. In some embodiments, a "self-cleaving" tag is used. See, e.g., PCT/US05/05763. In some embodiments, a tag comprises a fluorescent polypeptide (e.g., GFP or a derivative thereof such as enhanced GFP (EGFP)) or an enzyme that can act on a substrate to produce a detectable signal, e.g., a fluorescence or colorimetric signal. Luciferase (e.g., a firefly, Renilla, or Gaussia luciferase) is an example of such an enzyme.

Examples of fluorescent proteins include GFP and derivatives thereof, proteins comprising chromophores that emit light of different colors such as red, yellow, and cyan fluorescent proteins, etc. A tag, e.g., a fluorescent protein, may be monomeric. In certain embodiments a fluorescent protein is e.g., Sirius, Azurite, EBFP2, TagBFP, mTurquoise, ECFP, Cerulean, TagCFP, mTFPl, mUkGl, mAGI, AcGFPl, TagGFP2, EGFP, mWasabi, EmGFP, TagYPF, EYFP, Topaz, SYFP2, Venus, Citrine, mKO, mK02, mOrange, mOrange2, TagRFP, TagRFP-T, mStrawberry, mRuby, mCherry, mRaspberry, mKate2, mPlum, mNeptune, mTomato, T- Sapphire, mAmetrine, mKeima. See, e.g., Chalfie, M. and Kain, SR (eds.) Green fluorescent protein: properties, applications, and protocols (Methods of
biochemical analysis, v. 47). Wiley-Interscience, Hoboken, N.J., 2006, and/or Chudakov, DM, et al, Physiol Rev. 90(3): 1103-63, 2010 for discussion of GFP and numerous other fluorescent or luminescent proteins. In some embodiments a tag may comprise a domain that binds to and/or acts a sensor of a small molecule (e.g., a metabolite) or ion, e.g., calcium, chloride, or of intracellular voltage, pH, or other conditions. Any genetically encodable sensor may be used; a number of such sensors are known in the art. In some embodiments a FRET-based sensor may be used. In some embodiments different genes are modified to incorporate different tags, so that proteins encoded by the genes are distinguishably labeled. For example, between 2 and 20 distinct tags may be introduced. In some embodiments the tags have distinct emission and/or absorption spectra. In some embodiments a tag may absorb and/or emit light in the infrared or near-infrared region. It will be understood that any nucleic acid sequence encoding a tag may be codon-optimized for expression in a cell, zygote, embryo, or animal into which it is to be introduced.

[00170] In some embodiments it may be of interest to express fragments or domains of a protein, which may act in a dominant negative manner and may, for example, disrupt normal function or interaction of the protein.

[00171] In some embodiments a gene of interest encodes a protein the aggregation of which is associated with one or more diseases, which may be referred to as protein misfolding diseases. Examples include, e.g., alpha-synuclein (Parkinson’s disease and related disorders), amyloid beta or tau (Alzheimer’s disease), TDP-43 (frontotemporal dementia, ALS).

[00172] In some embodiments a gene of interest encodes a transcription factor, a transcriptional co-activator or co-repressor, an enzyme, a chaperone, a heat shock factor, a heat shock protein, a receptor, a secreted protein, a transmembrane protein, a histone (e.g., H1, H2A, H2B, H3, H4), a peripheral membrane protein, a soluble protein, a nuclear protein, a mitochondrial protein, a growth factor, a cytokine (e.g., an interleukin, e.g., any of IL-1 - IL-33), an interferon (e.g., alpha, beta, or gamma), a chemokine (e.g., a CXC, CX3C, C (or XC), or CX3C chemokine). A chemokine may be CCL1 - CCL28, CXCL1 - CXCL17, XCL1 or XCL2, or CXCL11. In some embodiments a gene encodes a colony-stimulating factor, a hormone (e.g., insulin, thyroid hormone, growth hormone, estrogen, progesterone, testosterone), an extracellular matrix protein (e.g., collagen, fibronectin), a motor protein (e.g.,
dynein, myosin), cell adhesion molecule, a major or minor histocompatibility
(MHC) gene, a transporter, a channel (e.g., an ion channel), an immunoglobulin (Ig)
supersnall (IgSF) gene (e.g., a gene encoding an antibody, T cell receptor, B cell
receptor), tumor necrosis factor, an NF-kappaB protein, an integrin, a cadherin
superfamily member (e.g., a cadherin), a selectin, a clotting factor, a complement
factor, a plasminogen, plasminogen activating factor. Growth factors include, e.g.,
members of the vascular endothelial growth factor (VEGF, e.g., VEGF-A, VEGF-B,
VEGF-C, VEGF-D), epidermal growth factor (EGF), insulin-like growth factor
(IGF; IGF-1, IGF-2), fibroblast growth factor (FGF, e.g., FGF1 - FGF22), platelet
derived growth factor (PDGF), or nerve growth factor (NGF) families. It will be
understood that the afore-mentioned protein families comprise multiple members.
Any such member may be used in various embodiments. In some embodiments a
growth factor promotes proliferation and/or differentiation of one or more
hematopoietic cell types. For example, a growth factor may be CSF1 (macrophage
colony-stimulating factor), CSF2 (granulocyte macrophage colony-stimulating
factor, GM-CSF), or CSF3 (granulocyte colony-stimulating factors, G-CSF). In
some embodiments a gene encodes erythropoietin (EPO). In some embodiments, a
gene encodes a neurotrophic factor, i.e., a factor that promotes survival,
development and/or function of neural lineage cells (which term as used herein
includes neural progenitor cells, neurons, and glial cells, e.g., astrocytes,
oligodendrocytes, microglia). For example, in some embodiments, the protein is a
factor that promotes neurite outgrowth. In some embodiments, the protein is ciliary
neurotrophic factor (CNTF) or brain-derived neurotrophic factor (BDNF).

[00173] In some embodiments a gene of interest encodes a polypeptide that is a
subunit of any protein that is comprised of multiple subunits.

[00174] An enzyme may be any protein that catalyzes a reaction of a type that has
been assigned an Enzyme Commission number (EC number) by the Nomenclature
Committee of the International Union of Biochemistry and Molecular Biology (NC-
IUBMB). Enzymes include, e.g., oxidoreductases, transferases, hydrolases, lyases,
isomerases, ligases. Examples include, e.g., kinases (protein kinases, e.g., Ser/Thr
kinase, Tyr kinase), lipid kinases (e.g., phosphatidylmositide 3-kinases (PI 3-kinases
or PI3Ks)), phosphatases, acetyltransferases, methyltransferases, deacetylases,
demethylases, lipases, cytochrome P450s, glucuronidases, recombinases (e.g., Rag-
1. Rag-2). An enzyme may participate in the biosynthesis, modification, or degradation of nucleotides, nucleic acids, amino acids, proteins, neurotransmitters, xenobiotics (e.g., drugs) or other macromolecules.

[00175] The mammalian genome encodes at least about 500 different kinases. Kinases can be classified based on the nature of their typical substrates and include protein kinases (i.e., kinases that transfer phosphate to one or more protein(s)), lipid kinases (i.e., kinases that transfer a phosphate group to one or more lipid(s)), nucleotide kinases, etc. Protein kinases (PKs) are of particular interest in certain aspects of the invention. PKs are often referred to as serine/threonine kinases (S/TKs) or tyrosine kinases (TKs) based on their substrate preference.

Serine/threonine kinases (EC 2.7.1 1.1) phosphorylate serine and/or threonine residues while TKs (EC 2.7.10.1 and EC 2.7.10.2) phosphorylate tyrosine residues. A number of "dual specificity" kinases (EC 2.7.12.1) that are capable of phosphorylating both serine/threonine and tyrosine residues are known. The human protein kinase family can be further divided based on sequence/structural similarity into the following groups: (1) AGC kinases - containing PKA, PKC and PKG; (2) CaM kinases - containing the calcium/calmodulin-dependent protein kinases; (3) CK1 - containing the casein kinase 1 group; (4) CMGC - containing CDK, MAPK, GSK3 and CLK kinases; (5) STE - containing the homologs of yeast Sterile 7, Sterile 11, and Sterile 20 kinases; (6) TK - containing the tyrosine kinases; (7) TKL - containing the tyrosine-kinase like group of kinases. A further group referred to as "atypical protein kinases" contains proteins that lack sequence homology to the other groups but are known or predicted to have kinase activity, and in some instances are predicted to have a similar structural fold to typical kinases.

[00176] Receptors include, e.g., G protein coupled receptors, tyrosine kinase receptors, serine/threonine kinase receptors, Toll-like receptors, nuclear receptor, immune cell surface receptor. In some embodiments a receptor is a receptor for any of the hormones, cytokines, growth factors, or secreted proteins mentioned herein. Numerous G protein coupled receptors (GPCRs) are known in the art. See, e.g., Vroeling B, GPCRDB: information system for G protein-coupled receptors. Nucleic Acids Res. 2011 Jan;39(Database issue):D309-19. Epub 2010 Nov 2. The GPCRDB can be found online at http://www.gpcr.org/7tm/.

G protein coupled receptors include, e.g., adrenergic, cannabinoid, purinergic receptors, neuropeptide receptors,
olfactory receptors. Transcription factors (TFs) (sometimes called sequence-specific DNA-binding factors) bind to specific DNA sequences and (alone or in a complex with other proteins), regulate transcription, e.g., activating or repressing transcription. Exemplary TFs are listed, for example, in the TRANSFAC® database, Gene Ontology (http://www.geneontology.org/) or DBD (www.transcriptionfactor.org) (Wilson, et al, DBD - taxonomically broad transcription factor predictions: new content and functionality Nucleic Acids Research 2008 doi: 10.1093/nar/gkm964). TFs can be classified based on the structure of their DNA binding domains (DBD). For example in certain embodiments a TF is a helix-loop-helix, helix-turn-helix, winged helix, leucine zipper, bZIP, zinc finger, homeodomain, or beta-scaffold factor with minor groove contacts protein. Transcription factors include, e.g., p53, STAT3, PAS family transcription factors (e.g., HIF family: HIF1A, HIF2A, HIF3A), aryl hydrocarbon receptor.

[00177] In some embodiments it may be of interest to genetically modify multiple genes that function in the same biological pathway or process, e.g., signal transduction pathway, biosynthetic pathway, xenobiotic metabolizing pathway, anabolic or catabolic pathway, apoptosis, autophagy, endocytosis, exocytosis. In some embodiments an animal generated according to inventive methods is useful for studying drug metabolism. For example, it may be of interest to genetically modify multiple enzymes involved in xenobiotic metabolism (e.g., multiple P450s). In some embodiments an animal generated according to inventive methods is useful for studying the immune system and/or for generating animals that have a humanized immune system or that are immunocompromised and may serve as hosts for cells or tissues from other organisms of the same species or different species.

[00178] The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. Various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims. The advantages and objects of the invention are not necessarily encompassed by each embodiment of the invention. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments described herein, which fall within the
scope of the claims. The scope of the present invention is not to be limited by or to embodiments or examples described above.

Section headings used herein are not to be construed as limiting in any way. It is expressly contemplated that subject matter presented under any section heading may be applicable to any aspect or embodiment described herein. Embodiments or aspects herein may be directed to any agent, composition, article, kit, and/or method described herein. It is contemplated that any one or more embodiments or aspects can be freely combined with any one or more other embodiments or aspects whenever appropriate. For example, any combination of two or more agents, compositions, articles, kits, and/or methods that are not mutually inconsistent, is provided.

Articles such as "a", "an", "the" and the like, may mean one or more than one unless indicated to the contrary or otherwise evident from the context. The phrase "and/or" as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause. As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when used in a list of elements, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but optionally more than one, of list of elements, and, optionally, additional unlisted elements. Only terms clearly indicative to the contrary, such as "only one of" or "exactly one of" will refer to the inclusion of exactly one element of a number or list of elements. Thus claims that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present, employed in, or otherwise relevant to a given product or process unless indicated to the contrary. Embodiments are provided in which exactly one member of the group is present, employed in, or otherwise relevant to a given product or process. Embodiments are provided in which more than one, or all of the group members are present, employed in, or otherwise relevant to a given product or process. Any one or more claims may be amended to explicitly exclude any embodiment, aspect, feature, element, or characteristic, or any combination
thereof. Any one or more claims may be amended to exclude any agent, composition, amount, dose, administration route, cell type, target, cellular marker, antigen, targeting moiety, or combination thereof.

[00183] Embodiments in which any one or more limitations, elements, clauses, descriptive terms, etc., of any claim (or relevant description from elsewhere in the specification) is introduced into another claim are provided. For example, a claim that is dependent on another claim may be modified to include one or more elements or limitations found in any other claim that is dependent on the same base claim. It is expressly contemplated that any amendment to a genus or generic claim may be applied to any species of the genus or any species claim that incorporates or depends on the generic claim.

[00184] Where a claim recites a composition, methods of using the composition as disclosed herein are provided, and methods of making the composition according to any of the methods of making disclosed herein are provided. Where a claim recites a method, a composition for performing the method is provided. Where elements are presented as lists or groups, each subgroup is also disclosed. It should also be understood that, in general, where embodiments or aspects is/are referred to herein as comprising particular element(s), feature(s), agent(s), substance(s), step(s), etc., (or combinations thereof), certain embodiments or aspects may consist of, or consist essentially of, such element(s), feature(s), agent(s), substance(s), step(s), etc. (or combinations thereof). It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited. Any method of treatment may comprise a step of providing a subject in need of such treatment, e.g., a subject having a disease for which such treatment is warranted. Any method of treatment may comprise a step of diagnosing a subject as being in need of such treatment, e.g., diagnosing a subject as having a disease for which such treatment is warranted.

[00185] Where ranges are given herein, embodiments in which the endpoints are included, embodiments in which both endpoints are excluded, and embodiments in which one endpoint is included and the other is excluded, are provided. It should be assumed that both endpoints are included unless indicated otherwise. Unless otherwise indicated or otherwise evident from the context and understanding of one
of ordinary skill in the art, values that are expressed as ranges can assume any specific value or subrange within the stated ranges in various embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. "About" in reference to a numerical value generally refers to a range of values that fall within ±10%, in some embodiments ±5%, in some embodiments ±1%, in some embodiments ±0.5%> of the value unless otherwise stated or otherwise evident from the context. In any embodiment in which a numerical value is prefaced by "about", an embodiment in which the exact value is recited is provided. Where an embodiment in which a numerical value is not prefaced by "about" is provided, an embodiment in which the value is prefaced by "about" is also provided. Where a range is preceded by "about", embodiments are provided in which "about" applies to the lower limit and to the upper limit of the range or to either the lower or the upper limit, unless the context clearly dictates otherwise. Where a phrase such as "at least", "up to", "no more than", or similar phrases, precedes a series of numbers, it is to be understood that the phrase applies to each number in the list in various embodiments (it being understood that, depending on the context, 100% of a value, e.g., a value expressed as a percentage, may be an upper limit), unless the context clearly dictates otherwise. For example, "at least 1, 2, or 3" should be understood to mean "at least 1, at least 2, or at least 3" in various embodiments. It will also be understood that any and all reasonable lower limits and upper limits are expressly contemplated.

[00186] Exemplification
[00187] Example 1
[00188] EXPERIMENTAL PROCEDURES
[00189] Procedures for generating sgRNAs expressing vector
[00190] Bicistronic expression vector expressing Cas9 and sgRNA (Cong et al., Science 339:819-823 (2013)) were digested with BbsI and treated with Antarctic Phosphatase, and the linearized vector was gel-purified. A pair of oligos (Table 6) for each targeting site was annealed, phosphorylated, and ligated to linearized vector.
[00191] Cell culture and Transfection
[00192] V6.5 mESCs (on a 129/Sv x C57BL/6 F1 hybrid background) were cultured on gelatin-coated plates with standard mESC culture conditions. Cells were
transfected with a plasmid expressing mammalian codon optimized Cas9 and sgRNA (single targeting), or three plasmids expressing Cas9 and sgRNAs targeting Tet1, Tet2, and Tet3 (triple targeting), or five PCR products each coding for sgRNA targeting Tet1, Tet2, Tet3, Sry, and Uty, along with a plasmid expressing PGK-puroR using FuGENE HD reagent (Promega), following manufacturer's instructions. 12 hours after transfection, mESC were re-plated at a low density on DR4 MEF feeder layers. Puromycin (2µg/ml) was added one day after replating and taken off after 48 hours. After recovering for 4 to 6 days, individual colonies were picked and genotyped by RFLP and Southern blot analysis, and the leftover mES cells on plate were collected for Suveryor assay.

[00193] Suveryor assay and RFLP analysis for genome modification

[00194] Suveryor assay was performed as described by (Guschin et al., Methods Mol Biol, 649:247-256 (2010)). Genomic DNA from treated and control ES cells or targeted and control mice was extracted. Mouse genomic DNA samples were prepared from tail biopsies. PCR was performed using Tet1, 2, 3 specific primers (Table S3) under the following conditions: 95°C for 5 min; 35x (95°C for 30 s, 60°C for 30 s, 68°C for 40 s); 68°C for 2 min; hold at 4°C. PCR products were then denatured, annealed, and treated with Suveryor nuclease (Transgenomic). DNA concentration of each band was measured on an ethidium bromide-stained 10% acrylamide Criterion TBE gel (BioRad) and quantified using Image J software. The same PCR products for Suveryor assay were used for RFLP analysis. 1Oul of Tet1, Tet2, or Tet3 PCR product was digested with SacI, EcoRV, or Xhol respectively. Digested DNA was separated on an ethidium bromide-stained agarose gel (2%). For sequencing, PCR products were cloned using the Original TA Cloning Kit (Invitrogen), and mutations were identified by Sanger sequencing.

[00195] Dot blot

[00196] DNA was extracted from pre-plated mESCs following standard procedures. DNA was transferred to nylon membrane using BioRad slot blot vacuum manifold apparatus. Anti-5hmC (Active Motif 1:10000) was used to detect 5hmC following manufacturer's protocol.
[00197] Production of Cas9 mRNA and sgRNA

[00198] T7 promoter was added to Cas9 coding region by PCR amplification using primer Cas9 F and R (Table 6). T7-Cas9 PCR product was gel-purified and used as the template for in vitro transcription (IVT) using mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). T7 promoter was added to sgRNAs template by PCR amplification using primer Tetl F and R, Tet2 F and R, Tet3 F and R (Table 6). The T7-sgRNA PCR product was gel-purified and used as the template for IVT using MEGAscript T7 kit (Life Technologies). Both the Cas9 mRNA and the sgRNAs were purified using MEGAClear kit (Life Technologies) and eluted in RNase-free water.

[00199] One cell embryo injection

[00200] All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at MIT. B6D2F1 (C57BL/6 X DBA2) female mice and ICR mouse strains were used as embryo donors and foster mothers, respectively. Super-ovulated female B6D2F1 mice (7-8 weeks old) were mated to B6D2F1 stud males, and fertilized embryos were collected from oviducts. Cas9 mRNAs (from 20 ng/µl to 200 ng/µl) and sgRNA (from 20 ng/µl to 50 ng/µl) was injected into the cytoplasm of fertilized eggs with well recognized pronuclei in M2 medium (Sigma). For oligos injection, Cas mRNA (100 ng/µl), sgRNA (50µl / µl) and donor oligos (100 ng/µl) were mixed and injected into zygotes at the pronuclei stage. The injected zygotes were cultured in KSOM with amino acids at 37°C under 5% C02 in air until blastocyst stage by 3.5 days. Thereafter, 15-25 blastocysts were transferred into uterus of pseudopregnant ICR females at 2.5 dpc.

[00201] Southern blotting

[00202] Genomic DNA was separated on a 0.8% agarose gel after restriction digests with the appropriate enzymes, transferred to a nylon membrane (Amersham) and hybridized with 32P random primer (Stratagene)-labeled probes.

[00203] Prediction of potential off-targets

[00204] Potential targets of CRISPR sgRNAs were found using the rules outline in Mali et al., Science, 339:823-826 (2013). For a 20nt sgRNA sequence of nnnnn nnMMMM MMMMMM MMMMMM (SEQ ID NO: 135), where M are the seed bases
preceding the PAM sequence NGG, four search sequences (MMM MMMM MMMM AGG (SEQ ID NO: 136); MMM MMMM MMMM CGG (SEQ ID NO: 137); MMM MMMM MMMM GGG (SEQ ID NO: 138); MMM MMMM MMMM TGG (SEQ ID NO: 139)) were generated. Exact matches to these search sequences in the mouse genome (mm9) were found using bowtie and reported as potential targets of the CRISPR sgRNA.

RESULTS

Simultaneous targeting up to five genes in ES cells

To test the possibility of targeting functionally redundant genes from the same gene family, sgRNAs targeting the Ten-eleven translocation (Tet) family members, Tet1, Tet2 and Tet3 were digested (Fig 1A). Tet proteins (Tet1/2/3) convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in various embryonic and adult tissues and mutant mice for each of these three genes have been produced by homologous recombination in ES cells (Dawlaty et al. Cell Stem Cell, 9:166-175 (2011); Gu et al, Nature, 477:606-610 (2011); Li et al, Blood, 118:4509-4518 (2011); Moran-Crusio et al, Cancer Cell, 20:11-24 (2011)). To test whether the CRISPR/Cas system could produce targeted cleavage in the mouse genome, plasmids expressing both the mammalian codon optimized Cas9 and a sgRNA targeting each gene (Cong et al, Science, 339:819-823(2013); Mali et al, Science, 339:823-826 (2013)) were transfected into mouse ES cells and determined the targeted cleavage efficiency by the Surveyor assay (Guschin et al, Methods Mol Biol, 649:247-256 (2010)). All three Cas9-sgRNA transfections produced cleavage at target loci with high efficiency of 36% at Tet1, 48% at Tet2, and 36% at Tet3 (Fig 1B). Because each target locus contains a restriction enzyme recognition site (Fig 1A), a ~500bp fragment around each target site was PCR amplified, and the PCR products were digested with the respective enzyme. A correctly targeted allele will lose the restriction site, which can be detected by failure to cleave upon enzyme treatment. Using this restriction fragment length polymorphism (RFLP) assay, 48 ES cell clones from each single targeting experiment were screened. Consistent with the Surveyor analysis, a high percentage of mESC clones were targeted, with a high probability of having both alleles mutated (Fig 5A). The results summarized in Table 1 demonstrate that between about 65% and about 81% of the tested ES cell
clones carried mutations in the Tet genes with up to about 77% having mutations in both alleles.

[00208] The high efficiency of single gene modification prompted testing targeting of all three genes simultaneously. For this ES cells were co-transfected with the constructs expressing Cas9 and three sgRNAs targeting Tet1, 2 and 3. Of 96 clones screened using the RFLP assay, 20 clones were identified as having mutations in all six alleles of the three genes (Fig 1C, 5B, Table 1). To exclude that a PCR bias could give false positive results, Southern blot analysis was performed and complete agreement with the RFLP results was confirmed (Fig 1C). The PCR products of Tet1, Tet2, and Tet3 targeted regions were sub-cloned and sequenced to verify that all of eight tested clones carried biallelic mutations in all three genes with most clones displaying two mutant alleles for each gene with small insertions or deletions (indels) at the target site (Fig ID). To test whether these mutant alleles would abolish the function of Tet proteins, the 5hmC level of targeted clones were compared to wt mES cells. Previously, a depletion of 5hmC in Tet1/Tet2 double knockout mES cells derived using traditional gene targeting methods was reported (Dawlaty et al, Cell Stem Cell, 9:166-175 (2011)). As expected from loss of function alleles, a significant reduction of 5hmC levels in all clones carrying biallelic mutations in the three genes was found (Fig IE).

[00209] Recently efficient targeting of two Y-linked genes, Sry and Uty, using TALENs was demonstrated (Wang et al., in press). To further test the potential of multiplexed gene targeting by CRISPR/Cas system, sgRNAs targeting these two Y-linked genes were designed (Figure 5C). Short PCR products encoding sgRNAs targeting all five genes (Tet1, Tet2, Tet3, Sry, and Uty) were pooled and co-transfected with a Cas9 expressing plasmid and the PGK puroR cassette into ES cells. Of 96 clones that were screened using the RFLP assay, 10% carried mutations in all eight alleles of the five genes (Figure 5D, Table 4), demonstrating the capacity of the CRISPR/Cas9 system for highly efficient multiplexed gene targeting.

[00210] One step generation of single gene mutant mice by zygote injection
[00211] Whether mutant mice could be generated in vivo by direct embryo manipulation was tested. Capped polyadenylated Cas9 mRNA was produced by in vitro transcription and co-injected with sgRNAs. Initially, to determine the optimal concentration of Cas9 mRNA for targeting in vivo, varying amounts of Cas9-
encoding mRNA were injected with Tet1 targeting sgRNA at constant concentration (20 ng/µl) into pronuclear (PN) stage one-cell mouse embryos and the frequency of altered alleles at the blastocyst stage was assessed using the RFLP assay. As expected, higher concentration of Cas9 mRNA led to more efficient gene disruption (Fig. 6A). Nevertheless, even embryos injected with the highest amount of Cas9 mRNA (200%/µl) showed normal blastocyst development, indicating low toxicity.

To investigate whether postnatal mice carrying targeted mutations could be generated, sgRNAs targeting Tet1 or Tet2 were co-injected with different concentrations of Cas9 mRNA. Blastocysts derived from the injected embryos were transplanted into foster mothers and newborn pups were obtained. As summarized in Table 2, about 10% of the transferred blastocysts developed to birth independent of the RNA concentrations used for injection indicating low fetal toxicity of the Cas9 mRNA and sgRNA. RFLP, Southern blot, and sequencing analysis demonstrated that between 50 and 90% of the postnatal mice carried biallelic mutations in either target gene (Figs. 2A, 2B, 2C, Table 2).

Surprisingly, specific Δ9 Tet1 and specific Δ8 and Δ15 Tet2 mutant alleles were repeatedly recovered in independently derived mice. Preferential generation of these alleles is likely caused by a short sequence repeat flanking the DSB (see Figure 6B) consistent with a previous report demonstrating that perfect microhomology sequences flanking the cleavage sites can generate microhomology-mediated precise deletions by end repair mechanism (MMEJ) (McVey and Lee, Trends Genet, 24:529-538(2008); Symington and Gautier, Annu Rev Genet, 11:636-646 (2011)) (Fig 6B). A similar observation was also made when TALEN mRNA was injected into one cell rat embryos (Tesson et al, Nat Biotechnol, 29:695-696 (2011)).

Blastocysts were also derived from zygotes injected with Cas9 mRNA and Tet3 sgRNA. Genotyping of the blastocysts demonstrated that of eight embryos three were homozygous and three were heterozygous Tet3 mutants (two failed to amplify) (Fig 6C). Some blastocysts were implanted into foster mothers and, upon C-section, multiple mice of smaller size (Fig 6D), many of which died soon after delivery, were readily identified. Genotyping shown in Fig 6E indicated that all pups with mutations in both Tet3 alleles died neonatally. Only two out of 15 mice survived that were either Tet3 heterozygous mutants or wt (Fig 6F). These results
are consistent with the lethal neonatal phenotype of Tet3 knockout mice generated using traditional methods (Gu et al, Nature, 477:606-610 (2011)), although which of the Tet3 mutations produced loss of function rather than hypomorphic alleles has not been established.

[00215] One step generation of double gene mutant mice by zygote injection

[00216] To test whether Tet1/Tet2 double mutant mice could be produced from single embryos, Tet1 and Tet2 sgRNAs were co-injected with 20 or 100 ng/µl Cas9 mRNA into zygotes. A total of 28 pups were born from 144 embryos transferred into foster mothers (21% live birth rate) that had been injected at the zygote stage with high concentrations of RNA (Cas9 mRNA at 100 ng/µl, sgRNAs at 50 ng/µl), consistent with low or no toxicity of the Cas9 mRNA and sgRNAs (Table 3). RFLP, Southern blot analysis and sequencing identified 22 mice carrying targeted mutations at all four alleles of the Tet1 and Tet2 genes (Fig 2D, 2E) with the remaining mice carrying mutations in a subset of alleles (Table 3). Injection of zygotes with low concentration of RNA (Cas9 mRNA at 20 ng/µl, sgRNAs at 20 ng/µl) yielded 19 pups from 75 transferred embryos (about 25% live birth rate), which is a higher survival rate than from embryos injected with 100ng/µl of Cas9 RNA. Nevertheless, more than about 50% of the pups were biallelic Tet1/Tet2 double mutants (Table 3). These results demonstrate that postnatal mice carrying biallelic mutations in two different genes can be generated within one month with high efficiency (Fig 2F).

[00217] Although the high live birth rate and normal development of mutant mice indicate low toxicity of CRISPR/Cas9 system, the off-target effects in vivo were determined. Previous work in vitro, in bacteria, and in cultured human cells suggested that the protospacer-adjacent motif (PAM) sequence NGG and the 8-12 base "seed sequence" at the 3' end of the sgRNA are most important for determining the DNA cleavage specificity (Cong et al, Science, 339:819-823(2013); Jiang et al, Nat Biotechnol, 31:233-239 (2013); Jinek et al, Science, 337:816-821 (2012)). Based on this rule, only three and four potential off targets exist in mouse genome for Tet1 and Tet2 sgRNA respectively (Table 5, Experimental procedures), with each of them perfectly matching the 12 bp seed sequence at the 3' end and the NGG PAM sequence of the sgRNA (there is no potential off target site for Tet3 sgRNA)
using this prediction rule). From seven double mutant mice produced from injection with high RNA concentration ~400bp fragments from all seven potential off-target loci were PCR amplified and no cleavage was found in the Surveyor assay (Figs. 7A-7B), indicating a high specificity of CRISPR/Cas system.

Multiplexed precise HDR-mediated genome editing in vivo

The NHEJ-mediated gene mutations described above produced mutant alleles with different and unpredictable insertions and deletions of variable size. The possibility of precise homology directed repair (HDR)-mediated genome editing by co-injecting Cas9 mRNA, sgRNAs and single stranded DNA oligos into one-cell embryos was explored. For this an oligo targeting Tet1 so as to change two base pairs of a Sacl restriction site and create instead an EcoRI site and a second oligo targeting Tet2 with two base pair changes that would convert an EcoRV site into an EcoRI site were designed (Fig 3A). Blastocysts were derived from zygotes injected with Cas9 mRNA and sgRNAs and oligos targeting Tet1 or Tet2, respectively. DNA was isolated, amplified and digested with EcoRI to detect oligo mediated HDR events. Six out of nine Tet1 targeted embryos and nine out of 15 Tet2 targeted embryos incorporated an EcoRI site at the respective target locus, with several embryos having both alleles modified (Fig 8A). When Cas9 mRNA, sgRNAs, and single stranded DNA oligos targeting both Tet1 and Tet2 were co-injected into zygotes, out of 14 embryos, four were identified that were targeted with the oligo at the Tet1 locus, seven that were targeted with the oligo at the Tet2 locus and one embryo (#2) that had one allele of each gene correctly modified (Fig 8B). All four alleles of embryo #2 were sequenced, confirming that one allele of each gene contained the 2bp changes directed by the oligo, while the other alleles were disrupted by NHEJ-mediated deletion and insertion (Fig 8C).

Blastocysts with double oligo injections were implanted into foster mothers and a total of 10 pups were born from 48 embryos transferred (21% live birth rate). Upon RFLP analysis using EcoRI, seven mice containing EcoRI sites at the Tet1 locus and eight mice containing EcoRI sites at the Tet2 locus, with six mice containing EcoRI sites at both Tet1 and Tet2 loci were identified (Fig 3B). RFLP analysis using Sacl and EcoRV to Tet1 and Tet2 loci respectively was also applied showing that all alleles not targeted by oligos contained disruptions, which is in consistent with the high biallelic mutation rate by Cas9 mRNA and sgRNAs
injection. These results were confirmed by sequencing demonstrating mutations in all four alleles of mouse #5 and #7 (Fig 3C). The results herein demonstrate that mice with HR-mediated precise mutations in multiple genes can be generated in one step by CRISPR/Cas mediated genome editing.

Table 1. CRISPR/Cas mediated gene targeting in V6.5 mES cells

Plasmids encoding Cas9 and sgRNAs targeting Tet1, Tet2, and Tet3 were transfected separately (single targeting) or in a pool (triple targeting) into mES cells. The number of total alleles mutated in each mES cell clone is listed from 0 to 2 for single targeting experiment, and 0 to 6 for triple targeting experiment. The number of clones containing each specific number of mutated alleles is shown in relation to the total number of clones screened in each experiment.

<table>
<thead>
<tr>
<th>Sg RNA</th>
<th>Mutant alleles per clone/Total clones tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Tet1</td>
<td>NA</td>
</tr>
<tr>
<td>Tet2</td>
<td>NA</td>
</tr>
<tr>
<td>Tet3</td>
<td>NA</td>
</tr>
<tr>
<td>Tet1+</td>
<td>20/96</td>
</tr>
<tr>
<td>Tet2+</td>
<td>20/96</td>
</tr>
<tr>
<td>Tet3+</td>
<td>20/96</td>
</tr>
</tbody>
</table>

Table 2. CRISPR/Cas mediated single gene targeting in BDF2 mice.

Cas9 mRNA and sgRNAs targeting Tet1, Tet2, or Tet3 were injected into fertilized eggs. The blastocysts derived from injected embryos were transplanted into foster mothers and newborn pups were obtained and genotyped. The number of total alleles mutated in each mouse is listed from 0 to 2. The number of mice containing each specific number of mutated alleles is shown in relation to the total number of mice screened in each experiment.

<table>
<thead>
<tr>
<th>Sg RNA</th>
<th>Dose of Cas9/gRNA mRNA (ng/µl)</th>
<th>Blastocyst/Injected zygotes</th>
<th>Transferred embryos (recipients)</th>
<th>Newborns (dead)</th>
<th>Mutant alleles per mouse/Total mice tested*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Tet1</td>
<td>200/20</td>
<td>38/50</td>
<td>19(1)</td>
<td>2(0)</td>
<td>2/2</td>
</tr>
<tr>
<td>Tet1</td>
<td>100/20</td>
<td>50/60</td>
<td>25(1)</td>
<td>3(0)</td>
<td>2/3</td>
</tr>
<tr>
<td>Tet1</td>
<td>50/20</td>
<td>40/50</td>
<td>40(2)</td>
<td>8(3)</td>
<td>4/7</td>
</tr>
</tbody>
</table>
*Some of the pups were cannibalized

**Table 3.** CRISPR/Cas mediated double gene targeting in BDF2 mice.

Cas9 mRNA and sgRNAs targeting Tet1 and Tet2 were co-injected into fertilized eggs. The blastocysts derived from the injected embryos were transplanted into foster mothers and newborn pups were obtained and genotyped. The number of total alleles mutated in each mouse is listed from 0 to 4 for Tet1 and Tet2. The number of mice containing each specific number of mutated alleles is shown in relation to the number of total mice screened in each experiment.

<table>
<thead>
<tr>
<th>SgRNA</th>
<th>Dose of Cas9/gRNA mRNA (ng/µl)</th>
<th>Blastocyst/Injected zygotes</th>
<th>Transferred embryos</th>
<th>Newborns (dead)</th>
<th>Mutant alleles per mouse/Total mice tested*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1 + Tet2</td>
<td>100/50</td>
<td>194/229</td>
<td>144(7)</td>
<td>31(8)</td>
<td>4/3, 2, 1, 1, 0</td>
</tr>
<tr>
<td>Tet1 + Tet2</td>
<td>20/20</td>
<td>92/109</td>
<td>75(5)</td>
<td>19(3)</td>
<td>11/19, 1/19, 2/19, 3/19, 2/19</td>
</tr>
</tbody>
</table>

*Some of the pups were cannibalized

**Table 4.** Plasmids encoding Cas9 and five PCR products expressing sgRNAs targeting Tet1, Tet2, Tet3, Sry, and Uty were co-transfected into mES cells. The number of clones containing mutations in all six Tet alleles is listed in the Tet1, 2, 3 column; the number of clones containing mutations in all six Tet alleles and Sry allele is listed in the Tet1, 2, 3 + Sry column; the number of clones containing mutations in all six Tet alleles and both Sry and Uty allele is listed in the Tet1, 2, 3 + Sry + Uty column.

**Table 4.** The increased efficiency of generating Tet1, 2, 3 triple targeted mES clones in this quintuple targeting experiment, compared to the triple targeting experiment (Table 1), is likely due to the use of short PCR products instead of...
plasmids that express sgRNAs. The much smaller size of pooled PCR products may ensure more efficient delivery into transfected cells. Table 4 is related to Table 1.

<table>
<thead>
<tr>
<th>Mutant Genes</th>
<th>Tet 1, 2, 3</th>
<th>Tet 1, 2, 3 + Sry</th>
<th>Tet 1, 2, 3 + Sry + Uty</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Mutant alleles</td>
<td>6 and More</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>No. Mutant clones/Total clones</td>
<td>54/96</td>
<td>37/96</td>
<td>10/96</td>
</tr>
</tbody>
</table>

Table 5 Potential off targets of Tet1 and Tete2 sgRNAs

<table>
<thead>
<tr>
<th>MatchName</th>
<th>Coordinate (mm9)</th>
<th>Strand</th>
<th>SEEDPAM</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>chr10:6229629-3-62296308</td>
<td>-</td>
<td>ggcgtgcgtCAAGGAGCCTACGAG (SEQ ID NO: 140)</td>
<td>Tet1</td>
</tr>
<tr>
<td>Tet1</td>
<td>chr16:8891797-8891794</td>
<td>+</td>
<td>cgcgttgGTCAAGGAGCTCAAGG (SEQ ID NO: 141)</td>
<td>33kb 3’ of 1810013L24Rik</td>
</tr>
<tr>
<td>Tet1</td>
<td>chr18:7513031-8-75130333</td>
<td>-</td>
<td>GgcggacGTCAAGGAGCTCAAGG (SEQ ID NO: 142)</td>
<td>9.4kb 5’ of Lipg</td>
</tr>
<tr>
<td>Tet1</td>
<td>chr2:36287584-36287599</td>
<td>+</td>
<td>gtttagtGTCAAGGAGCTCAAGG (SEQ ID NO: 143)</td>
<td>9.8kb 3’ of Olfr339</td>
</tr>
</tbody>
</table>

Tet2

<table>
<thead>
<tr>
<th>MatchName</th>
<th>Coordinate (mm9)</th>
<th>Strand</th>
<th>SEEDPAM</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet2</td>
<td>chr3:13314861-7-133148632</td>
<td>-</td>
<td>gaaaggtgCACAACAGATATCCAGG (SEQ ID NO: 144)</td>
<td>Tet2</td>
</tr>
<tr>
<td>Tet2</td>
<td>chr2:12069659-9-120696614</td>
<td>+</td>
<td>gcaagaacCACAACAGATATCCAGG (SEQ ID NO: 145)</td>
<td>Intron of Ubr1</td>
</tr>
<tr>
<td>Tet2</td>
<td>chr10:9520632-6-95206341</td>
<td>+</td>
<td>aagaaacCACAACAGATATCCAGG (SEQ ID NO: 146)</td>
<td>2.4kb 5’ of AK169506</td>
</tr>
<tr>
<td>Tet2</td>
<td>chr19:3909853-9-39098554</td>
<td>+</td>
<td>ccacctcCACAACAGATATCCCTG (SEQ ID NO: 147)</td>
<td>Intron of Cyp2c55</td>
</tr>
<tr>
<td>Tet2</td>
<td>chr15:5918889-2-59188907</td>
<td>+</td>
<td>gagataaCACAACAGATATCCCTG (SEQ ID NO: 148)</td>
<td>Intron of E430025E21Rik</td>
</tr>
</tbody>
</table>
Table 6 Oligonucleotides used in this study. oligonucleotides used for cloning sgRNA expression vector

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>CACCGGCTGCTGCTAGGGAGCTCA (SEQ ID NO: 149)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 150)</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 151)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 152)</td>
</tr>
<tr>
<td>Tet3</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 153)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 154)</td>
</tr>
<tr>
<td>Sry</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 155)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 156)</td>
</tr>
<tr>
<td>Uty</td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 157)</td>
</tr>
</tbody>
</table>

Oligonucleotides used for Suveryor assay and RFLP analysis

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 159)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 160)</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 161)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 162)</td>
</tr>
<tr>
<td>Tet3</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 163)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 164)</td>
</tr>
<tr>
<td>Sry</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 165)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 166)</td>
</tr>
<tr>
<td>Uty</td>
<td>F</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 167)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGCAGCTCTCAGGAGCCAGCC (SEQ ID NO: 168)</td>
</tr>
</tbody>
</table>

Oligonucleotides used for making template for in vitro transcription

<table>
<thead>
<tr>
<th>Template</th>
<th>Direction</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9</td>
<td>F</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 169)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 170)</td>
</tr>
<tr>
<td>Tet1 sgRNA</td>
<td>F</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 171)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 172)</td>
</tr>
<tr>
<td>Tet2 sgRNA</td>
<td>F</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 173)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 174)</td>
</tr>
<tr>
<td>Tet3 sgRNA</td>
<td>F</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 175)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TTAGTTCTTCTTCTGACTGC (SEQ ID NO: 176)</td>
</tr>
</tbody>
</table>
DISCUSSION

The genetic manipulation of mice is a crucial approach for the study of development and disease. However, the generation of mice with specific mutations is labor intensive and involves gene targeting by homologous recombination in ES cells, the production of chimeric mice, and, after germ line transmission of the targeted ES cells, the interbreeding of heterozygous mice to produce the homozygous experimental animals, a process that may take 6 to 12 months or longer (Capecchi, 2005). To produce mice carrying mutations in several genes requires time-consuming intercrossing of single mutant mice. Similarly, the generation of ES cells carrying homozygous mutations in several genes is usually achieved by sequential targeting, a process that is labor-intensive necessitating multiple consecutive cloning steps to target the genes and to delete the selectable markers.

As summarized in Figures 4A-4B and described herein, three different approaches for the generation of mice carrying multiple genetic alterations have been established. Demonstrate d herein is that CRISPR/Cas -mediated genome editing in ES cells can generate the simultaneous mutations of several genes with high efficiency, a single-step approach allowing the production of cells with mutations in five different genes (Figure 4A). Three Tet genes were chosen as targets because the respective mutant phenotypes have been well defined previously (Dawlaty et al, Cell Stem Cell, 9:166-175 (2011); Gu et al, Nature, 477:606-610 (2011)). Cells mutant for Tett, 2 and 3 were depleted of 5hmC as would be expected for loss of function mutations of the genes (Dawlaty et al, Dev Cell, 24:310-323 (2013)). However, which of the Cas9-mediated gene mutations produced loss of function rather than hypomorphic alleles has not been established.
Also shown herein is that mouse embryos can be directly modified by injection of Cas9 mRNA and sgRNA into the fertilized egg resulting in the efficient production of mice carrying biallelic mutations in a given gene. More significantly, co-injection of Cas9 with Tet1 and Tet2 sgRNAs into zygotes produced mice that carried mutations in both genes (Figure 4B, upper panel). It was found that up to about 95% of new-born mice were biallelic mutant in the targeted gene when single sgRNA was injected, and when co-injected with two different sgRNAs, up to about 80% carried biallelic mutations in both targeted genes. Thus, mice carrying multiple mutations can be generated within 4 weeks, which is a much shorter time frame than can be achieved by conventional consecutive targeting of genes in ES cells and avoids time-consuming intercrossing of single mutant mice.

The introduction of DSBs by CRISPR/Cas generates mutant alleles with varying deletions or insertions in contrast to designed precise mutations created by homologous recombination. The introduction of point mutations into human ES cells, cancer cell lines, and mouse by ZNF or TALEN along with DNA oligo has been demonstrated previously (Chen et al, Nat Methods, 8:753-755 (2011); Soldner et al, Cell, 146:318-331 (2011); Wefers et al, PNAS, USA, 110:3782-3787 (2013)). Demonstrated herein is that CRISPR/Cas mediated targeting is useful to generate mutant alleles with predetermined alterations, and co-injection of single stranded oligos can introduce designed point mutations into two target genes in one step, allowing for multiplexed gene editing in a strictly controlled manner (Figure 4B, lower panel). This targeting system allows for the production of conditional alleles, or precise insertion of larger DNA fragments such as GFP markers so as to generate conditional knockout and reporter mice for specific genes.

It is likely that a much larger number of genomic loci than targeted in the present work can be modified simultaneously when pooled sgRNAs are introduced. The methods presented here provide for systematic genome engineering in mice, facilitating the investigation of entire signaling pathways, of synthetic lethal phenotypes or of genes that have redundant functions. A particularly interesting application is the possibility to produce mice carrying multiple alterations in candidate loci that have been identified in GWAS studies to play a role in the genesis of multigenic diseases. In summary, CRISPR/Cas mediated genome editing
allows for the generation of ES cells and mice carrying multiple genetic alterations and facilitates the genetic dissection of development and complex diseases.

Example 2

RNA-programmable DNA binding enzymes (CRISPRzymes)

Reported herein is the generation of an RNA-guided, programmable transactivator based on CRISPR/Cas system, CRISPRa, which provides a tool for modulation of a (one or more) nucleic acid sequence, e.g., gene activation, and serves as a proof of principle for CRISPR-based RNA-guided DNA binding enzymes (CRISPRzymes).

Results

dCas9ta guided by sgRNA targeting tet binding site activates TetO promoter

To build a CRISPR/Cas-based transcriptional activator, H840A of the human codon-optimized Cas9 nickase was mutated to generate nuclease-deficient dCas9 [PMID: 23452860] and a 3x minimal VP16 transcriptional activation domain (TAD) was fused to the C-terminal of the dCas9 protein (Figure 10A) to generate dCas9ta. dCas9ta was first tested on a tdTomato reporter under the control of Tet-inducible promoter with seven copies of tet binding site upstream of a CMV minimal promoter (TetO::tdTomato) (Figure 10B). HeLa and NIH3T3 cells with TetO::tdTomato transgene were generated by PiggyBac transposition [PMID: 17576687]. As a positive control for the reporter activity, these cells also constitutively expressed the rtTA-M2 transactivator that can induce tdTomato expression upon doxyeyelene treatment (Figure 11C panel ii; Figure 12B). Transfection of dCas9ta with sgRNA (sgTetO) complementary to tet binding site activated TetO::tdTomato reporter in the absence of doxycycline (Figure 11C panel iv; Figure 12D). Transfection of dCas9ta without sgRNA did not activate tdTomato expression, indicating the dCas9ta depends on sgRNA to bind to the target tet binding sites to activate tdTomato expression (Figure 11C panel iii; Figure 12C).

dCas9ta with sgRNA targeting Nanog promoter can activate both the NanogGFP reporter and the endogenous Nanog expression in NIH3T3 cells

To test whether dCas9ta can activate endogenous gene expression, dCas9ta chimeric expression construct was designed and cloned with 8 different
sgRNAs targeting Nanog promoter (sgmNanog) and transfected in NIH3T3 cells. As a comparison, a NanogGFP plasmid [PMID: 18594521] containing 1.2kb promoter of Nanog was co-transfected. Transfection of dCas9ta without sgRNA did not activate the exogenous NanogGFP reporter (Figure 11C; panel ii) or the endogenous Nanog gene (Figure 11B, column ii) while transfection of 8 constructs expressing dCas9ta and sgmNanog activated the NanogGFP reporter (Figure 11C, panel iii) and endogenous Nanog expression (Figure 11B, column iii).

[00247] dCas9 fusion with P-TEFb components also activate gene expression
[00248] To test whether dCas9 can be used to bring other protein domains to DNA to regulate gene expression, dCas9 was fused to Cdk9 and CycT, two components of the P-TEFb complex involved in the transcriptional pause release [PMID: 22986266] and their transactivation was tested activity on the TetO::tdTomato with or without dCas9ta (Figures 13A-13D). Transfection of dCas9ta resulted in 10% of tdTomato positive cells. Transfection of both dCas9Cdk (pAC72) and dCas9CycT (pAC73) also activated tdTomato expression, though to a lesser extent (2%). Co-transfection of three plasmids, pAC5 (dCas9ta), pAC72 (dCas9Cdk9), pAC73 (dCas9CycT), with sgTetO resulted in 13% tdTomato positive cells. This additive effect indicates that co-transfection with or fusion of additional transcriptional activators or transactivation domains to dCas9ta likely further augment dCas9ta transactivation activity.

[00249] Materials and Methods
[00250] Cloning
[00251] A two-step fusion PCR was used to amplify Cas9 Nickase ORF without stop codon from the pX335 vector, incorporate H840A mutation, EcoRI-Agel restriction site on the 5’ end as well as an Fsel site on the 3’end (EcoRI-Agel-dCas9-Fsel fragment). The 3x minimal VP16 activation domain coding fragment (TAD) was excised from a vector (Addgene: 20342) containing NLSM2rtTA coding sequence by Fsel and EcoRI digestion (Fsel-TA-EcoRI fragment). The two fragments were ligated into pCR8/GW/TOPO (Invitrogen) vector digested by EcoRI to generate pAC1 which contains the dCas9ta gene. The dCas9ta coding sequence was subsequently excised from pAC1 and cloned into pX355 vector (Addgene: 42335) by Agel-EcoRI digestion to replace dCas9 Nickase to create a chimeric vector pAC2 that expresses both the dCas9ta and the sgRNA. sgRNA spacers were
cloned into the Bbsl-digested pAC2 vector. For example, sgRNA targeting TetO (sgTet) was cloned by ligating phosphorylated and annealed oligos sgTet-F: caccGCTTTTCTCTATCACTGATA (SEQ ID NO: 179) and sgTet-R: aaacTATCAGTGATAGAGAAAAGC (SEQ ID NO: 180) onto Bbsl-digested pAC2 vector to generate pAC5. To replace the 3x minimal activation domain (3xmTAD) in dCas9ta protein for other protein domains, Fsel-EcoRI fragment from pAC5 or pAC1 was replaced by PCR amplicons of different domains or genes with Fsel and EcoRI added on the primer sequences. dCas9 was cloned by PCR amplification of dCas9ta with reverse primer before the 3xTA domains and cloned into pCR8GWTOPO to create pAC84 and pAC5 to create pAC89. Non-chimeric versions of dCas9 fusions were generated by LR Clonase-mediated recombination to a pmax-DEST vector (pAC90).

<table>
<thead>
<tr>
<th>Domains/ Genes</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>3xVP16(TA)</td>
<td>N/A (directly excised from a plasmid)</td>
<td>N/A</td>
<td>pAC1, pAC2</td>
</tr>
<tr>
<td>mCdk9</td>
<td>AAAGGAGCGGGGATGGCCA AGCAGTACGACTC (SEQ ID NO: 181)</td>
<td>AaaanGAATTCTcaGAAGACA CGTTCATAATCCG (SEQ ID NO: 182)</td>
<td>pAC72 (dCas9Cd9, U 6- sgTetO)</td>
</tr>
<tr>
<td>mCycT</td>
<td>AAAGGAGCGGGGATGGGAGGAAGAAGAACA (SEQ ID NO: 183)</td>
<td>AaaanGAATTCTCACTTAGGAAGAGGTGGAGAAGTGGTG (SEQ ID NO: 184)</td>
<td>pAC73 (dCas9CycT, U 6-sgTetO)</td>
</tr>
</tbody>
</table>

[00252] A reporter assay for dCas9ta activity
[00253] A TetO::tdTomato (plasmid pAC3) transgene and a EF la::NLSM2rtTA (plasmid pAC4) transgene were delivered into NIH3T3 (mouse) and HeLa (human) cells by PiggyBac transposition. sgRNAs were designed to target TetO binding site (sgTetO). pmaxGFP (Clontech) was used as a transfection control. Transfection was done using FuGene HD following manufacturer's instructions.

[00254] qRT expression analysis
[00255] Pellets were snap-frozen and stored at -80C. RNA were prepared from the pellets by RNeasy kit (QIAGEN). cDNA were produced by Superscript III RT (Life Technology). qRT were done in triplicates using Gapdh as a control.
qRT primers:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>mGapdh</td>
<td>TGTGTCCGTGATGGATCTCTGA (SEQ ID NO: 185)</td>
<td>CCTGCTTCACCCCTTTTTGTA (SEQ ID NO: 186)</td>
</tr>
<tr>
<td>mNanog</td>
<td>TTGCTTACAGGGTCTGTACT (SEQ ID NO: 187)</td>
<td>AACTGTTAGGAAGAATCTAGGGCT (SEQ ID NO: 188)</td>
</tr>
</tbody>
</table>

sgRNA designs, DNA targets, oligos, and plasmids used to target different DNA. Last three bases are PAM (5'-NGG-3') motif. Lowercase letters in the target sequences indicate changes made (first g) to allow efficient U6 transcription or for mutational analysis (other changes). Lowercase letters in the oligo sequences indicate overhang compatible to the BbsI-digested vectors. Target gene names with m prefix indicate mouse gene while those with h prefix indicates human genes.

<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Target sequences</th>
<th>Forward oligo</th>
<th>Reverse oligo</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgTet</td>
<td>Tet binding site</td>
<td>Gcttttactactatactgtaggg (SEQ ID NO: 189)</td>
<td>CaccGCTTTTCTCTATCTGATA (SEQ ID NO: 190)</td>
<td>AaaACATTAGTATAGAGAAAGC (SEQ ID NO: 191)</td>
<td>pAC5</td>
</tr>
<tr>
<td>sgmNanog-1</td>
<td>mNanog promoter</td>
<td>GTAATGCAAAGAAAGCTGTAAAG (SEQ ID NO: 192)</td>
<td>CaccGTAATGCAAAAGAAAGCTGTAAAG (SEQ ID NO: 193)</td>
<td>AaaACATACGCTCTTTTGCATTAC (SEQ ID NO: 194)</td>
<td>pAC34</td>
</tr>
<tr>
<td>sgmNanog-2</td>
<td>mNanog promoter</td>
<td>GATCTCTATGTGGAAGTTTC (SEQ ID NO: 195)</td>
<td>CaccGATCTCTATGTGGAAGTTTC (SEQ ID NO: 196)</td>
<td>AaaACGAAACTCCCACTAGAGATC (SEQ ID NO: 197)</td>
<td>pAC35</td>
</tr>
<tr>
<td>sgmNanog-3</td>
<td>mNanog promoter</td>
<td>GCTCTCCATTTGGAAACCTGTG (SEQ ID NO: 198)</td>
<td>CaccGCTCTCCATTTGGAAACCTGTG (SEQ ID NO: 199)</td>
<td>AaaACGTGGTTTCAATGTGAAGAGC (SEQ ID NO: 200)</td>
<td>pAC36</td>
</tr>
<tr>
<td>sgmNanog-4</td>
<td>mNanog promoter</td>
<td>GAGTTTTAAGATTAATGTAGAG (SEQ ID NO: 201)</td>
<td>CaccGAGTTTTAAGATTAATGTAGAG (SEQ ID NO: 202)</td>
<td>AaaACCTACATTTAATTTAACACTC (SEQ ID NO: 203)</td>
<td>pAC37</td>
</tr>
<tr>
<td>sgmNanog-5</td>
<td>mNanog promoter</td>
<td>GAGTTTTCCATGACCCGAGACTG (SEQ ID NO: 204)</td>
<td>CaccGAGTTTTCCATGACCCGAGACTG (SEQ ID NO: 205)</td>
<td>AaaACGTCTCGGGTACGTGAAGAC (SEQ ID NO: 206)</td>
<td>pAC38</td>
</tr>
<tr>
<td>Plasmid name</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC1</td>
<td>dCas9ta on pCR8GWTOPo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC2</td>
<td>Dual expression construct expressing both dCas9ta and sgRNA from U6 promoter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC3</td>
<td>TetO::tdTomato PiggyBac</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC4</td>
<td>EFla::NLSM2rtTA PiggyBac</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC5</td>
<td>Dual expression construct expression both dCas9ta and sgRNA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC34</td>
<td>Dual expression construct expression both dCas9ta and sgRNA3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC35</td>
<td>Dual expression construct expression both dCas9ta and sgRNA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC36</td>
<td>Dual expression construct expression both dCas9ta and sgRNA5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC37</td>
<td>Dual expression construct expression both dCas9ta and sgRNA6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC38</td>
<td>Dual expression construct expression both dCas9ta and sgRNA7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC39</td>
<td>Dual expression construct expression both dCas9ta and sgRNA8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC47</td>
<td>Dual expression construct expression both dCas9ta and sgRNA9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC48</td>
<td>Dual expression construct expression both dCas9ta and sgRNA10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC71</td>
<td>Dual expression construct expression both dCas9Apobecl and sgRNA11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC72</td>
<td>Dual expression construct expression both dCas9Cdkl and sgRNA12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC73</td>
<td>Dual expression construct expression both dCas9CycT and sgRNA13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC76</td>
<td>Dual expression construct expression both dCas9Gadd45a and sgRNA14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC78</td>
<td>Dual expression construct expression both dCas9mCBPHAT and sgRNA15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC80</td>
<td>Dual expression construct expression both puroR and sgRNA16 (control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC81</td>
<td>dCas9ER on pCR8GWTOPO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC82</td>
<td>dCas9p300HAT on pCR8GWTOPO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC83</td>
<td>dCas9TetlCD on pCR8GWTOPO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC84</td>
<td>dCas9 on pCR8GWTOPO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC85</td>
<td>dCas9EnR on pmax expression vector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC86</td>
<td>dCas9p300HAT on pmax expression vector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC87</td>
<td>dCas9TetlCD on pmax expression vector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC88</td>
<td>dCas9ta on pmax expression vector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAC89</td>
<td>Dual expression construct expression both dCas9 and sgTetO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[00259] CRISPRzyme dCas9 fusion peptides

[00260] Sequences: dCas9TA peptide sequence is shown below. The underlined sequence indicates the 3x VP16 minimal transactivation domains.

[00261] dCas9TA peptide sequence (Underlined sequence indicate the 3x VP16 minimal Transactivation domains)

[00262] MYPYDVPDYASPKKKRKKVEASDKKYSIGLAIGTNVS
GWAVIDEYKVPSKKFKVLGN
TRHSIKK
LNIGALLFDGTA
EAETRLKRTARRRRY
TR
RNRICYLQEFINNEMAKVD
DSFFHRLEESFLV
EEDKK
HERHPFIGNVDEVA
YHEKYPTIYHLRRKLV
DSTKD
LRLIY
LALAHMIKRFRGHFLIEG
DLNPDNS
DV
DKLFIQL
VQTYINQL
FEENP
VGDA
LSARLSKRRLENLI
AQLPGE
KKNGLFGNLI
ASLGTLPNFKSNF
DAL
EDAKLQLSK
KD
TYDDDLDNNL
AQIG
DQYADLFLAAKN
LSDAIL
LSDILR
VNT
ITEITK
APL
SASM
KRY
DEH
HQDLTL
KALVRQ
QLPEKYKEI
FDQSKNGYAG
YIDG
AS
QEEFYKFIK
PILEKMDGTE
EL
V
KL
NR
LRRKQR
TFDNGSIPH
H
LHA
LRQQDFYPF
K
D
N
EKIL
FRIPYYVG
PLARGNSRFAWMT
RKEETIT
PN
EVVD
KAG
A
QSFIE
RMTNFDKNLNP
EKLPKHSLLYEYFTV

YNTK
KY
VTEGMRKPAFL
G
QKKA
VIDLFLK
R
KV
VK
Q
K
EDYFKKIECFDSVEIS
GDVF
ASL
G
TYH
DLLKIKKD
F
LDNENEDILEIDIV
LTLF
DEREMIEERLKTYAWLFD
DKVM
KQLKRR
YTGWG
LSRKL
GIRDQKS
GK
TILD
KSDGFA
RNFMQLIHD
DSLTFKEDIQK
ACQVSGQG
DS
LHEIANLAGSPA
K
GI
QTV
V
DELV
V

MGRHKPENI
VEMARENQTTQKG
K
NRSRERMKR
IEEGIKELG
QLKEHPVE
NTQLQNEKLYLYL
QNGRDMYV
QDL
INL
SYDV
AIVPQFSLKD
S
ID
NKVLTRS
DK
R
GS
N
PSEEEVV
K
MKNYWRQLN
KLITQRF
DNLTKA
ERGGLSELDKAGFI
RQL
VETRQITKH
A
QIL
D
SMNT
YD
KLI
REV
K

ITALKS
KLVSDFRKDFQFYK
VREIN
YHA
HDAYLNAV
GTALIKKYPKLESE
FVYGDYKVYDVRKMIACKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLMPQVNIYKKTEVTGFGSKEISI
PKRNSDKLARKKDWDPPKYGFGDSPTVAYSVLVAVAKVEKGSKKLKSVK
ELLGITIUMRSSFEKNPFPIDLEAKGYKEVKKDLIIKLPYSLFELENGRKRML
ASAGELQKNGELALPSYVNFLASHYEKLKSPEDNEQKQLFVEQHKH
YLDEIEEQISEFSKRVILADANLDKVLSAYNKHRDPKREQAENIIHLFTLTLN
GAAAFAKYFDDTIDRKRTYSTEKLMDAITHIQSITGLYETRIDLSQLGGDSPK
KKRKVEASPADALLDFDLMLPAADLDDDFDLMLPAPADLDDDFDLMLP

G- (SEQ ID NO: 1)

[00263] dCas9Apobec1

[00264] MYPYDVPDYASPKKRRKVEASDKKYSIGLAIGTNSVGVAVIT
DEYKVPSKKFKVLGNTRRDHSSIKKLGLALLFDGSGTAEEATRLKRTARRRYTR
RKNRCYCLQEIFSNAKAVDSSFFHRLESFLVEEKKHERHPFGNVDAVEY
YHEKYPTJYHLRRKLVLVSTDKADLRLYIHLAHIKFRGFLIEGDLNPDNS
DVDKLFQILQVQNYLFEENPASGVDAKAILLSRLLKRRELNLQQLPGGE
KKNGLFGNIALSLGLTPNFKSNFDLAEADKQLLSKDTYDDDLDNLAAQIG
DQYADFLAANKLSDAILLSDLRVNEITKAPLSASIMKYDHEHQDLDLTL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGQASQQUIFYFKFIPKLMGDTE
ELLVKNREDLLRKRQTFDNPSIPHIQLHELHAILRQRQDFYPFLLKDNREK
EKILTFRIPYYVGPLARNGNSRFAWMTRKSEETITIPFNEEVVDKGASQSFIE
RMTNFDKLNPKVLPKHALLEYFTYVNETKVKYVTGMRKFAFLSAGE
QKKAIVDLDLFKTNRKTVPQKVLKEDYKIKFCCCCFSDVESIGVEDRFNASLTYH
DLLKIKDKDFDLDNEENILEDIVLTLTFLFEDREMIEERLKYAHFLFDKVM
KQLKRRRSTGWGRSLKRKLINGRDQSGKTILDVLKSDKGFANRNNFMQLIHD
DSLTFKEDIQKAQVSGQGDSLHELHIANLAGSPAIKKGILQTVKVVDLELVK
MGHRHKPENIVIEMARENQQTQGKQKNSRERMKRKIEEGIKELGSQILKEHPE
NTQLQNEKLRLYLLYQNGRDMDYVDQELDINRLSDYDVDAIIPQSFLKDDSID
NKVLTRSDKNSKRGSNDPSEEVVKMKHNYWRQLLNAKLITQKRKFNLTLKA
ERGGLSELDKAGFIQKRLVETRQITKHVAQILDSRMNTKYYENDKLIREVKV
ITALSKLVSDFRDFQFQYKREINNYHHAAHAYLNAVGTALIKKYPKLEE
FVYGDYKVYDVRKMIACKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLMPQVNIYKKTEVTGFGSKEISI
PKRNSDKLARKKDWDPPKYGFGDSPTVAYSVLVAVAKVEKGSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKDLIKLKPYSYLFELENGRKRML
ASAGELQKGNELALPSKYVNFLASHYKEKLGSPEDNEQKQLFVEQHKH
YLDEIIIEQISEFSKRVILADNLKVSAYNKHRDKPIREQAENIIHLFTL
GAPAAFKYFDTTIDRKRYSTKEVLDATLHIQSITGLYETRIDLQLQGDSPK
KKRKVEASPAMSSETGPAVAVDPTLRRRIEPHEFEVFFDPRELKRKETCL
LYEI
NWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERYFRPNTRCSITWFLSW
SPC
GECSRAITEFLSRHPYVTITIARLYHTHDQRNRQGLRLISSGVITIQIMTE
EQYCYCWRNFVNYPSNEAYWPRYPHLWVYKLYLEYCIILGLPPCLKLRR
KQPQLTFTTTLTQCTHYQRIPPHLWATGLK (SEQ ID NO: 2)

[00265] dCas9Cdk9
[00266] MYPYDVPDYPASPKKRRKVEASDDKKYSIGLAIAGTNVSFGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKKNLIGALLFDGSETAEATRLKRTARRRY
TRKRNRICYLFQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAKHEKYPTIYHLRKLVDSTDKDLRLIYLALAHMIKFRGHFLIEGDLIPD
NSDVDKLFIQLYQTYNQLFEENPASGFDAKALSRLSKSRRLLENIAQLPG
EKKNLFGNLIALSLGTLNPFLKSNFDLAEADKLQKSCTYDDDLDNLAQIG
DQYADLFLAAAKNSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD
TTLKALVRQQLPEKYKEIFFDSQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKRQRTFDINGSPHQLHIELHAILRRQEDFYPLK
DNREKIEKILTFRIPYYVGPLARNSRFAWMTRKSETITTPWNEEVVDKGA
SASSIFIE
RMTNFDKNLNPVEKLPKHLLLEYFTVINELTVKÝTVEGMRKPAFLS
GEGKKAIVDLLFKTRKVTVKQLKEDYFKKIECFDSVEISGVEDI
FKNASLGTYHDLKIKKDQLDNEENDELDEIVLTLTFEDREMIEERKTYAH
LFDQDKVMMQLKRRRTYGWGRSRLKGINGIRDKQSGKTILDLSKDG
ANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGIL
QTVKVVDKV
MGHHPENIVIEMARENQTTKQKGGNSREMRRKRIEGIKELGSQLKEHPV
ENTQQLNEKLKLYLYQNRDMYVQEDLINRLSDYDVDAIVPQSFKLDDS
DNSKVLTRSDKNKSDKSNPVSEEVVKMKNYWRQLLNAKLITQRRKFDN
LTKAERGLSLEDZAKFIRQLVETRQITKHVAQILDSRMNTKYENDKLIR
VEKVFILTSLVSDKFTRDFQFYKREINYYHHAIHAYLNAVGVTLAIJKK
YPKLESEFVYGYKDYVFVRKMIAKSEQEIGKATAKYFFYMSNMNFKTEIT
LANEIRKRPLIETGETGEIVWDKGRDFATVRKVLSPMQVNIVKKEVQT
GGFSKESIL
PKRNSDLIRKKDDWPKKYYGGDFSPTVAYSVLVVAKEKGSK
KLSKV
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIEIQISEFSKRVILADANLDKLVSAYNKHRDKPIREQAENIIHLFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLATLHIQSITGLYETRIDLSQLGGDSPK
KKRKVEASGPAMAPYSDECPCFCDVTKYEKLAQKGQGTGFEGVFKAKHR
QTTQKVALKVLMENEKEGFPITALREIKILQLKHENVVNLIEICRTKASPYY
RCKGSIYLVDFCEHLAAGLLSNVLVKFTLSEIKRVMQMMLONGLYYIHRN
KILHRDMDKANVLITRDGVLKLDGFLARAFSLAKNSQPNRYYTNRVVTLW
YRPPELLLGERDYGGPIDLWAGACIMAEMWTRSPMQNGTEQHQLALISQL
CGSITPEWVPNKYELEFKLEVKQOKRVDKLYVRDPYALDLIDKL
LVLPQAQRISDDALNHDFFWSDPMSDMLKGMSTHTSLMFEYLAPPARRKG
SQITQQSTNQRNPATTNQTEFERVF(SEQ ID NO: 3)

[00267] dCas9CycT

[00268] MYPYDVPDYASPKKRKVEASDDKYSGLAI GTNSVGAWAVIT
DEYKVPSKKFKVLGNTDRHIKKN LI GallFDSGETA EATRLKRTARRRYTR
RKNRICYLQEISNEAKVDDSFFHRLEESFLVEEDKKHERHIPFIGNVDEVA
YHEKYPTIYHRLRKKLVDSTDSDLRLIYALAKHMFKRGFLIEGLDNLPSDS
DVKLFIFQLVQTYNQLFEENPASGVDAKASARLSSRRENLIAQLPGKGE
KKNCLFGNLIALSLGLTPNFSNFSLADEAKLQLSMDTDDDDLDNLLAQIG
DQYADMLFLAKNLSEAILLSDLRKNTEITKAPSAMIKRYDEHHQDILTT
KALVRQQLPEKYEI FFDSQKNGYAGY IDGGA SGQEEFY KIFPILMKDM TGE
ELLVKLNREDLRKQRTFDNSIHPQHIGHLHELHRLRRQEDFYFPLKDNREKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETTPWNEVDDKGA SAGAQSFIE
RMNFDKNLPNEKVLKHKLVENYETVYNELTVKVYVTEGMKPAFLSGE
QKKAIVDLFLFKTNKRVTVKQLKEDFYKKEICFDSVEISVGEDRNASLGT YH
DLLKIJKDKFLDNEENILEDIVTLTLFDREMIEERLKTAYHLFDDKV
MKQILKRRITYGWGRSRLKGIRDKQSGKTI DLDFKSGF ANRNFMQLIH
DSLT FKEIDIQKAAQVSGQGDSLSHEIANLGA SPAIKKGILQTVKVD ELKVK
MGHRKPPENIVEMARENQTTQKGQKNSRERMKRIEGIEKLGSILKEHPVE
NTQLQNEKLYLYQLNGRDMYVDQELDINRLSDYDVDAIPQSFKLDDSID
NKVLTRSDKRNGSDKDNVPSEEVVKMKNYWRQLNALKITQRFDLNLTKA
ERGG LSELDKAGFIKRQLVETRQITKHVAQILDSRMNKTYYDENDKLLREVKV
ITLKS VLSDFRKDFQFYKVREINYHHAHDAYLANAVGTA LKYPKLESE
FVYGDYKVYDVRKMIKAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLMPQVNIVKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWDPKKYGGFDSTPVAYSVLVVAKEGKSKKLKSVK
ELLGITIMERSSEFKNPIFDLEAKGYEVKVDLKIPLKYSFLFENNGRKRML
ASAGELQKGNELAPSYKNVFYLASHYEKLKGSPEDEQKQLFVEQHKH
YLDEIEQISEFSRKLVILADNLDKLVSYANKHRDKPIREPQAENIIHLFTLNL
GAPAAFKYFDTTIDRKYRSTKEVLATLIHQSITGLYETRIDLSQLGGDSPK
KRRKVEASGPAEMEGKRNKNNKRWYFRTQLENSPSRFGRVDSKELSYRQ
QAANLQLDMGQRLNVSLQINTAIVYMHRFYMIQSFTQFHRYSMAPAAFL
AAKVEEQPKKLEHVIKVAHTCLHPQESLDPTRSEAYLQQVQDVLVILESILQT
LGFLETIDHPHTHVVKCTQLVRASKDLAQTSYFMATNLSLHTTFSLQYTPPV
VACVCIHIAKWSNWIEPVSSTDGKHWEYVDATVTELLELDLTHEFLQILE
KTPSRLKRINRWRAYQAAAMKTPDGRADENTSEQTILNMISSQTSDDTTIAQ
LMSTMSTASTSAVPSLPSSESSSSTTSVDMLQGERWLSQSRPPFKLEAAQQHR
TSESALIGVDHSLQDGGSAFSGSQKASVSQPSAKVSLKEYRAKHAEEAA
QKRQLENENAVKSQYAAQNNLLSHDHSSVILKMPIESSENPERPFLDK
ADKSALKMRPLPSGDKAVSSKPEEIKMRKVHSAGDKHNSIERSVTSRE
HKEKQRTHPNSNNHHHHHNNHHRHSHLQPLAPGVSKRPSDPKHHSSQTSTLH
KTYSLSSLSSSSKRGRPPEETGAAVFDHAIAKSTKSLNFPFPPLPTMT
QLPGHSSDTSLPFFQPSCKTRVPHMKLDKGPGANGHATQSIDYQTVDN
MLHSLLSAQGVQPTQAPADEFVHSYGEYMNPRAIIS
RSGTTDKPRPPPLPSEPPPLPLPK (SEQ ID NO: 4)

[00269] dCas9EnR
[00270] MYPYDVDPYASPKKKRKVEASDKKYSGILAIGTNSVGWAIVT
DEYKVPSKFKVLGNTDRSHSIIKKNLIGALLFDSDGETAEATRLKRTARRRYTR
RKNRICYLQEFSNEMAKVDSSFHRLEESFLVEEDKKHERHPFIGNVEVA
YHEKYPTIYHLRRKLVSTDKADRLYLIHLAHMIKFRGHLFIEGDLNPDNS
DVDKFLIQIVQTQYNQLFEEPINCEVGDAKAILSARLSKRRLNQLIAQPLGE
KKNGLFGNIALSLGLTPFNKSFDLAEADKQLSKDTTYDDDLNNDLAQIQG
DQYADLFLAAKNLSDALLSDLRVTNTEITKAPLSAMIKRYDEHHQDITL
KALVRQQLPKYEKIEFQDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNRDLLLKQRQTFDNGSIPHQIHLGELHAILRRQEDFYPFPFLKDREK
EKILTFRIIPYYVGPLARGNSRFAWMTKSEITTPNFEEVDFKGAASAQSFIE
dCas9Gadd45a

MYPYDVPDYASPPKRRKVEASDVKYSIGLAIGNTSVGWAVIDEYKPSKKFKVGLNTDRSHIGKLITIGALLFDSGETAEATRLKRTARRTYTRRKNRICYLQEQIFSNEMAKVDDSFFHRLEESFLVEEDKHHERPIFIGNIVEVAYHEKTPYHRLRKKVLDSDRKLRLIALAHMIKFRGHFLIESGDNLPSNDSDVDKLFIQLVQYTVENPINASGVDAKAILSAKLSRKLSELNLIQALPSKKNLFGNLIALSLGTPNFKSNFDLAEADKLQLSPLNPSYQQQAATATAAMMLERANFLCNFPAAYPRIHEIVQSLRRAANAVPPMSSKSMDSANPEKALSGMPQKLEQKLLISEEDLN (SEQ ID NO: 5)
ELLVKLNREDLLRKQRTFNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNEVEVDKGASAQSIE
RMTNFDKLNPNKVLHKLYLEYFTVYNETLKVKY VEGRMRKPAFLSGE
QQKAI VDLFLFKTNKRKTVQLKEFYKKIECFDSVEISGVDE RFNASLGYTH
DLLKIIKDKFLDNEENEDILEDILTLTFLFEDREMIEERLKY AYHLFD DKVM
KQLKRRYRTGWRSLRKLINGRDKQSGKTDLKLSDKGANDFNRFMQLIH
DSLTFKEDIQKAQVSGQGDSHELHHEIANLAGSPAIIKGILQTVKVDRLVK
V MGRHKPENIVIEMARENQTTQKGQKNRERMKRIEGIKEGLSGQLKEHPVE
NTQLQNEKLXYLYLQNGRDMYVDQELDINRLSDYDVAI VPQFSLKDDSID
NKLVRTRSDKNRRGSNDVPEESVEVKKMKNYWRQLLLNKLITQRFKLNKTA
ERGLSE LDKAGF IKRQVL VETRQITKHAQI LDSRMTKYDENDKLIREVK
ITLKSLKVDSTDKDFQ FYKVR EINNYHHAHDAYLNAVGTALIKKYPKLESE
FYVGDYKVKVDVRKMIKA SSEQEIGKATAKYFFYSNMNFFKTEITLANEGIRK
R PLETNETGIEIVWDGKRDFAVFTRVKLMPQVNVKKTEVQ FTTGFSKESIL
PKRNSDKLRIA KDWDPKKG FYGDPS TAYSVLVVA KVEGKS KKLKV
ELLGITTIMERSSFKEKPIDFLEAKGYKEVKKDILKL PKYS LFELENGKRML
ASAGELQKNGNELAPS YVNFLASYE YKLGSPDENEQKQLFVEQKHK
YLDEIEE QISEFSKRVLADANL DKVLISAY NKHRDKPIR EQAENIIHFLTLN
L GAPAAFYFD TIDRKR YSTKEVLDA TI HQSITGLYETRIDLSQLGGDSPK
KKRKVEAS GPAMT LEFS AAEQKTERMDTVGAL EEVL SKARSQRTITVGV
YEAAKLLNVDPDNVVLCLLAAD EDDRRDVALQIHFTLRAC CENDINILRV
SNPGRLAELL LLENDAGPAESGGAAQTPDLHC LVTNP HS SQW KDPAQLSQL
ICFCRESRYMDQWVP VINLPER (SEQ ID NO: 6)

[00273] dCas9p300HAT

[00274] MYPYDVPDYASPPKKKVRVEASDKKYSIGLAI SGNTSVGVWAVIT
DEYKVS KKFVKLGN TDRHSIKK NLGALLFDG TAEATRLKR TARRR YTR
RKNRI CYLQEIFS NEMAKVDDSFFHRLEESFL VE EDD KHER HPIFGN IV DEVA
YHEKYPTIYHL RKKLV DSTDKADLRLIYLAHMIKFRGHFLEIGDLNPD N
DVDKL F IQL V Q T Y NQLE F E E PINA SGVD AKILS ARL S KS S RLLENLIAQLPG
EEKNFLGNLIALSLGLTP FKNSF DLA E ADKLQ LSKD TYD DDLDN LLAQ I G
DQYADL FLAA KNL SDA ILL LDS ILVRN TEITKAP LSAM KRYDEHH QDL TTL
KALVRQQLPEKYKEIFFDQS K N GYAGYIDGG ASQEEF Y KFIK PILE KMDGT E
ELLVKL NREDLLRKR QT F DNGS I PHQI H L GEL HAIL RRQEDFY PFL KDNREKI
dCas9mTet1CD

MYPYDVPDYASPPKRKVEASDDKYSIGLAIGTVNSVGVAVIT
DEYKVPSSKFKVLGNTDHRSIKKNLIGDFFDSGETAATRLKRTARRRYTR
RKNRICYLQIEFSQMAKVDSSFFHLRESFLVEEDKKHERHPFIGNVDEVA
YHEKYPTIYHLRKLVDSTDQADLRILYLALAHMIKRGFHFLIEGDNLPPDNS
DVKDGLIQVQTYNQLFEENPASGVDAKAILSRLSRSRLENLIALQPKGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSDKTDYDDDLNDALLAQIG
DQYADLFLAAKNSDAILSLDILVNTIETKAPLSAMIRKREDEHQLDRTKL
KALVRQQLPKEKYKEIFDFQSKNGYAGYIDGGASQEEFYKFKPIKLEKMDGTE
ELLVKLNRDELRRKQRTFDNGSIPHQIHLGELHAILRRQDFYFPLKDNEKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPPWNEEVDKGASAQFIE
RMTNFDKNLPNEKVLKHSLLYEEYFTVYNETKVKYVTEGMRKPAFLSGE
QKKAIVDDLKFTNREVTQKLVEYFKKIECFDSVEISGVDEADFNASLGTYP
DLLKIIKDFDLDNEENILEDIVLTTLFDREMEERLKTYAHLFDDDKVM
KQLKRRRTGWRSLRKLINGIRDKQSGKTIDLFLKSDGFANRNFMQLHID
DSLTFKEDIQAQVSQGDSELHEIANLAGSPAIIKGIQOTVDKVVDDELVKV
MRHHPENIVIEMARENQTQKQGKNSRERMKRREEGIKELGKSILKEHPVE
NTQLQNEKLYLYLQNQDRMDYVDQELDINRSLSDYDVDAIVPQSLKDDSID
NKVLTRSDKNQRGSKDNSQSEEVVKMKNLYWRQLLNAKLIKTEQRKDFDNSLTKA
ERGGLSELDKAGFIRQVLQETRQITKHVAQILDSRMNTKYDENDKMLEK
ITYLKSCLVSDFRKFDFQYFKVREINNYHHAHDAYLNHNAVGTALIKPKLESE
FVYGDKYVYDVRKMIAKSEQEIGKATAFYFFYSINMMFKEITLANGEIRK
RPLIETNGETGEIVWDGKRDFATVKVLSMPQVINVKTTEVQTVGKSSIESL
PKRNSDKLIARCKDWDPPKYGGFSDPTVAYSVLLVAKVEKGSKKLKSVK
EILLGITIMERSSSFKNDPDIFLEAKGYEVKWKLKDQYKLPSLFELENGRKRML
ASAGELQKGNELALPSSKYVNLASYEKLNGSEDNEQKQLFVQKCHRK
YLDEIEEQISEFSKRVLADADNLKVLASAYNHKDRKPIREQAENIILFTLTLN
GAPAAAFKQYDFTTDIRKRYTSTKEVLATLHQSITGLYETRIDQSLQLGDSPK
KKRKVEASGPAEAAPCDDCGGTQKEGPPYYTHLGAENGPSVAVRELMTRF
GQKGGKAIREKIVFTGKGKSSQGCPVAKWVIRRSGPEEKLICLVRERVDHH
CSTAVIVVLILLWEGIPRMLADRLYKELTENLRYSNHPTDRRTCLNKKRTC
TCQGIDPKTCGASFSGCSWSMYFNGCKGSRNSPRKFRALAPNYPLNYYK
RITGMSSEGSVKTGWIPDRKTLISREEKQLEKNLQELATVLAPLYKMAP
VAYQNQVEYEEVAGDCRLGNEEGRPFSGVTCCMDFCAHSHKDIHNMHNGS TVVCTRLADGRDTPENPDEQLHVLPLYRLADTDFGSVEGMKAKIKSGAIQ VNGPTRKRLRFTEPFPVRCGKRAKMKQHNKSHNSHNTKSFSSASSSTSHLK DEUSTDCFLQASSSTCTSYKSTAGGFAETSSILHCTMPGSAHSGGANAA GECTGTVPQPAEVAAAHPHQSLPTADSPVHAELTISPQELTNSQNQQLPLLS NSQKLASCQVEDERHPEADEPQHPEDDNLQPQDEFWSDSEEIYADPSFGGV AIAPIHGSVLIECARKELHATTSLRSPKRGVFPRVSLVFYQHKSLNKPHGF DINKIICKCKKVKTPKAPDRECPDVSPAEQLSHQPSVASTLTDNVTVSP YSLTHVAGPYNRUV (SEQ ID NO: 8)

[dCas9hACIDA]

[00277] MYPYDVPDYASPKKKRKVEASDKKYSIGLAIAGTNSGWAVITDE YKVPSSKKFVKVLGNTDHRSHKKNLIGALLFDGTAETEALKRTARHYPRR NRICYLQEIENMSNAKVDSSFHRLEESLVEEDKKHERHPFGNIVDEVAYH EKYPTYHRLKQLVDSTDKLRLIYLALAHMIKFRGFILEDLNPNSDV DLKLIQLVQTQYPFEEEPNASGVDKAILSARLSSKRRLENNILQPGK NGLFGNLIALSLGLTPNFSDKNFLAEDKQLSKTDDDLDNLQAIGQD YADLFLAAKNSLSDAILSDLRVEITEKAPLSASMKRYDEHHQDLDLTKAL VRQQLPEKYKEIFDFQSKNYAGYIDGBQEEFYKFIKPILEKMDGTEELL VKLNREDKIRKQRTFDPSHPQHQLHELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMT McNTRKSEETITPWNFEEVVDKGAQAQSFIERM TNDKLNPLNEKVLKPHSLLEYFTVYNETLKVKYVTGMRKPAFLSSEQK AIVDLFFKTNKRVTINKLQKEFYKKIECFDSVEISGVEDRFNASLGTYD LLL KIJKDKFIDNEEDELDIVLTLLFEDREMIERLKTAYAHFDKVMQKL KRRYTGWRGLSRKINGIRDQPSGKTLDFLDGFSANRMQILHDSSLT FKEDIQKAQVSGQGDSLLEHIANLGSAPKKGILQTVKVVDLKVGMRH KPNENIEMAREQTTQKKGKNSRMEKKKRIEIKELGQILKEHPVENTQ L QNEKLILYLQLNQGRDMYVDQELNRLSDYDVAIPQSFLKKDSIDNVKL TRSNDKNSKSDKVNPSEEEVFKKMKNYWRQLNKLITQRFKDNLTKAERGG LSELDKAGFIKRQLVETRQITKHVAQILDTSNTKDYDNEKLRVEKVTLL SKLABFRTDKDFQFYKREINNYHHABDAYLNAVGTALKYRPYKPLESEFYV GDYKVDVRKMIASEQIEGKATAKYFFYSNIMNNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGGRDFATVRKVLMPQVNIVKKTEVQTGGFSKESILKPR NSDKLIARKKWDPPKYYGGFDSTPVTAYSVLVVACKVEKGKSKKLSVKEU
dCas9hDMNT1

[00279] dCas9hDMNT1

[00280] MYPYDVPDYASPKKKRKVEASDDKKYSIGLAIGTNSVGWAVITDEYKVPSKFFKVLGNTRHISIKKNLIGALLFFDGETAEATRLKRTARRRYTRRKIRICYQLQIFSNEAMKVDSSFHRLEESFLVEEDKKHERHPFGNIVDEVAYHEKYPTYHRLKVLVDKADRLRIYLALAHMIFRGHFLIEGLDLNDNSDVKDFIQQLVQYQNLFEENPINASGVDKAILSLSKSRRLENLIAQLPGEKKNLFGNIALSLGLTPNFLKSNFDLAEDAKQLSKDYIDDLNDLAAIQIDQYADLFLAAKLNSDAILLSLDIRVNTETIKAPLSASMIKRYDEHHQDLT LTLTALVRQQLPEKYEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILKMDGTEELLVKLNRDELLRKRQTFDNGSIPHQIHLGELHAILRRQEDGEYPFLKDNREKEKILTFRIPYYVGPL ARGNSRFAMTRKSEETITPWNEEEVDKGASASQSFIERMTNFSDKLPNEKVLKPHELLE YETVYNEITKVKVYTEGMRKPAFLSGEQKKAIVDLFLKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFDLNEENEDILEDIVLTTLTFEDREMEERLKYTAHLFDDDKVMKQLKRRRTGWGRSLRKLINGIRDKQSGNTKTLDFLKSDFGANRNFMLIQHD DSNLTFKEDIQKAQVSGQGDSLHEHIANLAPAIKKGILQTVKVDLKV MGRHKPENIVIEMARENQITTQKGGKNRSERMKRIEIGKILEGSOI KEHPVE NTQLQNEKLYLYLYQNRDMDYVLQDLEINLSDYDVDAIYPVQSPFLLDSSIDNKLVTIRSDKNKSDNPSEEEVKKKMKNYYLRQNNAKLITQRKFDNLTKA EREGGLSELDKAGFIRKQLVETRQITKHVAQILDSRMNKTQENDKELREVKVI TLKSKLVDKFKDFQFYKVREINNHHAHDAVLNAVGTA LIKYK PLESEFVYGDYKVVDVRKMIAKSEQEIGKATAKYYFFYSNMNFKFTITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKLSMPQNVKKTEVQTGGFSKESILPKRNSD KLIARKKDDWDPKYYGGFSDSTVAYSVLVVAKEKGSK KKLKSVK
dCas9hDNMT3a

MYPYDVPDYASPKKKRVEASDDKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIIKKNLIGALLFDGTAEEATRLKRTARRRYTR
RKNRICYQLQFFSNEMAKVDSSFFHRLEESFLVEEDKKHERHIPFGNVIDEVA
YHEKYPTIYHLRKLVDSTDKDCLRLLYIYALAHMIKFRGHLIEGDLNPDNS
DVKLFIQQLQYQTNPEPASVGDASLARKLSSRLLENLIAQLPGK
KKNGLFGNLIALSLGTLPHNKSNDLAEADAKQLSukturDYNLDNNLAQIG
DQYADILFAAKNLSSDAILRVTNIEITPLAPSLASMIKRYDEHHQDLTLL
KALVRQQLPHEYKEIFDDQSKNGYAGYIDGGASEQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKKQFTDNGSIPHIQLGENHAILLRQEDFYFLKINNREKI
EKILTFRIPYYGVPLARGNSRFCAWMTRKSETITTPNVFEEVDKGASASQFIE
RMTNFDKKLNPENKLVPKHSSLYYEYTFTYVEMLTKVKYVTGMRKPAFLSGE
QKKAIVDDLKFTNKVTQKVLKDYYFKKIEFDSVEISVGVEDFNASLGTY
DLLKLKDHFNLDEENEDIHTLTLTFEDREMIEELKTYALHFDDKVM
KQLKRRRRTGWGRLSRKINGRDKQSGKTIIDLKSDGFANRNFQMILHID
DSTLFKEDIQKAAQVSGQGDLSHEHIANLAPAIKKGILQTVKVLDELVKV
MGRHKPENIVIEMARENQTTQKGGNSRERMKRIIEEEKGLSGQKEHPVE
NTQLQNEKLYLYQLYQNGRDMYDQELDINRLSDYDVDAIVPQSLKDSDID
NKVLTRSDKRNKSDNPESSEEVKMKNYWRQLLNALKATQRKFDNLTKA
ERGGLSDELKAGFIKRQVLETRQITEKVAQILDSRMNTKYEDDEKLVREKV
ITLKSLLLLSFLRDQFYKVIENNYHHADAYNAVGTSGDLTVK
FVYGGDYKVYDVrMKIAKSEQEIGKATAKYFFSNIMNFFKTEITLANEIRK
RPLIETNTGEIVWDGKGRDFATVRKVLSPMPQVNIKKNTEQFTGGSKESIL
PKRNSDKLIRKDGDPKKGFFDSPTVAYSVLVVAKEAKGGSKKLSVK
ELLGITIMERSSFNKPDIFLEAKGYEKVKKDLIKLPYSLFEENGRKRM
ASAGELQKGNELALPSKYNVFYLASHYEKLGKSGPDNEQKVQLFVEQHKH
YLDEIEQISEFSKRVLADANLKDVSAYNHKRDKPIREQQAOENIIHLFTLNL
GAPAASKFYFDTITTDRKRYTSTKEVLDAHLHQSITGLYETRIDLSQLGGDSPK
KKRVEASGPAMPAMPSSPGDTSSSAAEEDRKKDGEEQEEPGRKREEQ
PSTTARKVGRPGRKRHKPPVESGDTPKDPAVISKSPSMAQDSGASELPNGD
LEKRSEPQPEEGSPAGGQKGGAPAEGGAAETLPEASRAVENGCCTPKEGR
GAPAEAGKEQKETNIESMKMEGSGRGLRGLMGESSLQRPMRPLTFQAG
DPYISKKRDEWLRKREAEEKAKVIAFMNEENQGPGESQKVEEAS
PPAVQQPTDPASPTVATTPVEPGSDAGDKNATKAGDEPEYEDGRGFGIGE
LVWGKLRFSGWLPGRJVSWWMTGRSRAAEGTRWVMFGDGKFSVCCVE
KLMPLSSFCSAFHQTATYNKQPMYRKAIYEVLQASSRAKLFVPHCDSDES
DTAKEVEQNKPMIEWALGGFQPSGPKGLEPEEPKEKNYPEKYITMWMVEP
EAAAYAPPPPAPKPKSTAEKPVKEIIDEERTRELRVYEVQKCRNEDICIS
CGSLNVTLEHLPFGVGMQNCNNCFLECAAYQYDDGDYQSYCTICCGGREGV
LMCGNNNCRCFCVCCVECVDDLVPGPAQAQAIAKEDPWNCYMCGHKGTGGL
RRREDWPSRLQMFFANNHDQEFDPKKYPVPAPAEKRKIRVLSLDFGIATGL
LVKLDLGIQVDRYIASEVECDSTSVGMRHQKGKMYGDVRSTQKHIQEW
GPFDLTVGSPCNDSLIVNPARKGLYEGTGRLLFEFYRLHDARPKEGDDRP
FFWLFENVAMGVSDKRDISRFLESNPMIDADEVSAAHRAFYWHNLPG
MNRPLOSTVNFDKLEQCELHGRIKFSKVRTITTSNSIKQGKQDKQHPVF
NEKEDILWCTEMERVFGFPHYTDVSNMSLRQRLLLGRSWSVPVIRHLFA
PLKEYFACVID (SEQ ID NO: 11)

[00283]  dCas9hDNMT3b

[00284]  MYPYDVPDYASPKKKRKVEASEDKKYSIGLAIAGTNSVGWAVIT
DEYKVPSSKFFKVGLGNTDRHLSKNIKLIGALLFDSGETAEATRLKRTARRYTR
RKNRICLYQEFSNEMAVKVDSSFFHRLEESFLVEEDKHERHPFGNIVDEVA
YHEKYPTIYHLRKLVDSTKADRLRLYIALHAAHMIKRFGHFLIEGDLNPDNS
DVKDLFIQLVQTYNQLFEENPINASGVDAKILSARLSKSRRENLIAPLGE
KKNGLFGNLIALSLGLTPNFKSNFDELADAKLQLSKDTDYDDLDNLQAIG
DQYADLFLAAKNLSDAILSLDRLVNEITKAPLSMIKRYDEHQLDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFKIPLEKMDGTE
ELLVKLNREDLLRKRQRTFDNGSiPHIHLGHELHAILRRQEDFYFPFLKDNREKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPNFEEVVDKGASAQSFIE
RMTNFDKNLNPENKVLKSHLYEFYFTYNELTKVYVTEGMRKPAFLSGE
QKKAIVDLFLKFTNRKVTKVQKLEDYFKKIECFDSVEISGVEDRFNASLGYH
DLLKIIDKDFDLNEENEDILEDIVLTLTLDREMIIEELKTYAHLFDKVM
KQLKRRRITYTGWRGLSRKLINGIRDQKQSGKTILDFLKDGSFANRNFMQLIHĐ
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVEMARENQTQKGGKNSRERMKRIEGIKELGSQILKEHPVE
NTQLQNEKLYLYLQNGRMDYVDQELDINRLSVDVDAIPQFSLKDDSID
NKVLTRSDKRNKSDNPSEEVVKMKNYWRQLNNAKLITQRKFDNLTKA
ERGGLSELDKAFIQVLRETQIKTHVAQILDSRMTNKYDENDKLIREVKV
ITLKSKLVSDFDRQFYKVKREINNYYHAADYLANAVGRTALIKKYPKLESE
FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFSYNINNFKEITEILTANGEIRK
RPLIETNGETGIEIVWDKGFRDFATVRKLMPQVPNIVKKEVTQTGGFSKESIL
PKRNSDKLIAIRKRKWDPPKYGFDSPTVAYSLVVAKEGSKKLLKSVK
ELLGITIMERSSFKEKPIDFLEAKGKEVKKDLIIKPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNLASHYEKLKGSPDENEQKQLLFVEQHKH
YLDIEIEIQSEIFSKRVILADANLKDVSAYNKHRDKPIREQAENIIHFLTLNL
GAPAAFKYFDPTTDIRDKRYPTSTEVLDAATLHIQSTIGLYETRIDLQSLGGDSPK
KKRKVEASGPAMKGDTRHLNGEADAGGREDISLYNVGACSDFQSSDSPPLEAIP
RTPEIRGRRSRSRLKREVSSLSYQTQDLTDGDGDGDGDGDSDTPVMPKLFRE
TRTRTSEPAVRTRNNNSVSSRERHRPSPRSTRGQRGRNHVEDSEPVEFFATRL
RRRATASAGTPWSPSPSSYLTIDLTDDTEDTHGTPQSSSTPYARLAAQDQQG
GMESPQVEADSGDGDSSEYDQGKEFGIDLVWGGIKGFSSWPAMVSSWK
ATSKRRQMGRMWRVQFQDGKSFSEVSAKLVALGLFQFLNATFLNKLV
SYRKAMYHALEKARVRAKTFPSSPGDSLEDQLKPMLEWAHGGFKEPTGIE
GLKPNNTPENKTRRTADDSDATSCHYCPAPKRLKTNCYNNKDRGDGDEDS
REQMASDVANKSSLEDGCLSCGRKNPVSHFPLFEGGCLQCTCRDFLEELFY
MYDGGYQSYCTVCCEGRELLLCSNTSCCRFCVECELEVLVGTGTAEEAK
LQEPWSCYMCLPQRCCHGLRRKWDNVRLQAFFTSĐTGLEYEAPKLYPAIP
AARRRPVRVLSLDFGIATGYLVKELGIVGKYVASEVCEESSAVGTVKHNEWI
KYVNDRNITKKNNIEWLPFDLVIGGSCDLSNVPARKGLYEGTGRILFFEF
YHLLNYSRPKEGDDRPFFWMFMENVVAMKVGDKRDISRFLECNPVMIDAIKV
SAAHRARYFWGNLPGMNRIFGFPVHYTDVSNMGGRGARQKLLGRWSVPPIVI
RH LFAPIKDYFACEID (SEQ ID NO: 12)
VT\textsubscript{1}TKGDPGSLEEWETVVGDDFSLYYDSYSVDERVDSKSEVEALTEQLS
EEEEEEEEEEEEEEEEEEEEEEEEEEEEEESGNQSDRSGSSGGRKACKKWRKDSP
WVKPSKRRKREPPRAKEPRGVSNDTSLETHERGFEELPLCSCRMEMAPKIDRI
SERAGHKCMATESVDGELSCNAAILKRETMRPSSRVALMVLCTHHRARM
VKHHCCPGCGYCTAGFTGELFVPHARHFKACVQLNGMVFCPHCE
DASEAQEVTIPRGDGVTTPAGTAAPAPPPLSVDQVPGRADTSQPSARMRHG
EPRRPPCDPLADTIDSSGSLTLPGNGGCLSAVGLPGPGEALEKALVIQESE
RRKLRFHPQRQLYSVKQGELQKVLMLDNLDPFQSDQSQKRTPLHAAA
QKGSVEICHVLLQAGANINAVDKQRTPLMEAVNHLEVARYMVQRGCG
VYSKEEDGASTCLHAAKIGNLEMSLLLSSTGQVDVQAQDSGGWTPIWAAE
HKHIEVIRMLLTRGADVTLTDNEENICLHWASFTGAAIAEVLNNARCDLHA
VNYHGDTPHLIAARESYHDVLPFLLSRGANPELNKEDGTAWDLTPERSDV
WFALQLNRKRLRGLGNRAIRTEKICRDVARGYENVPICVNGVDGEPED
FYIYISENCTSTMNIDRNITHLQHCTVDDCSSNLCGQLSIRCWYDKDGRL
LQEFNKIEPLIFECNQASCWRNCKNRVQVSGIKVRQLYRTAKMGWGV
RALTIPQGTFICEYVGEILSDAEADVREDSSYLFDLDNKDGEYVICDARYY
GNISRFINHLCNPNIIPVRVFMLHQDLRFPRIAFFSSRDIRGTEGELGFDYGDRF
WDIKSKYFTCGSQGSEKCHSAEAIALEQSRLALDPHELPEGLSLPPVNT
D (SEQ ID NO: 13)

[00287] dCas9VP64
[00288] MYPYPDVDPYASPKKKRRKVEASDKKYSIGLAIMTSVNGWAVIT
DEYKVPSSKFKVLGNTDRHSIKNKLIGALLDSDGTAowlerLKTARRRYTR
RKNRICYLQEFINSNEMAKVDDSSFHRLEESFLVHEDKKHERHPIFNGIVDEVA
YHEKYPTIYHLRKLVDSTKDALKDLRLIYLAHMIKFRGHFLTLEGDLNDNS
DVDKLFIQLVQTYNLQENPINASGVDAAKALSARLSKRRLRENIAIQLGE
KKNQHFGNLIALGLSLGTNFKNFSDLKEDAKLQLSDKDYDDDDNLMAQIG
DQYADLFLAACKNIDAISDILRNVETIKPALSASMKRYDEHHQDLTL
KALVRQQLPEKYEIIFFDQSNGNYAGYIDGGASQEEFYKFIKPILKMDGTE
ELLVKNREDLLRKQRTFDNSIPHIQIHGELHAILRRQEDFYPFVLKDNREK1
EKILTFRIYYVGPLARGNSRFAWMTRKSEETITPWNEEVVDKGASASQSIE
RMTNFDKNLNPNEKVLKSHLLEYFTVYNETLTKVYVTEGMKPAFLSGEQ
KKAIVDLFLKTRKVRTKQLKEDYFKKIECFDSVEISGVEDRFNASLGYTNH
LLKI+DKFLDNEENEDILEDIVLTLTLEDEMIERLKTAYHNLFDKVMK
QLKRRRYTGWRGLSRKLINGIRDQKSGKTLDFLKSDFGANRFNMQILIHDDS
LTFKEDIQKAVSQQGDLSHEHIANLAGSPAIIKKGILQTVKVDVDELVVKVMG
RHKPNENIEMARENQTTQKQKNSRERMKRIEIGIKLGSQILKEHPVENT
QLQNEKLYLYYLQNGRDYMVDQELDINRLSDYDAVPQFSLKDDSIDNKV
LTRSDKNRGKSDNPSEEVVKKMKNYWRQNLAKITQRKFDNLTKAERG
GLSLEDKAGFIKRQLVETRQITKHAQILDSRMTKDYENDKLERKVITL
KSKLVSDFRDFQFYKVREINNYHHADAYLNAVGTALIKYKPLESEFY
YGDYKVYDVRKMIaketoseqegkatakyyffysshmnfkdteilangeirk
LIEITNETGEINVWDDKGRDFATVRKVLSPMQVNYKKTEVTQGFFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLSKSVKEL
LGITIMERSSFKEPNPDLFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLAS
AGELQKGKNEALPSKYVNFYLASHYEKLLGKSGPEDEKQKQLFVEQHKHYL
DEIEIQISEFSKRVILADANLKDVLVAYNKHRDKPIREQQAENIIHFLTLNLAG
PAAFKYFDTTIDRKRYSTKEYLDATLHALQSITGLYETRDLSQLGGDSPK
RKVEASGPAGSGRADALDDDFDLMLGSDALDDDFDLMLGSDALDDDFLD
MLGSDALDDDFDLMLYID (SEQ ID NO: 14)

[00289] dCas9VP96

[00290] MYPYDVPDYASPKKKRRKVEASDDKKSIGLAIGTNSVGWAVIT
DEYKVPSKKKFVLGNTDRHSIKKLIGALLFDSEGETAEATRLKRTARRRYTR
RKNRCYLQIEFSNEMAKVDDSFFHRLEESFLVEEDKHERHPGIGNVDEVA
YHEKYPTIYHLRKLVDKSTDKADLRLIYLALAMHIKFRGHFLIEGDLNPDNS
DVKLEFIQLVQTYNLQFENPINASGVDAKSARLSKSSRENLIAQLPGE
KKNGLFGNNLIALSLGTLPNFKSNFDLAEDAKLQLSDKDYTEDDDLNLLAQIG
DQYADLFLAAKNSLDSAILDLRVNTEITKAPLSAMIKRYDEHHQDLTLL
KALVRQQLPEKYKIEFDQSKNGYAGYIDGGASQEEYFKIKPILEKMGTE
ELLVKLNREDLRLKQRTDFDNSISPHQIHLGHELHIALRQEDFPYFLKDNREKI
EKILTFRIPYVGPLARGNSRFAWMTRKSEETITPWNFEENVVDKGAQAQSIE
RTMTNFSDKLNPEKVLKHSLLYEYFTVNYELTKKVYVTEGRMKPAFLSSE
QKKAIVDFLKFTRNKTQVTVQLKEDFYKKIEFDSVEISGVEDRFNASLGTYH
DLLKIIKDDFLDNEENEDILEDIVTLLTFDREMIEERLKTYAHLFDDKVM
KQLKRRRYTGWRGLSRKLINGIRDQKSGTILDFLKSDGFMANRFNMQILIH
DLSLTFKEDIQKAQVSGGQDLSHEHIANLAGSPAIIKKGILQTVKVDVDELV
MVGRHKPENIEMARENQTTQKQNGKSREMRKRIEIGIKLGSQILKEHPVE
[00291] dCas9VP160
[00292] MYPYDVPDYASP KKKRKVEASDDKKYSIGL AIGTNSVGWAVIT
DEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSDGETAEATRLKRTARRRYTR
RKNRICYLQEFISNEAMKVDSSFFHRLEESFLVEEDKKHERHPFIGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
DVKLFIQLVQTYNQLFEENPINASGVDAKAILSRSLKSRRLENLIAQPGE
KKNLFLGNNLIALSLGTPNFKNSFDLAEDAKQLKSDTYDDDDLDNLAAQIG
DQYADLFLAAKNLSDAILSDDLVRNTEITKAPLSAMIKRYDEHHPQLTLLL
KALVRQQLPEKYKEFFDQSNSNQAYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVLKLNREDLLRRKQRTFDNSISPHQIHLGELHAIIRQQEDFYFPFLKDNRKEI
EKILTFRIPYYVGPLARGNSRFARWMTKSETTIPWNFFEEVDKGASASQFIE
RMTNFDKLNLNEKVLKPVHSLLEYFYFTVYNELTKVYVTEGMRKPAFLSAGE
QKKAIVDLLFTKTVKJQDCQKIEFSDVESIGVEDRFNASLGTYH
DLLKIJKDKDLDNEENIDILITLTLTEDREMIERLKTAYHLFDDKVM
KQLKRRYTGWRDSLKINGIDKQSGKTLIADFDSGFANRNFMQLIHD
DSLTFKEDQKAQVSGQGDSLHEHIANLAGSPAIIKGQILQQTQKVKVDSELVKV
MGHRHPENIVIEMARENQTQKQKNSRERMKRIEIGIKELGQILKEHPIE
NTQLQNEKLYLYQLQGRMDMYVDQELDINRLSDYDVAIPQSFNLKDDSID
NKVLTTRSDKRKGSDNVPSEEVVKKMKNYWRQLNKLITQRKFDNLTKA
ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVDSDFKDFQFYKVREINNYHHAHDAYLANAVGTLALKYKPKLESE
FVYGDYKVDYVRKMIK AESQ IEKAKAYYY YNSIMNFKTEITLANGEIRK
RPLIETNGETGEIVWDGRDFATVVRKVLSMMPQVNIVKKEVTQGGFSKESIL
PKRNSDKLIRAKKDWDPPKKYGGFDSTPVAYSVLVAVKEGSKKLLKSVK
ELLGITYMERSSSFENPIFDLEAKGKYKEVKKDILKLPKLYSLELENGRKRML
ASAGELQKGNELALPSYKVNYLFYLASHYEKLKGSPEDNEQKQLFVEQKH
YLDEIEEQISEFSDKRVLADANLKDVLASYKNHDKPIREQAENIIHFTLTNL
GAPAAFKYFDTTIDRKRYSTKEVLDATLIHQSITGLYETRIDLSQLGDSPK
KKRKEVSAGPGASGRADALDDFDLMLGSDALDDDFDLGSDALDDFDL
DMLGSDLDDFFDLMLGSDALDDDFDLGSDLDDDFDLGSDLDDDFDLGSDALDD
FDLMLGSDLDDDFDLMLGSDLDDDFDLGSDLDDDFDLMLGSDLDDDFDLMYID
(SEQ ID NO: 16)

[00293] dCas9hINII
[00294] MYPYDVDPYASPKKKRKVEASDDDYSGIAGLCGTVGAVIT
DEYKVPSSKFKVVLGNTRHISIKKNNLIGGALLFDSGETAAEATRLKRTAARRY
RKNRICYLQEFINSNEKAVDSSFHHLRLESFLVEEDKKHERHPFIGNIVDEVA
YHEKYPTIYHLRKKVLVDSDKALRLJYLYLALAHMIKFRGHEIDGALPNDS
VDKIFTCILNQVTVNQFPNAPASEGDAAKAILSAFLSRRLENLIAQLPE
KKNLFGNLALSLGGLTVNFKSNFDFLAEDAKLQLSREQDDDDNQLAQIG
DQYADLFLAARNLSDAISSLDRVTNTEITKAPLSASMIKARYDEHHQDLTL
KALVRQQLPKSYKEIFFDQSKKNQAGYIDGGASQFEEFKIFKIPLEKMDGTE
ELLVKLNRDELLRKRQRTFDNGSTHQLHELHAILRRQEDFYPLKDNREKI
EKILTFRIPYYVGPLARGNFSFAWMTRKSEQETTPWFEEVVDKGAASQFIE
RMNTFNDKLNPENKVLPHKLLEYESFTYVNELTKKVYVTEGMPAFLSGE
QKKAIVDLDLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRFNALSHTY
DLLKIJKDFLDNEENEDILEDIVTLTLFFEDREMIERLKYAWLFFDKV
QKLKKRRYTGWRSLRKLINGIRDKQSGKTLDFLKDGSFAFRNNMQLIHID
DSLTFKEDIQDKAQCYSQQGDSLHEIANLAGSPAIIKGILQTVKVVDLKV
MGRHKPENIVIEMARENQITQKGQKNSREMKRIEEGIKELGSQLKEHPVE
NQLQNEKLYLYLPQGRMDVMQDELDINRSLDYDVAIPQSLFDSDS
NKVLTRSDKNRGKSDNVPSVEEVVKKMKNYWRQLNKLITQRKFDNLTCA
ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVEINNYHAHDAYLNAVVGTALIKKYPKLESE
FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNNFKTEITLANGEIRK
RPLIETNGETGEIVDWKDGRDFATVRKLVSMQPVNIVKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWPDKYGGFSDPTAYSLVVAKEVKGSKKLKSVK
ELLGITIMERSSSFENPFPDFELEAKGYEVKKDLIILPKYSLFELENGKRML
ASAGELQKGNELALSYYNVALSHYEKLKGSPEDNEQKQLFVEQHWH
YLDELIEEQISEFSKRVILADANLKVSLAAYNKHRDKPIREQAENIIHLFTLNL
GAPAAKFYFDTTIDKRYSTKVELDATLAHQISGTYETRIEDLSQLGGDSPK
KKRVEASGPAMMMALSSTFQPKVFQLEDDGEFYMIGEVGNYLRM
FRGSLYKRYPSLWRLATVEERKKIVASSHHGKTKPNTKHGTYDLLTATS
LLKASEVEEILDGNDEYKAVSISTEPPTYLREQKAKRNSQWVPTLPNSSSH
LDAVPCSTTINRNMRGRDKKCRTFPLCFDDHDPAVHESAQPEVLVPILDM
EIDGQKLRDAFTWNNMNEKLMTPLMFSIEILCDDDLNLPTFVPAISAIRQIQIE
SYPYSDLEDQSDQRVIIKLNIHVGNISLVDFQFEWDMSEKENSPEKFALKCS
ELGLGGEVFTTAYSIRGQSLSHQKYAFSENPLPTVEAIRNTGADQWCPL
LLETLTDAEMKIKRDQRNRTRMRRLANTAPA (SEQ ID NO: 17)

[00295]  dCas9hMBD4
[00296]  MYPYDVPDYASPKKKRKVEASDKKKYSIGLAIAGTNVSVGWAVIT
DEYKVSFFKVLGNTDRHSIKKLNSGALLFDSGETAEATRLKRTARRRYTR
RKNRICYLQIEFSNEMAKVDDSFFHRLIESFLVEEDKHERHIPFGNVIDEVA
YHEKYPTIYHRLKVIKVDSTDKALRLIYLALAHMIKRHGHLIEGDLNPDNS
DVDKLFIQQLVQTYNQLFEENPINASGVDAKALSARLSSRRLENLIAQLPGE
KKNGLFGNLIALSGLTPNFKSNFDLAEADKQLSKDLYDLDLNNQIQ
DQYADLFLAAKNSLDAIQILSDDLVRNTEITKAPSAMKRYDEHHQDLTLL
KALVQRQLPEKYKEIFDFDSQKNGYAGYIDGQASQEEFYKIFKIPLEMDGTE
ELLVALKRDLLRKQRTDFNSGIPHIHLGHELHAILRRQDFYPFLKDRNEKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPPWNEVEVVDKGASAQSFIE
RMNTDKNLNPNEKVLPHSSLLYEYFTVYNELTKVYVTEGRMRRPFAFEGE
QKKAIVDLFFKTNKTVTVQKLEDYFKKIECFDSVEISGVEDRFNASLGTYH
DLLKIIKDSDKLFNEENIDLEDIVITLTLFDREMIERLKTAYAHLDKKVM
KQLKRRRTGWRGSLRKLINGIRDKQSGKTIDFLKSDGFANRNFQMQLIH
DSLTFKEDIQKQAQVSGQGDSLHHEIANLAGSPAIKKGILQTAKVVDDELVKV
MGRHKPENIVIEMARENQTQKQGKNSRERMKRIEGIKEKELGSQIKAHPYE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSID
NKVLTRSDKRNGSDNPSEEVVVKKMKNYWRQLNNAKLITQRFDNLTKA
ERGGLELDKAGFIRQLVETRQITKHVAQILDSRMNTKYNEDKLIREVKV
ITLKSKLVSDFRDFKFQYKREINNYHHADLAYNNAVGTALIKKYPKLESE
FVGYGYKVYDVRMIakiASEQEIGKAKTYFFYSNMNNFKTETLTANGEIRK
RLIEHTNGETGEIVWDGKRDFATVRKLMPQVNNVKKTEVQITGGFSKESIL
PKRNDSKLIARKKDWDPPKYYGGFDSTPVAYSVVLVAKVEKGSKKLKSVK
ELLGITIMERSSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYFNVLASHYEYELKKGSPDENQKQLFVEQVHKH
YLDEIIIEQISEFSKRVILADANKLVSAYNKHRDKPIREQAENIIHFLTLNL
GAPAAFYFDTITDRKRYTSTKEVLDATLIHQTSITGLYETRIDLSQLGGDSPK
KKRKVEASGPAMGTTGLESLSLGDGRGAAPTVTSERLVPPNPDLRKEVIDVA
MELERVGEDEEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVP CGWE
RVTQKQRLFGKTAGRFVDVFYISPQGLKFRSKSSLANYLHKNGETSLKPEDFDF
VRLSKRGIKSYKDCSMAALTSHLNQSNNSNWNLTRSKCKKDVMPRESS
SELQESRGLSNFTSTHLKKEDEGVVDDVNFRKRPGKVTILKIGPIKKT
KGCRKSCSGFVQPSDKRESVNCNADAESEPVAQKSQLDRTVCIISDAGACGE
TLSVTSEENLSVKKERSSLSSGSNFSCSEQKTSGIINKFCSAKDSEHNKYEDET
FLESEEIGTKVEVERKEHHTLDTILKRGSEMNDNCSPRTKDFGTGEKIFQEDTI
PRTQIERKRTSLYFSSKYNKEALSPPRRAFKKWTPPRSPFNLVQTELHDP
WKLLIATIFLNRTSGKAIPVWLKFLEKYPASEVARTADWRDVSELLKPLG
LYDLRAKTVKSDEYLTQKWKypiELHGICKYGNDSYRFVNEWKQVHP
EDHKLKNKYHDLWLENHEKL S (SEQ ID NO: 18)

[00297] dCas9hTDG

[00298] MYPYDVPDYASPDDKKRKVEASDDKKYSIGLAI GSTNVGWARIT
DEYKVKPSKKFKVLGNTDRHISIKKNLIGALLFDGS TAAE ATRKLKRTARRYTR
RKNRICYLQEFISNEMAKVDSFFHRLEESFLVEEDKKHERHPFIGNIVEDVA
YHEKYPTIYHLRKLVDSTDKADLRLIYLALAHMIKRGRHFLIEGDLNPDNS
DVKDLFIIQLVQTYNQLFEENPINASVGDAKAILSA RLKSRRLENLIAQLPG
EKKNLGFLNALSLGLTNPKNSFDLAEADKLQLSKDTYDDDDLDNLLAQIG
DQYADLFLAANKLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLL
KALVRQQLPEKYKEIFFDQSKNYAGYAGIDGGASQEEFYKFIPKILEMDGTE
ELLVKLNRDDLRKQRTFDNSIPHQIHGLHELHAILRRQEDYFPLKDNREKI
dCas9hTETICD

MYPYDVPDYASPKKKRKVEASDKKYSIGLAI GNTNVSGWAVIT
DEYKVPSKKFKVLGNTRHISIKKNLIGALLFDA GTEAEATRLKRTARRRYTR
RKNRICYLQEIQNEMAKVDDSFFHRLEESFLVEN DKKHERPHGNIYVEVA
YEHEKPYTIHRLKVLVDSTDADLRLLTIYLALAHMIKFRGHLIEGDNLPDNS
DVKDLFIQLVQTYNQLFEENPASGVDAKAIMALSRSKSLRRENMLAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSAILSDILRVNTEITKAPLSASMKRYDEHHQDLTLL
KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNRDLLLKRQRTFDNGSPHQLHGEHLHRLRQEDFYPFLKDRNEKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQAQSIE
RMTNFDKNLPNKEVLPHSLLEYYFTVYNELTKVKYVTGEMRKPAPFLSAGE
QKKAIVDLLFKTNRKVTVKQLEDYFKKIECFDSVEISGVEDRFNASLGYTYH
DLLKIIKDKFILDNEENEIDILEDIVLTLTLEDREMIEERLKTYAHLFDKVM
KQLKRRRTGTWGRSLKRLINGIRDQSKGTILTDLKSDGFANFMRMQLHID
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQQKGNRSERMKRIEEGIKELGSQILKEHPVE
NTQLQNEKLYLYLQNGRDMDYVDQELDINRLSDYDVDAIVPQSLKDDSID
NKVLTRSDKGNRSNDVPSEEEVKKMKNYWRQLLNAKLITQKFDNLTKA
ERGLLSELDKAGFIKRQVLTRQITKHAQILDSRMNTKYENDKLIievKv
ITLKSKLVSDFRKFDFQFYKVREINNYHHAHAYLNAVGTALIKKYPKLESE
FVYGDVYKVYDVRKMIAKSEQIEIGKAKAYYFFYSNIMNFKTEITLANGEIRK
RPLIETNGETEIVWEDKGDFATVRKVLSPMQVNIVKKEVTQGIFGSKESIL
PKRNSDKLIARKKDWDPPKYGGFSOFTVAYSVLVLVAVKVEGKSKLKSVK
ELLGITIMERSSFKNDPIFDLEAKGYKEVKKDILKLPYSFLJELEGRKRML
ASAGELQKGNELALPSKYYNFLAYSHYEKLKGSPEDNEQKQLFVEQVQKHH
YLDEIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLNL
GAPAASKYFDDTDIRDKRYSKEVLDATLHQSITGLYETRIDLSQLLGGDSPK
KKRKVEASGPAELPTSCLDRVQIDKGPYYTHLLAGPSVAAIVREIMENRY
QGKGNARIEIVYVGYKEGKSSHGCPIAKWVLRSSDEEKVLCLVRQRTGHH
CPTAVMVLVMVWDGIPLMADMRLritteNKLNSYNGHPTDRCTLNENRT
CTCQGIDPETGCASFSFGCSWSMYFCGKFGSPSRLFRIDPSSPLHEKMLE
DNLQLSLATRLAPIYQKYAPVAYQNQVEYENVARECLGSKEGRPSGVTAC
LDFCAHPHRDIHMNNSTVVTCTLTDNRSGLVIPQDEQLHVLPYKLSD
TDEFGSKEGMEAIKSGAEVLAPRKKRTCTQPVPRSGKAKRAAMMTEVL
AHKIRAveKPKPRKKNSTTTINNSKPSLPTLGSNTETVQPEVKSETEPHF
ILKSSDNTKTYSLMPASAPHPVEASPGFWSMPTASATAPLKNLSDACGF
SERSSTPHCTMPSGRLSGANAAADGPGISQLGEVAPLTLSPAVMEPLINSE
PSTGVTEPLTPHQPNHQSFLTSPQDSLAPMEDEQHSEADEPPSDEPLSDD
PLSPAEEKLPHIDEYWSDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATT
VEHPNRRNHPTRLSLVFYQHKNLNPKQHFELNKIKFEAKEAKNNKMKASE
QKDQAANEGPEQSSEVNELNQIPSHEYKALTHTHDNVVTVSPYALTHVAGYPN
HWVID (SEQ ID NO: 20)

[00301] dCas9hPRMTI

[00302] MYPYDVPDAYSPPKKRKVEASDKKYSIGLAIGTNSVGWAVITDE
YKVPSKKFKVLGNTDHRHSIIKKNLIGALLFDGSGETAEATRLKTARRRYTRR
KRCYLQEIFSNEMAVKDDSFFHRLEESFLVEEDKKHERHPFIGNVDEVAYH
EKYPITYHLRKKLVDSTDKADRLRIYLALAHMIKFRGHLIEGDNLNDVD
DKLFIQLVQVTNLFEENPASGVDAKAILSLKSRRLENLIAQLPGEC
NGLFGLIALSLGLTPNFKNFDLAEDAKQLSKDYDDDLNNLAAQIDGQ
YADLFLAAKNLSDAISLLDLVRNTEITEKAPSAMIRYDEHHQDLTLKALK
VRQQLPEKEYKEIFFDQSKNGYAGYIDGGAQCEEYKFIIKPILEKMDGTTEELL
VKNREDLLRKRQTFDNGSIPHQIHLGELHAILRRQEDFYFLKDNRKEIEK
LTFRIPYYVGPLARGNSRFAWMTRKSEITIPFNSFEEVVDKGAASQFIERT
TNFDKNLNEKVPKHLLYEYFTVYNETLKVKYVGMRKPAFLSGEQK
AIVDLLFKTNRKVTVKQLKEDYFFKIEFCDFSVEISGVEDRFNASLGTYHDLL
KIIKDKFLDNEENEDILEDIVTLTLFDREMIEERLKTYAHIFDDKVKKQL
KRRRTGWRGRLSRKLINGIRDKSQGKTIDFLLKDSGFANRNFMLQIHDDSLT
FKEDIQKAAVSQGSQGDSLHEHIANLAGSPAIIKGLQTJKVVDLVKVMGHR
KPNENIVIEMARENQTTQKGGKSRSREMKRRIIEGIGELGSQILKEHPVENTQL
QNEKLYLYQNNGRDMYQVQELDINRLSVDYDIAVPQSFGLKSDSNKVL
TRSDKNRGKSDNVPSEEVKKMKNYWRQIILNKLITQRFQDNFTKAERG
LSELDKAGFIKRQLVETVRITKHAQVILDSSRMNTKYDENDKLIREVKVITLK
SKLVSDFRKDFQFYKVEHNNYHHADAYLNAVGTALIKKYKLESEFYV
GDYKVYDVKRMIKASEQIEIGKATAKYFFYNSIMNFTEITLANEIRKPLI
ETNGETGEIIVWDKGRDFATVRKVLMPQVNIKTEVQTGGSFESILPKR
NSDKLIAARKKDWDPKKYGGFSDPSVTAYSVLVLVKAKEGKSKKLKSVELL
GITIMERSSEFKNPIDFLEAAGKYEKVKKDLIKLPKYSLEEFNGRKRMLAS
GELQKGNELALPSKYNVNLYASHYEKLGSPEDNEQKQFLVEQKHYPYLDI
IEEQISEFSKRPILADANLKLVSAYNHKRDKPIREQAENIHLFHTLTLNGAPA
AFKYFDTTIDRKRSTKEVLDAATLHQSITGLYETRIDLSQGGDSPSSKRK
VEASGPAMAAAEAAANCIMENFVATLAMGMSLQPPLVEVSCQGAESSEKPN
EDMTSKDYYFDSYAHFGIHEEMLKDEVRTLTYRNSMFHNRHLFKDKVVLD
VGSGTGILCMFAAKAGARKVIGIECSSISDYAVKIVKANKLDHVVTIIGKGV
EEVELPVEKVDIIIIEWMGYCLFYESMLNTVLYARDKWLAPDGLIFPDRTL
YVTAIEDRQYKDYIKHWWENVYGFDMSCIKDVAIKEPLVDVVPKQLVTN
ACLKEVIDITYVKVEDLFTSPFCLQVKRNDYVHALVAYFNIETFRCRKHRTG
FSTSPESPYTHWKQTVFYMEDELTVKTGEIFGTIGMRPNAKNRRDLDFCID
LDKFGQLCELSCSTDYRMRID (SEQ ID NO: 21)

[00303] dCas9hSET7

[00304] MYPYDVPYASP99KRRKVEASDKKYSIGLAIGTNSVGWAVITDE
YKVPSSKFKVLGNTDRHISIKKNLIGALLFDGETAEATRLKRTARRRYTRRR
NRICYLQEIFSNEMAKVVDSFFHRLESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTYHLKRLKDSTDKDCLAIRLILAHAMIKFRGHFLIEGLDNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLKSRRLENLIAQLPGKK
NGLFGNIALSLGLTPNFKNFSDLAEADKLQKSKDYDDDLNDNLAQIQDQ
YADLFLAAAKNLSDAADLDRLRVENTITEKAPLASSMISKRYDEHHQDLDLKL
VRQQLPEKYKKEFDQSKNGAYAGIDGASQEEFYFKFPIKILEKMDGTEELL
VKLNRDULLKQRTFDNGSIHPQHIGHLGELHAHRRQEDFYFPKLKDNEKIEK
LTFRIPYVGPLARGNSRFKAMTRKSETITPNFEEVVDKGASAQSFIERM
TNFDKLNLPEKVLPKHSSLVEYFTYNELTKVYVTENMKPAPFLSLGEQKK
AIVDLNLKTNKVTKVLKDYFKKIECFDSVEEQGDVFQFNASLGODYHDLL
KIIKDKDLDNEENILEDIVLTLTLEDREMIERLKTYAHLFDDKVMQKL
KRRRYTWGWRLSRRKLIRGDKSEQGTILDQIDSGFANRNFQMLIHDDSLLT
FKEDIQKAVQSGQDSLHEHIANLAGSPAIIKQILQTVKVVDLKVKGMHR
KENIVIEMARENQTQFGQKSVRVRMRKSPHRKIIDIEKILGQILKEHVPENTQL
QNEKLYYLYQNGRMDYVQDSDINLSDYDIVDAPRSFQFLDDSIDNKVL
TRSDKNRGSNDVPSEEVKMMKYVRQLNNALITQRKFDNLKTAERGG
LSHLDKAFIKRQVLTRQIKTHVAQILDSRMNKTYDENDKLRVEKVITLKL
SKLVSDFRKDFQFYKVRINNYHAHHDAYLNAVGTALIKYKPKLESEFYVY
GDYKQYVDRKMIASKESEQIEGKATAKYFFYSINMNFFKETILANGEIRKRIPL
ETNGETGEIVWDKGRDFATVRKVLSPMQVNIVKDTDQTVGTGFSKESILKPR
NSDKLIARKKDDWDPKKYGGFDSTPVAYSVLVVAKVEKGRSKSKLSVKELL
GITIMERSSFEKNNPIDFLEAKGYKEVKKDLJIKLIPKSYSLFELNKRKMRMLASA
GELQKGNELALPSKYVNFYLASHYEKLKSGPDEQKQLFVEQKHYLDE
IIEQISEFSKRVLADANEKLSAYNKHRDKPIREQAENIIHLFTLTLNLGAPA
AFKYFDTTTDRKRYSTKEVLDATLHQSITGLYETRIDLQLGGDSPKKKRK
VEASGPAMDSDDDEEMVEAAVEGHLLDDGLPHGFECTVYSTSTDRFGNFVHG
EKNGRGGFKFFFDGSTLEGYYVDDALQGQGVYTEDGGVQLQTYVDGEIGN
PAQEYDGDGRLFKGQYKDNNHRGVCIWWYYYPDGSLSVGVEVNEDGMTEGI
AYYVDPERTALYKGIDGEMIEKLSTMLSTEGRPHFLMPGNSVYHFDK
STSSCINTNAALPPDYPESERYVVAESLISAGLEGLFSKVAVGPTVMSFYNGV
RITHQEVDSDWALKNGTSLDEETVIDVPEYNVHVKASCALGHKANHSF
TPNCYDMFVHPFPGPKICRTLRAVEADEELTVAYGDHSPGPAGPEAPE
WYQVVELKAFEQATQKID (SEQ ID NO: 22)

[00305] dCas9SID4x
[00306] MYPYDVPDYASPKKRKVEASDSDKYSIGLAIGTNSVGVAVITDE
YKVPSSKFKVLGNTDHRHIIKNNLGALLDSGETAATRLKRTARRRYTRRK
NRCYLQEIIFSNEMAKVDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEAYH
EKYPTIYHLRKLVDSTKDADLRPAQNALAHMIKFRGHLIEGDGLNDNPSDV
DKLFQVLQVQYQNLFEENPASGVDAKAILSARLSKRNENLIAQLPGEKK
NGLFGNLIALSLGLTPNFSNFDAEDAKLQLSKDYDDDDLNLQAQIDQQ
YADLFLAAKNLSDLAILSDLRVNEITKAPLSAMKRYDEHHQDLTLKCAL
VRQQLPEKYEIFFDQSKNGYAGYIDGGAQEEFYFKIFKPILEKMDGTEELL
VKNREDLLRKRQTFDNGSIPHQLHELHAILRRQEDFYPFFLKDNREKIEKI
LTFRIPYYVGPLARGNSRFOWMTRKSEETITPNFEEVVDKGASAQSFIERM
TNFDKNLPNEKVLKPHSLLEYEYFTYVEYNTKVYTVETGRMKPAPFLSGEQKK
AIVDLKFNTKRVKTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLL
KIJKDKDFLDNEEDEDILEDIVLTTLTFEDREMIEERLKTAYHFLDDKVMKQL
KRRTYTGWGRSLRKLINGRDQKSGKTLDFLSDGFGANRNFMQMLHDDSULT
FKEDDIQKAQVSGQGDSLHEHIANLGSAPAIKKGILQTKVVDVELVKVMRH
KPNENIEMARENQTTQKQGKNSREMKRIEEGIEKLSQILKEPHVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIFPQSFLKDSIDKNV
LTRSDKNRGSNDNPSEEVVKKMKNYWRQLLNAKLITQRFKDNLTCAERGG
LSELDKAGFIKRQLVETRQITKHVQILDSRMNKTYENDKLIKEVKITLK
SKLVSDFRKFQFYKVPVREINNYHAHDAVYLANVGTALIKKYPKLESEFYV
GDYKVDYDVRKMIAKSEQEIGKATAKYFFYSNINNFKTITEILTANGEIRKPLI
ETNGETGEIVWDKGDRFATVRKLSPQVNIVKKEVQFTGGFSKESILPKR
dCas9hsSssIM

[00307] MYPYDVPDYASPKKKRKVEASDKKYSIGLAIGTN SVGWAVITDE

[00308] YKVPSKKFKVLGNTDRHISIKNLIGALLFDSGETAETRLKRTARRRYYTRRK

RIDID (SEQ ID NO: 23)

[00307] dCas9hsSssIM

[00308] MYPYDVPDYASPKKKRKVEASDKKYSIGLAIGTN SVGWAVITDE

[00307] dCas9hsSssIM

[00308] MYPYDVPDYASPKKKRKVEASDKKYSIGLAIGTN SVGWAVITDE
[00309] dCas9hsSssIMA (split SssIM fragment A fusion)

[00310] MYPYDVDPDYNASPKKKRKVEASDKKKYSIGLGAIGNTSNSVWAVITDE

KYVPSSKKFKVLGNTRSHIKKLNLLGALDFSGETEAATLKRTRARRYYTRR

NRICYLQIEFNSEMAKVDSSFFHRLEESFLVEEDKKHERHPFIGNIVDEVAYH

EKYPTIYHLRKKLDSTKDADLRLYIYALAMHIKFRGHLIEGDLNPDNSDV

DKLFQILQVQTYNLQFEEPINASGVDAKALSAKLSPRLELIAQLPGEK

NGLFNGNIALSLGLTPFNSDFLDAEDAKLQLSKDYYDDDLDNLLAQIGDQ

YADLFLAAKNSDAILLSDILRVNTEITKAPLSAMIRKDEHHQDLTLKLAL

VRQQQLPEKYKEIFDFQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL

VKLNREDLKRRQTFDNGSIPHQHLGHELHRRQQDFYFPLKDNREKIEKI

LTFRIPYVVGPLARGNSRFRAWMTRKSEITIPWNEEVVDKGAQSFIERM

TNFDKNLPEKVLKSHLLYEYFTVNYELTKVKYVTEMKRPAFLSEQKK

AIVDLDLFKTNRKVTQKQLKEDYFKKIECFSEISGFEDFNASGLGTYHDLL

KIIDKDFLDNEENEDILEDIVLTTLFEDREMIEERLKTYAHLFDDVKMKQL

KRRRYTGWGRSLRKLINGIRDQKQSGKTIDFGLKSDGFAANFNMQLIHDSSLT

FKEDIQQAQVGSQGDSLHEHIANLSAgPAIKKLGQITVKEVDELVKVMGRH

KPENIVIEMARENQTTQKPQKNSRERMKRIEIGIKELGSLQIHEPVENTQL

QNEKLYLYQLNGRDVMYVDQELDINRLSVDYDAIVPOLSDKDDSNKVL

TRSDKNRGKSDNVPSSEEVKKMKNYWQQLNAKLITQRFKDNLTKAERGG

LSELDKAGFIKRQLVETRQITKHVAQILDSMNTKYEYDNEDKLIREVKVIITL

SKLVSDFRKDFQFYKVEINNYYHHAHDAYLNAVGTALIKKYKPLESEFVY
dCas9hsSssIMB (split SssIM fragment B fusion)
[00313] Example 3

[00314] One step generation of mice carrying reporter and conditional allele by CRISPR/Cas mediated genome editing

[00315] Here, reporter and conditional mutant mice were created by co-injection of zygotes with Cas9 mRNA, different guide RNAs (sgRNAs) as well as DNA vectors of different sizes. Using this one step procedure, mice carrying a tag or a fluorescent reporter construct in the Nanog, the Sox2 and the Oct4 gene as well as Mecp2 conditional mutant mice were generated. In addition, using sgRNAs targeting two separate sites in the Mecp2 gene, mice harboring the predicted deletions of about 700 bps were produced. Finally, potential off-targets of four sgRNAs in gene-modified mice and ESC lines were analyzed and off-target mutations were identified in only rare instances indicating high specificity of genome editing by the CRISPR/Cas system.

[00316] EXPERIMENTAL PROCEDURES

[00317] Production of Cas9 mRNA and sgRNA

[00318] Bicistronic expression vector px330 expressing Cas9 and sgRNA (Cong, L., et al., Science 339, 819-823 (2013)) was digested with Bbsl and treated with Antarctic Phosphatase, and the linearized vector was gel-purified. A pair of oligos
Table 1) for each targeting site was annealed, phosphorylated, and ligated to the linearized vector.

**[00319]** T7 promoter was added to Cas9 coding region by PCR amplification using primer Cas9 F and R (Table 11). T7-Cas9 PCR product was gel-purified and used as the template for in vitro transcription (IVT) using mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). T7 promoter was added to sgRNAs template by PCR amplification using primer listed in Table 11. The T7-sgRNA PCR product was gel-purified and used as the template for IVT using MEGAshortscript T7 kit (Life Technologies). Both the Cas9 mRNA and the sgRNAs were purified using MEGAclear kit (Life Technologies) and eluted in RNase-free water.

**[00320]** Single stranded and double stranded DNA donors

**[00321]** All single stranded oligos were ordered as Ulramer DNA oligos from Integrated DNA Technologies. Nanog-2A-mCherry vector was modified from previously published targeting vector Nanog-2A-mCherry-PGK-Neo (Faddah et al, 2013). Nanog-2A-mCherry-PGK-Neo was digested with Pad and Ascl to drop out the PGK-Neo cassette, the 9.7kb fragment was gel purified and blunt-ended using T4 DNA polymerase (New England Biolabs), then self-ligated using T4 DNA ligase (New England Biolabs). Oct4-IRES-eGFP-PGK-Neo vector is previously published (Lengner et al, 2007).

**[00322]** Suveryor assay and RFLP analysis for genome modification

**[00323]** Suveryor assay was performed as described (Guschin, D.Y., et al., Methods Mol Biol 649, 247-256 (2010)). Genomic DNA from targeted and control mice or embryos was extracted and PCR was performed using gene specific primers (Table 11) under the following conditions: 95°C for 5 min; 35x (95°C for 30 s, 60°C for 30 s, 68°C for 40 s); 68°C for 2 min; hold at 4°C. PCR products were then denatured, annealed, and treated with Suveryor nuclease (Transgenomic). DNA concentration of each band was measured on an ethidium bromide-stained 10% acrylamide Criterion TBE gel (BioRad) and quantified using Image I software. For RFLP analysis, IQμI of Tetl, Tet2, Mecp2-Rl, R2 PCR product was digested with EcoRI, IQμI of Mecp2-Ll, L2 PCR product was digested with Nhel. Digested DNA was separated on an ethidium bromide-stained agarose gel (2%). For sequencing,
PCR products were cloned using the Original TA Cloning Kit (Invitrogen), and mutations were identified by Sanger sequencing.

[00324] **One cell embryo injection**

[00325] All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at MIT. B6D2F1 (C57BL/6 X DBA2) female mice and ICR mouse strains were used as embryo donors and foster mothers, respectively. Super-ovulated female B6D2F1 mice (7-8 weeks old) were mated to B6D2F1 stud males, and fertilized embryos were collected from oviducts. Cas mRNA (100 ng/µl), sgRNA (50%/ µl) and donor oligos (100 ng/µl) were mixed and injected into the cytoplasm of fertilized eggs with well-recognized pronuclei in M2 medium (Sigma). The injected zygotes were cultured in KSOM with amino acids at 37°C under 5% C0₂ in air until blastocyst stage by 3.5 days. Thereafter, 15-25 blastocysts were transferred into uterus of pseudopregnant ICR females at 2.5 dpc.

[00326] **Southern blotting**

[00327] Genomic DNA was separated on a 0.8% agarose gel after restriction digests with the appropriate enzymes, transferred to a nylon membrane (Amersham) and hybridized with ³²P random primer (Stratagene)-labeled probes. Between hybridizations, blots were stripped and checked for complete removal of radioactivity before rehybridization with a different probe.

[00328] **In vivo Cre recombination**

[00329] A 20-µl reaction containing 1 µg of genomic DNA and 10 units of recombinant Cre recombinase (New England Biolabs) in 1 x buffer was incubated at 37 °C for one hour. For all targets, 1 µl of the Cre reaction mix was used as template for PCR reactions with gene-specific primers. For each target, primers DF and DR were used for detecting the deletion products, and primers CF and CR were used to detect the circle product. All products were sequenced.

[00330] **Immunostaining and Western blot analysis**

[00331] For immunostaining, cells in 24-well were fixed in PBS supplemented with 4% paraformaldehyde for 15 min at room temperature (RT). The cells were then permeabilized using 0.2% Triton X-100 in PBS for 15 min at RT. The cells were blocked for 30 min in 1% BSA in PBS. Primary antibody against V5 (ab9137, abeam) was diluted in the same blocking buffer and incubated with the samples overnight at 4 °C. The cells were treated with a fluorescently coupled secondary
antibody and then incubated for 1 hr at RT. The nuclei were stained with Hoechst 33342 (Sigma) for 5 min at RT.

For western blot, Cell pellets were lysed on ice in Laemmli buffer (62.5 mM Tris-HCl, pH 6.8, 2% sodium dodecyl sulfate, 5% b-mercaptoethanol, 10% glycerol and 0.01% bromophenol blue) for 30 min in presence of protease inhibitors (Roche Diagnostics), boiled for 5-7 min at 100 °C, and subjected to western blot analysis. Primary antibodies: V5 (1:1,000, ab9137, abeam), beta-actin (1:2,000). Blots were probed with anti-goat, or anti-rabbit IgG-HRP secondary antibody (1:10,000) and visualized using ECL detection kit (GE Healthcare).

ESC derivation and differentiation

Morulas or blastocysts were selected to generate ES cell lines. The zona pellucida was removed using acid Tyrode solution. Each embryo was transferred into one well of a 96-well plate seeded with ICR embryonic fibroblast feeders in ESC medium supplemented with 20% knockout serum replacement, 1,500 U/ml leukemia inhibitory factor (LIF), 3 M CHIR99021, and 1 M PD0325901. After 4-5 days in culture, the colonies were trypsinized and transferred to a 96-well plate with a fresh feeder layer in fresh medium. Clonal expansion of the ESCs proceeded from 48-well plates to 6-well plates with feeder cells and then to 6-well plates for routine culture.

For ESC differentiation, cells were harvested by trypsinization and transferred to bacterial culture dishes in the ES medium without or LIF. After 3 days, aggregated cells were plated onto gelatin-coated tissue culture dishes and incubated for another 3 days.

Prediction of potential off-targets

Potential off-targets were predicted by searching the mouse genome (mm9) for matches to the 20-nt sgRNA sequence allowing for up to 4 mismatches (Nanog) or 3 mismatches (Sox2, Mecp2-L2 and Mecp2-RI) followed by NGG PAM sequence. Matches were ranked first by ascending number of mismatches, then by ascending distance from the PAM sequence.

Results

Targeted insertion of short DNA fragments

As described herein, precise introduction of base pair mutations into the Tet1 and Tet2 genes was done through homology directed repair (HDR)-mediated
genome editing following co-injection of single stranded mutant DNA oligos, sgRNAs and Cas9 mRNA (Wang, H., et al., Cell 153, 910-918 (2013)). To test whether a larger DNA construct could be inserted at the same DSBs at Tetl exon 4 and Tet2 exon 3, oligos were designed containing the 34bp loxP site and a 6bp EcoRI site flanked by 60bps sequences adjoining the DSBs (Fig. 19A). Cas9 mRNA was co-injected with sgRNAs and single stranded DNA oligos targeting both Tetl and Tet2 into zygotes. The restriction fragment length polymorphism (RFLP) assay shown in Fig. 19B identified six out of 15 tested embryos carrying the loxP site at the Tetl locus, eight carrying the loxP site at the Tet2 locus and three had at least one allele of each gene correctly modified. The correct integration of loxP sites was confirmed by sequencing PCR products of sample 2, 9 and 14 (Fig. 19C) These results demonstrate that HDR-mediated repair can introduce targeted integration of 40bp DNA elements efficiently through CRISPR/Cas mediated genome editing (summarized in Table 7).

[00341] Mice with reporters in the endogenous Nanog, Sox2 and Oct4 genes

[00342] Since the study of many genes and their protein products are limited by the availability of high quality antibodies, the potential of fusing a short epitope tag to an endogenous gene was explored. sgRNA was designed to target the stop codon of Sox2 and a corresponding oligo to fuse the 42bp V5 tag into the last codon (Fig. 16A). After injection of the sgRNA, Cas9 mRNA and the oligo into zygotes, in vitro differentiated blastocysts were explanted into culture to derive ES cells. PCR genotyping and sequencing identified seven out of 16 ES cell lines carrying a correctly targeted insert (Fig. 16B and 16C). Western blot analysis revealed a protein band at the predicted size using V5 antibody in targeted ES cells but not in the control cells (Fig. 16D). As expected from a correctly targeted and functional allele, Sox2 expression was seen in targeted blastocysts and ES cells using V5 antibody (Fig. 16E and 16F). 12 of 35 E 13.5 (Embryonic Day 13) embryos and live bom mice derived from injected zygotes carried the V5 tag correctly targeted into the Sox2 gene as indicated by PCR genotyping and sequencing (Table 7).

[00343] To assess whether a marker transgene could be inserted into an endogenous locus, Cas9 mRNA, sgRNA and a double stranded donor vector which was designed to fuse a p2A-mCherry reporter with the last codon of the Nanog gene were co-injected (Fig. 17A). A circular donor vector was used to minimize random
integrations. To assess toxicity and to optimize the concentration of donor DNA, different amounts of Nanog-2A-mCherry vector were microinjected. Injection with high a concentration of donor DNA (500 ng/µl) yielded mCherry-positive embryos with high efficiency with most blastocysts being retarded, whereas injection with a lower donor DNA concentration (10 ng/µl) yielded mostly healthy blastocysts most of which were mCherry-negative. When 200 ng/µl donor DNA was used, 75% of the injected zygotes developed to blastocysts, 9% of which were mCherry-positive (Figure 17C, Table 9). mCherry was mainly expressed in the inner cell mass (ICM) consistent with targeted integration of the mCherry transgene into the Nanog gene. Six ES cell lines were derived from mCherry positive blastocysts, four of which uniformly expressed mCherry with the signal disappearing upon cellular differentiation (Fig. 17C). The other two lines showed variegated mCherry expression with some colonies being mCherry positive and others negative (Fig 20A) consistent with mosaic donor embryos, which would be expected if transgene insertion occurred at a later than the zygote stage as has been previously observed with ZNF and TALEN - mediated targeting (Cui, X., et al., Nat Biotechnol 29, 64-67 (2011); Wefers, B., et al., Proc Natl Acad Sci USA 110, 3782-3787 (2013); Brown, A.J., et al. Nat Methods (2013)). Correct transgene integration in ES cell lines was confirmed by Southern blot analysis (Figure 17B). Postnatal mice from injected zygotes were also generated. Southern blot analysis (Fig. 20B and 20C) revealed that eight out of 86 E13.5 embryos and live born mice carried the mCherry transgene in the Nanog locus. One targeted mouse was mosaic, since the intensity of targeted allele was lower than the wild type allele (Fig 20B, #6). Two of the mice carried an additional randomly integrated transgene (Fig. 20C, #3). As summarized in Tables 7 and 9, the efficiency of targeted insertion of the transgene was about 10% in blastocysts and postnatal mice derived from injected zygotes.

Finally, sgRNA targeting the Oct 4' UTR was designed, which was co-injected with a published donor vector designed to integrate the 3kb transgene cassette (IRES-eGFP-loxP-Neo-loxP; Figure 17D) at the 3' end of the Oct4 gene (Lengner, C.J., et al. Cell Stem Cell 1, 403-415 (2007)). Blastocysts were derived from injected zygotes, inspected for GFP expression and explanted to derive ES cells. About 20% of the blastocysts displayed uniform GFP expression in the inner cell mass (ICM) region. Three of nine derived ES cell lines expressed GFP (Figure
17E), including one showed mosaic expression (Table 10). Three out of ten live
born mice contained the targeted allele (Table 7). Correct targeting in mice and ES
cell lines was confirmed by Southern blot analysis (Figure 17F). Conventionally,
transgenic mice are generated by pronuclear instead of cytoplasmic injection of
DNA. To optimize the generation of CRISPR/Cas9 targeted embryos, different
concentrations of RNA and of the Nanog-mCherry and the Oct4-GFP DNA vectors
were compared as well as three different delivery modes: (i) simultaneous injection
of all constructs into the cytoplasm, (ii) simultaneous injection of the RNA and the
DNA into the pronucleus and (iii) injection of Cas9/sgRNA into the cytoplasm
followed 2 hours later by pronuclear injection of the DNA vector. Table 9 shows
that simultaneous injection of all constructs into the cytoplasm at a concentration of
100 ng/µl Cas9 RNA, 50 ng/µl of sgRNA and 200 ng/µl of vector DNA was
optimal, resulting in 8% to 18% of targeted blastocysts. Similarly, the simultaneous
injection of 5 ng/µl Cas9 RNA, 2.5 ng/µl of sgRNA and 10 ng/µl of DNA vector
yielded between 10% and 18% targeted blastocysts. In contrast, the two step
procedure with Cas9 and sgRNA simultaneous injected into the cytoplasm followed
2 hours later by pronuclear injection of different concentrations of DNA vector
yielded no or at most 3% positive blastocysts. Simultaneous injection of RNA and
DNA into the cytoplasm or nucleus is an efficient procedure to achieve targeted
insertion.

[00345] Conditional MeCP2 mutant mice

[00346] Whether conditional mutant mice can be generated in one step by
insertion of two loxP sites into the same allele of the MeCP2 gene was also
investigated herein. To derive conditional mutant mice similar to those previously
described using traditional homologous recombination methods in ES cells (Chen,
R.Z., et al., Nat Genet 27, 327-331 2001)), two sgRNAs targeting MeCP2 intron 2
(L1, L2), and two sgRNAs targeting intron 3 (R1, R2) as well as the corresponding
loxP site oligos with 60 bp homology to sequences surrounding each sgRNA
mediated DSB were designed (Fig. 21A) To facilitate detection of correct insertions,
the oligos targeting intron 2 were engineered to contain an Nhel restriction site and
the oligos targeting intron 3 to contain an EcoRI site in addition to the loxP
sequences (Fig. 18A, 21A). To determine the efficiency of single loxP site
integration at the MeCP2 locus, Cas9 mRNA was injected and each single sgRNA
and corresponding oligo into zygotes, which were cultured to the blastocyst stage and genotyped by the RFLP assay. As shown in Figure 21B, the L2 and R1 sgRNAs were more efficient in integrating the oligos with 4 out of 8 embryos carrying the L2 oligo and 2 out of 6 embryos the carrying R1 oligo (there was no amplification in DNA from 2 embryos). Therefore, L2 and R1 sgRNAs and the corresponding oligos were chosen for the generation of a floxed allele (Fig. 21A).

A total of 98 E 13.5 (Embryonic Day) embryos or mice were generated from zygotes injected with Cas9 mRNA, sgRNAs, and DNA oligos targeting the L2 and R1 sites. Genomic DNA was digested with both Nhel and EcoRI, and analyzed by Southern blot using exon 3 and 4 probes (Fig. 18A and 18B). The L2 and R1 oligos contained, in addition to the loxP site, different restriction sites (Nhel or EcoRI). Thus, single loxP site integration at L2 or R1 will produce either a 3.9 kb or a 2 kb band, respectively, when hybridized with the exon3 probe (Fig. 18A and 18B). It was found herein that about 50% of the embryos or mice carried a loxP site at the L2 site and about 25% at the R1 site. Importantly, integration of both loxP sites on the same DNA molecule, generating a floxed allele, produces a 700 bp band as detected by exon 3 probe hybridization (Fig. 18A and 18B). RFLP analysis, sequencing (Fig. 22A and 22B) and Southern blot analysis (Fig. 18B) showed that 16 out of the 98 mice tested contained two loxP sites flanking exon 3 on the same allele. Table 8 summarizes the frequency of all alleles and shows that the overall insertion frequency of an L2 or R1 insertion was slightly higher in females (32/38) than in males (38/60) consistent with the higher copy number of the X-linked MeCP2 gene in females. To confirm that the floxed allele was functional, genomic DNA was used for in vitro Cre-mediated recombination. Upon Cre treatment, both the deletion and circular products were detected by PCR in targeted mice, but not in DNA from wild type mice (Fig. 18C). The PCR products were sequenced and confirmed the precise Cre-loxP mediated recombination (Fig. 22C).

Some pups, herein, carried large deletions but no LoxP insertions, raising the possibility that two cleavage events may generate defined deletions. To confirm this, Cas9 mRNA, MeCP2-L2 and R1 sgRNAs were coinjected but without oligos. PCR genotyping and sequencing (Fig 18D and 18E) revealed that eight out of 23 mice carried deletions of about 700bp spanning the L2 and R1 sites removing exon 3. This was confirmed by Southern analysis (not shown). Because DNA breaks are
repaired through the non-homologous end joining (NHEJ) pathway, the ends of the breaks are different in different deletion alleles (Fig 18E).

[00349] Mosaicism

[00350] As mentioned above, some animals were mosaic for the targeted insertion. The frequency of mosaicism in Mecp2 targeted mice by Southern blot analysis was characterized. Since Mecp2 is an X-linked gene, in males more than one allele and in females more than two different alleles suggest mosaicism, which would be expected if integration occurred later than the zygote stage. For example, as shown in Fig. 18B, female mouse #2 contained three different alleles (one WT allele, one floxed allele, and one L2-loxP allele), and female mouse #4 contained four different alleles (one WT allele, one floxed allele, one L2-loxP allele, and one R1-loxP allele). Male mouse #5 contained two different alleles, with each allele carrying a single loxP site (Fig. 18B). Eight mosaics out of 16 mice were identified containing a Mecp2 floxed allele. The frequency of mosaicism among 49 embryos and mice containing loxP site was about 40% (Table 10). Since Southern blot analysis cannot detect small in-del mutations caused by NHEJ repair, it is possible that this underestimates the overall mosaicism frequency.

[00351] Off-target analysis

[00352] Recent studies identified a high level of off-target cleavage in human cell lines using the CRISPR/Cas system, with Cas9 targeting specificity being shown to tolerate small numbers of mismatches between sgRNA and target DNA in a sequence and position dependent manner (Fu, Y., et al., Nat Biotechnol. (2013); Hsu, P.D., et al., Nat Biotechnol (2013)). Potential off-target (OT) mutations in mice and ES cell lines were characterized derived from zygotes injected with Cas9 and sgRNAs targeting the Sox2, the Nanog and the Mecp2 gene. All genomic loci containing up to three base pair mismatches were identified compared to the 20 bp sgRNA coding sequence (Table 11). All 13 potential OT sites of Sox2 sgRNA was amplified in six mice and four ES cell lines carrying the Sox2-V5 allele and was tested for potential off target mutations using the Surveyor assay. No mutation was detected in any locus. When nine Nanog sgRNA potential OT sites were tested in five correctly targeted mice and four targeted ES cell lines, mutations were found in seven samples at OT1 (Table 11). Since Nanog OT1 has only one base pair difference at the very 5’ end of the sgRNA, it may be not surprising to find such a
high frequency of mutations at this locus. In contrast, no off-target mutation was seen in any other Nanog OTs, which contain three or four base pair difference. Finally, four potential off-targets sites for Mecp2 L2, and ten sites for Mecp2 R1 were analyzed in ten mice carrying a Mecp2 floxed allele. Only one off-target mutation was identified in one mouse at the Mecp2 R1 OT2 (Table 1). In summary, all potential off-target sites differing up to three or four base pairs in 29 mice or ES lines were tested and identified mutations in only one off-target site for Nanog (7/9 samples) and Mecp2 (1/10 samples). Thus, the off-target mutation rate is substantially lower than was observed in previous studies using cultured human cancer cell lines (Fu, Y., et al., Nat Biotechnol. (2013); Hsu, P.D., et al., Nat Biotechnol (2013)).

[00353] DISCUSSION

[00354] In this study, CRISPR/Cas technology can be used for efficient one-step insertions of a short epitope or longer fluorescent tags into precise genomic locations, which will facilitate the generation of mice carrying reporters in endogenous genes. Mice and/or embryos carrying reporter constructs in the Sox2, the Nanog and the Oct4 gene were derived from zygotes injected with Cas9 mRNA, sgRNAs and DNA oligo or vectors encoding a tag or a fluorescent marker. Also, microinjection of two Mecp2 specific sgRNAs, Cas9 mRNA and two different oligos encoding LoxP sites into fertilized eggs allowed the one-step generation of conditional mutant mice. In addition the introduction of two spaced sgRNAs targeting the Mecp2 gene produced mice carrying defined deletions of about 700 bp. Though all RNA and DNA constructs were injected into the cytoplasm or nucleus of zygotes, the gene modification events could happen at the one cell stage or later. Indeed, Southern analyses revealed mosaicism in 13% to 40% of the targeted mice and ES cell lines indicating that the insertion of the transgenes had occurred after the zygote stage (Table 10).

[00355] Previous experiments (Wang, H., et al., Cell 153, 910-918 (2013)) demonstrated herein is an efficiency of CRISPR/Cas sgRNA mediated cleavage that was high enough to allow for the one-step production of engineered mice up to 90% of which carried homozygous mutations in two genes (4 mutant alleles). The results reported here show that the sgRNA mediated DSBs occur at a significantly higher frequency than insertion of exogenous DNA sequences. Therefore, the allele not
carrying the insert will likely be mutated as a consequence of NHEJ-based gene disruption. Thus, the reporter allele would need to be segregated away from the mutant allele in order to produce mice carrying a reporter as well as a wt allele.

[00356] Two recent studies reported a high off-target mutation rate in CRISPR/Cas9 transfected human cell lines (Fu, Y., et al., Nat Biotechnol. (2013); Hsu, P.D., et al., Nat Biotechnol (2013)). The off-target rate for four different sgRNAs was analyzed and identified the cleavage of Nanog OT1 in 7 out of 9 samples and of MeCP2 R1 OT2 in 1/10 mice tested. Nanog OT1 has only one base pair difference from the targeting sequence at the extreme 5’ end (position 20, numbered 1-20 in the 3’ to 5’ direction of gRNA target site), while MeCP2 R1 OT2 has one base pair mismatch at position 20, and one mismatch at position 7. No mutations were detected in 34 potential OTs of Sox2, Nanog, Oct4 or MeCP2 containing 2, 3, or 4bp mismatches in a total of 29 mice and ES cell lines tested. This result is consistent with the previous findings that Cas9 can catalyze DNA cleavage in the presence of single-base mismatches in the PAM-distal region (Cong et al, 2013; Hsu et al, 2013; Jiang et al, 2013; Jinek et al, 2012). Consistent with the observation that three or more interspaced mismatches dramatically reduce Cas9 cleavage (Hsu et al, 2013), there were no observed off-target mutations at loci containing 3bp mismatches.

[00357] There are several possibilities to explain the significant difference in off-target cleavage rate seen in animals derived from manipulated zygotes and the results reported for CRISPR/Cas treated human cell lines (Fu, Y., et al., Nat Biotechnol. (2013); Hsu, P.D., et al., Nat Biotechnol (2013)). The off-target mutagenesis was analyzed based on the analysis of a "clonal genome" in animals derived from a single manipulated zygote, in contrast to the two previous reports that analyzed heterogeneous cell populations. The surveyor assay, based upon extensive PGR amplification, may identify any mutation, even very rare alleles that may be present in the heterogeneous population. The transformed human cell lines may have different DNA damage responses resulting in a different mutagenesis rate than the normal one cell embryo. In the experiments described herein, CRISPR/Cas was injecting as short-lived RNA in contrast to Fu et al. and Hsu et al. who used DNA plasmid transfection, which may express the Cas9/sgRNA for longer time periods leading to more extensive cleavage. Thus, this data suggests high specificity
of the CRISPR/Cas9 system for gene editing in early embryos aimed at generating gene-modified mice. Nevertheless, characterization of off-target mutagenesis of CRISPR/Cas system using whole genome sequencing would be highly informative and may allow designing sgRNAs with higher specificity.

[00358] In summary, CRISPR/Cas mediated genome editing represents an efficient and simple method of generating sophisticated genetic modifications in mice such as conditional alleles and endogenous reporters in one step. The principles described in this study could be directly adapted to other mammalian species, which provides sophisticated genome engineering in many species where ES cells are not available.

[00359] Table 7: Mice with reporters in the endogenous genes

<table>
<thead>
<tr>
<th>Donor</th>
<th>Blastocyst s / Injected zygotes</th>
<th>Targeted blastocysts / Total</th>
<th>Targeted ESCs / Total</th>
<th>Transferred embryos (recipients)</th>
<th>Knock-in pre- and postnatal mice / Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1-loxP + Tet2-loxP</td>
<td>65/89</td>
<td>Tex1-loxP 6/1 5</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tex2-loxP 6/1 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both 3/1 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sox2-V5</td>
<td>414/498</td>
<td>N/A</td>
<td>7/16</td>
<td>200(10)</td>
<td>12/35</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>936/1262</td>
<td>86/936</td>
<td>ND^</td>
<td>415 (21)</td>
<td>8/86</td>
</tr>
<tr>
<td>Oct4-GFP</td>
<td>254/345</td>
<td>47/254</td>
<td>3/9</td>
<td>100(4)</td>
<td>3/10</td>
</tr>
</tbody>
</table>

Cas9 mRNA, sgRNAs targeting Tet1, Tet2, Sox2, Nanog, or Oct4, and single stranded DNA oligos or double stranded donor vectors were injected into fertilized eggs. Targeted blastocysts were identified by RFLP or fluorescence of reporters. The blastocysts derived from injected embryos were derived ES cell lines or transplanted into foster mothers and E13.5 embryos and postnatal mice were obtained and genotyped.

^Only mCherry positive blastocysts were selected to generate ES cell lines.
ND, not determined; N/A, not applicable.
Table 8: Conditional Mecp2 mutant mice

<table>
<thead>
<tr>
<th>Donor</th>
<th>Blastocyst / Injected zygotes</th>
<th>Transferred embryos (recipients)</th>
<th>Sex</th>
<th>(Pre and post) Mice with loxP / Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total ^a</td>
<td>L2 ^b</td>
</tr>
<tr>
<td>Mecp2-L2 + Mecp2-R1</td>
<td>367/451</td>
<td>360 (18)</td>
<td>Male</td>
<td>28/60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Female</td>
<td>21/38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>49/98</td>
</tr>
</tbody>
</table>

Cas9 mRNA, sgRNAs targeting Mecp2-L2 and Mecp2-R1, and single stranded DNA oligos were injected into fertilized eggs. The blastocysts derived from the injected embryos were transplanted into foster mothers and pre- and postnatal mice were genotyped.

^aTotal mice containing loxP site integration in the genome.

^bMice containing loxP site integrated at L2 site.

^cMice containing loxP site integrated at R1 site.

^dThese male mice were mosaic.

Table 9: Efficiency of generation of reporter embryos by cytoplasm and pronuclear injection

<table>
<thead>
<tr>
<th>Donor</th>
<th>Dose of Cas9/sgRNA (ng/µl)</th>
<th>Dose of Donor vector (ng/µl)</th>
<th>Injected zygotes</th>
<th>Targeted Blastocysts / Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-step injection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>500(Cyto)</td>
<td>186</td>
<td>1/81</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>200(Cyto)</td>
<td>1262</td>
<td>86/936</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>50(Cyto)</td>
<td>402</td>
<td>7/308</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>10(Cyto)</td>
<td>333</td>
<td>1/278</td>
</tr>
<tr>
<td>Oct4-GFP</td>
<td>100/50 (Cyto)</td>
<td>200(Cyto)</td>
<td>345</td>
<td>47/254</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>5/2.5 (Nuc)</td>
<td>10 (Nuc)</td>
<td>98</td>
<td>7/75</td>
</tr>
<tr>
<td>Oct4-GFP</td>
<td>5/2.5 (Nuc)</td>
<td>10 (Nuc)</td>
<td>105</td>
<td>13/72</td>
</tr>
<tr>
<td><strong>Two-step injection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>50 (Nuc)</td>
<td>45</td>
<td>0/0</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>10 (Nuc)</td>
<td>91</td>
<td>1/34</td>
</tr>
<tr>
<td>Nanog-mCherry</td>
<td>100/50 (Cyto)</td>
<td>2 (Nuc)</td>
<td>85</td>
<td>1/68</td>
</tr>
</tbody>
</table>

Cas9 mRNA, sgRNAs targeting Nanog, or Oct4, and double stranded donor vectors were injected into cytoplasm or pronuclei of zygotes. In one-step injection, the RNA and the DNA were simultaneously injected into the cytoplasm or pronucleus. In two-step injection, Cas9/sgRNA were injected into the cytoplasm followed 2 hours.
later by pronuclear injection of the DNA vector. Targeted blastocysts were identified by fluorescence of reporters. Cyto, cytoplasm; Nuc, nucleus.

Table 10: Mosaicism in targeted mice

<table>
<thead>
<tr>
<th>Donor</th>
<th>Mosaic / Total targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanog-cherry</td>
<td>Mice 1/8, ESCs 2/6</td>
</tr>
<tr>
<td>Oct4-EGFP</td>
<td>Mice 0/3, ESCs 1/3</td>
</tr>
<tr>
<td>Mecp2-L2 +</td>
<td>Male 11/28, Female 9/21</td>
</tr>
<tr>
<td>Mecpe-Rl</td>
<td>Total 20/49*</td>
</tr>
</tbody>
</table>

Targeted mice or ESCs were identified by RFLP, Southern blot or Sequencing. The frequency of mosaicism in targeted mice was determined by fluorescent reporter or Southern blot analysis.

* These 49 mice contain at least one loxP integration.

Table 11: Off-target analysis

<table>
<thead>
<tr>
<th>Site name</th>
<th>Sequence</th>
<th>Indel mutation frequency (Mutant/Total)</th>
<th>Coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Sox2 Stop</td>
<td>TGCCCCCTGTGCGCAGATGGAGGG</td>
<td>/</td>
<td>chr3: 34550278-34550300</td>
</tr>
<tr>
<td>OT1 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 126636377-126636399</td>
</tr>
<tr>
<td>OT2 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 58830941-58830963</td>
</tr>
<tr>
<td>OT3 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 136641-136641 196</td>
</tr>
<tr>
<td>OT4 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 69081892-69081914</td>
</tr>
<tr>
<td>OT5 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr5: 130633965-130633987</td>
</tr>
<tr>
<td>OT6 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr5: 61611640-61611662</td>
</tr>
<tr>
<td>OT7 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr5: 136841014-136841036</td>
</tr>
<tr>
<td>OT8 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 141162344-141162456</td>
</tr>
<tr>
<td>OT9 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 48224249-48224451</td>
</tr>
<tr>
<td>OT10 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr7: 72596616-72596638</td>
</tr>
<tr>
<td>OT11 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr8: 56473819-56473841</td>
</tr>
<tr>
<td>OT12 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr6: 98776389-9877641</td>
</tr>
<tr>
<td>OT13 Sox2 Stop</td>
<td>TGCCCGTGTGCGCAGATGGAGGG</td>
<td>0/10</td>
<td>chr4: 148868089-14886811</td>
</tr>
</tbody>
</table>
Target Nanog_Stop CGTAAGTCTCATATTCACCTGG / chr6: 122663559-122663581
(SEQ ID NO: 230)
OT1 Nanog_Stop gTGAAGTCTATATTCACCTGG (SEQ ID NO: 231)
OT2 Nanog_Stop CgAAAGTCTATTTTCACCAG (SEQ ID NO: 232)
OT3 Nanog_Stop GaTAAGTCATATTCACCCGG (SEQ ID NO: 233)
OT4 Nanog_Stop TGcAATTCATATTTCACCTGG (SEQ ID NO: 234)
OT5 Nanog_Stop GcTAAGTCATATTTCCACCC (SEQ ID NO: 235)
OT6 Nanog_Stop TGcAATTCATATTTCACCTGG (SEQ ID NO: 236)
OT7 Nanog_Stop TGcAATTCATATTTCACCTGG (SEQ ID NO: 237)
OT8 Nanog_Stop TGcAATTCATATTTCACCTGG (SEQ ID NO: 238)
OT9 Nanog_Stop CgTAAGTCATATTTCCACCC (SEQ ID NO: 239)

Target Mecp2_Rl AGGAGTGAGGTCTAGTACTTGGG / chrX: 71282103-71282125
(SEQ ID NO: 245)
OT1 Mecp2_Rl AGGgGTGAGtTCTAGTACTTCGG (SEQ ID NO: 246)
OT2 Mecp2_Rl TGGAGTGAGGTCTtGTACTTGGG (SEQ ID NO: 247)
OT3 Mecp2_Rl AGGAGTctGGgCTAGTACTTGGG (SEQ ID NO: 248)

Mismatches from the on-target sequence are lower-case, boldface and underlined.

Indel mutation frequencies in targeted mice or ESCs were calculated by Surveyor assay. Coordinate in which sites were located are shown. OT, off-target; /, not tested. aNanog OT1 and 2 contain 3bp mismatches; OT3 to 9 contain 4bp mismatches lying in PAM distal region.

bPCR products were cloned and sequenced to confirm off-target mutations.

[00364] Table 12. Oligonucleotides used in this study.

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>Caccggtgtgctgtcagggactc (SEQ ID NO: 249)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Aaactgagctccctgccagcagc (SEQ ID NO: 250)</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>caccggaaagggccagactcagtagctcc (SEQ ID NO: 251)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaacggatctctgctgacacccagtcttgct (SEQ ID NO: 252)</td>
</tr>
<tr>
<td>Sox2</td>
<td>F</td>
<td>caccggtccctctgctgcacatgtga (SEQ ID NO: 253)</td>
</tr>
</tbody>
</table>
Oligonucleotides used for RFLP analysis and PCR genotyping.

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>ttgtctctctctgactgc  (SEQ ID NO: 267)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>tgattgataaagtgcctgc  (SEQ ID NO: 268)</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>cagatgcctagcgcaataaaag  (SEQ ID NO: 269)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aagaagcacaacatcagagatg  (SEQ ID NO: 270)</td>
</tr>
<tr>
<td>Sox2</td>
<td>SF</td>
<td>acagtgcatcagcctccttc  (SEQ ID NO: 271)</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>taatgattgaggtggtggtg  (SEQ ID NO: 272)</td>
</tr>
<tr>
<td>V5</td>
<td>V5F</td>
<td>acaggcccaagccccccttcc  (SEQ ID NO: 273)</td>
</tr>
<tr>
<td>Mecp2 L1,L2</td>
<td>LF</td>
<td>aatgctggccatcccttgac  (SEQ ID NO: 274)</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>ttctgatgttctttgtcctgc  (SEQ ID NO: 275)</td>
</tr>
<tr>
<td>Mecp2 R1,R2</td>
<td>RF</td>
<td>aagcgatgcgctgagccaaaaacc  (SEQ ID NO: 276)</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>ccgtgctcagaaggcacaaaaag  (SEQ ID NO: 277)</td>
</tr>
</tbody>
</table>

Oligonucleotides used for making template for in vitro transcription.

<table>
<thead>
<tr>
<th>Template</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9</td>
<td>F</td>
<td>Ttatacgactcataagggagagaattgacatggagctcag (SEQ ID NO: 278)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gcgagcttctggaatttcct  (SEQ ID NO: 279)</td>
</tr>
<tr>
<td>Tet1 sgRNA</td>
<td>F</td>
<td>Ttatacgactcataagggagagaattgacatctcagggagctc  (SEQ ID NO: 280)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaaagccacaggccgccgtgctc  (SEQ ID NO: 281)</td>
</tr>
<tr>
<td>Tet2 sgRNA</td>
<td>F</td>
<td>Ttatacgactcataagggagagaattgacatccagatcc  (SEQ ID NO: 282)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaaagccacaggccgccgtgctc  (SEQ ID NO: 283)</td>
</tr>
<tr>
<td>Sox2 sgRNA</td>
<td>F</td>
<td>Ttatacgactcataagggagagaattgacatctcagggagctc  (SEQ ID NO: 284)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaaagccacaggccgccgtgctc  (SEQ ID NO: 285)</td>
</tr>
<tr>
<td>Nanog sgRNA</td>
<td>F</td>
<td>Ttatacgactcataagggagagaattgacatctcagggagctc  (SEQ ID NO: 286)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>aaaagccacaggccgccgtgctc  (SEQ ID NO: 287)</td>
</tr>
<tr>
<td>Oct4</td>
<td>F</td>
<td>Ttatacgactcataagggagagaattgacatctcagggagctc  (SEQ ID NO: 288)</td>
</tr>
<tr>
<td>sgRNA</td>
<td>R</td>
<td>aaaagcaccgactcgggctgcc (SEQ ID NO: 289)</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Mecp2-L1 sgRNA</td>
<td>F</td>
<td>Ttaatacgactcactatagttggccccagcttgacc (SEQ ID NO: 290)</td>
</tr>
<tr>
<td>Mecp2-L2 sgRNA</td>
<td>F</td>
<td>Ttaatacgactcactatagggctggtaagagtatctctc (SEQ ID NO: 292)</td>
</tr>
<tr>
<td>Mecp2-R1 sgRNA</td>
<td>F</td>
<td>Ttaatacgactcactataggtctctagctgtacct (SEQ ID NO: 294)</td>
</tr>
<tr>
<td>Mecp2-R2 sgRNA</td>
<td>F</td>
<td>Ttaatacgactcactatagttggctggtggactagttct (SEQ ID NO: 296)</td>
</tr>
<tr>
<td>Mecp2-L1 sgRNA</td>
<td>R</td>
<td>aaaagcaccgactcgggctgcc (SEQ ID NO: 291)</td>
</tr>
<tr>
<td>Mecp2-L2 sgRNA</td>
<td>R</td>
<td>aaaagcaccgactcgggctgcc (SEQ ID NO: 293)</td>
</tr>
<tr>
<td>Mecp2-R1 sgRNA</td>
<td>R</td>
<td>aaaagcaccgactcgggctgcc (SEQ ID NO: 295)</td>
</tr>
<tr>
<td>Mecp2-R2 sgRNA</td>
<td>R</td>
<td>aaaagcaccgactcgggctgcc (SEQ ID NO: 297)</td>
</tr>
</tbody>
</table>

[00368] Oligonucleotides used for HDR-mediated repair through embryo injection

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetl- loxP</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 298)</td>
</tr>
<tr>
<td>Tet2- loxP</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 299)</td>
</tr>
<tr>
<td>Sox2-v5</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 300)</td>
</tr>
<tr>
<td>Mecp2-L1- loxP</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 301)</td>
</tr>
<tr>
<td>Mecp2-L2- loxP</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 302)</td>
</tr>
<tr>
<td>Mecp2-R1- loxP</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 303)</td>
</tr>
<tr>
<td>Mecp2-R2- loxP</td>
<td>Taccagaggggcccttaaccctggggcacggacacgtatctgggtgggtctacgggaaggtgtctggctctgggccccttacccactgtcctctcctcctggggccctctggggtcctctgggccccagcttgacc (SEQ ID NO: 304)</td>
</tr>
</tbody>
</table>
### Oligonucleotides used for off-targeted analysis

<table>
<thead>
<tr>
<th>Gene target</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>OT1_Sox2_Stop</td>
<td>F</td>
<td>Atgacatgacctagaagtaaaccc (SEQ ID NO: 305)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Ctcpactgtactagcccagcc (SEQ ID NO: 306)</td>
</tr>
<tr>
<td>OT2_Sox2_Stop</td>
<td>F</td>
<td>tgcaggttttgtgacttgctg (SEQ ID NO: 307)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatctatagtagaaatttg (SEQ ID NO: 308)</td>
</tr>
<tr>
<td>OT3_Sox2_Stop</td>
<td>F</td>
<td>aaactgagccagactgcttg (SEQ ID NO: 309)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gtaactgacacctgtaag (SEQ ID NO: 310)</td>
</tr>
<tr>
<td>OT4_Sox2_Stop</td>
<td>F</td>
<td>oaccttagggtcarttgcatcc (SEQ ID NO: 311)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gttgactacgatgtaagagc (SEQ ID NO: 312)</td>
</tr>
<tr>
<td>OT5_Sox2_Stop</td>
<td>F</td>
<td>accatgagctgtactaggcac (SEQ ID NO: 313)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>cattgacgcattagctagt (SEQ ID NO: 314)</td>
</tr>
<tr>
<td>OT6_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 315)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 316)</td>
</tr>
<tr>
<td>OT7_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 317)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>cattgacgcattagctagt (SEQ ID NO: 318)</td>
</tr>
<tr>
<td>OT8_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 319)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 320)</td>
</tr>
<tr>
<td>OT9_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 321)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>cattgacgcattagctagt (SEQ ID NO: 322)</td>
</tr>
<tr>
<td>OT10_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 323)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 324)</td>
</tr>
<tr>
<td>OT11_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 325)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 326)</td>
</tr>
<tr>
<td>OT12_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 327)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 328)</td>
</tr>
<tr>
<td>OT13_Sox2_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 329)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 330)</td>
</tr>
<tr>
<td>OT1_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 331)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 332)</td>
</tr>
<tr>
<td>OT2_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 333)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 334)</td>
</tr>
<tr>
<td>OT3_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 335)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 336)</td>
</tr>
<tr>
<td>OT4_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 337)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 338)</td>
</tr>
<tr>
<td>OT5_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 339)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 340)</td>
</tr>
<tr>
<td>OT6_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 341)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 342)</td>
</tr>
<tr>
<td>OT7_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 343)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 344)</td>
</tr>
<tr>
<td>OT8_Nanog_Stop</td>
<td>F</td>
<td>Ftgatggtttttggtgactgcc (SEQ ID NO: 345)</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>gacagatcatagatagaaaattg (SEQ ID NO: 346)</td>
</tr>
</tbody>
</table>
Example 4

Multiplexed activation of endogenous genes by CRISPR-on, a RNA-guided transcriptional activator system

As described in Example 2, a two-component transcriptional activator consisting of a nuclease-dead Cas9 (dCas9) protein fused with a transcriptional activation domain and single guide RNAs (sgRNAs) with complementary sequence to gene promoters. It is demonstrated that CRISPR-on can efficiently activate exogenous reporter genes in both human and mouse cells in a tunable manner. In addition, robust reporter gene activation in vivo can be achieved by injecting the system components into mouse zygotes. Furthermore, CRISPR-on can activate the endogenous ILIRN, SOX2, and OCT4 genes. The most efficient gene activation was...
achieved by clusters of 3 to 4 sgRNAs binding to the proximal promoters suggesting their synergistic action in gene induction. Significantly, when sgRNAs targeting multiple genes were simultaneously introduced into cells, robust multiplexed endogenous gene activation was achieved. Genome-wide expression profiling demonstrated high specificity of the system.

[00373] Materials and Methods

[00374] Cloning

[00375] A two-step fusion PCR was performed to amplify Cas9 Nickase ORF without stop codon from the pX335 vector (Addgene: 42335), incorporate H840A mutation, EcoRI-Agel restriction site on the 5’ end as well as an FseI site on the 3’ end (EcoRI-Agel-dCas9-FseI fragment). The 3x minimal VP16 activation domain coding fragment (VP48) was excised from a vector (Addgene: 20342) containing NLSM2rtTA coding sequence by FseI and EcoRI digestion (FseI-TA-EcoRI fragment). The two fragments were ligated into pCR8/GW/TOPO (Invitrogen) vector digested by EcoRI to generate a gateway compatible dCas9VP48 coding plasmid. The dCas9VP48 coding sequence was subsequently excised and cloned into pX355 vector (Addgene: 42335) by Agel-EcoRI digestion to replace dCas9 Nickase to create a chimeric vector that expresses both the dCas9VP48 and the sgRNA (dCas9VP48-U6-sgRNA-chimeric). sgRNA spacers were cloned into the BbsI-digested vector by annealing oligos as previously described (Cong et al, Science; 339 (6121):819-823(2013)). For construction of dCas9VP160 (SEQ ID NO: 16), a gBlocks gene fragment containing coding sequence for 10 tandem repeats of VP16 domains separated by Glycine-Serine (GS) linker was ordered from Integrated DNA Technology (IDT) and amplified by PCR primers containing FseI and EcoRI sites to replace VP48 fragment in pCR8-dCas9VP48 to generate pCR8-dCas9VP160. A pmax-DEST gateway destination vector was constructed by replacing GFP coding sequence in pmaxGFP (Clontech) by a gateway destination cassette (Invitrogen). The pCR8-dCas9VP160 vector was then recombined with pmax-DEST via LR Clonase-mediated to create pmax-dCas9VP160 expression plasmid. For the endogenous gene experiments, sgRNAs were cloned by oligo clonding method mentioned above to a PBneo-sgRNA expression vector.

[00376] Culturing and transfection of HeLa, HEK293T and NIH3T3
HeLa, HEK293T and NIH3T3 cells were cultured in DMEM with 10% inactivated FBS, 1% Penn/Strep, 1% Glutamine, 1% non-essential amino acids. Transfection was done using Fugene HD (Promega) using a 2:6 ratio (A total DNA amount of 2µg and 6µl of Fugene HD reagent) in 6-well plates. For TetO::tdTomato experiments, 2µg of the chimeric vector was used. For endogenous gene activation experiments, the U6 promoter - sgRNA - terminator sequence was amplified from the PBneo-sgRNA plasmids, purified by PCR purification kit (QIAGEN), and transfected as linear DNA ( Total sgRNA expressing DNA) with 1µg of pmax-dCas9VP160 plasmid. Where there are multiple sgRNAs for multiple genes, the amount per sgRNA was evenly divided among genes first, then among the sgRNAs targeting each gene.

Transgene activation experiment in mouse embryonic stem cells (mESC)

Mouse ESCs from mice carrying a Dox-inducible Musashi-1 (MSI1) allele in the CollAl locus (Kharas et al, Nature medicine; 16 (8):903-908 (2010)) were transfected with dCas9VP48 using Xfect mESC transfection reagent (Clontech) or were cultured in mouse ES medium with 2 µg/ml Doxycycline. 48 hours later, Protein lysates were prepared on ice from cell pellets in SDS-Tris lysis buffer (10% SDS, 10% Glycerol, 0.1M DTT, 0.12g/ml Urea) supplemented with protease and phosphatase inhibitor tables (1 tablet/10ml, Roche) and analyzed by western blot. Blots were probed with primary rabbit anti-MSI1 (Cell Signaling Technologies, #2154), mouse anti-Alpha-Tubulin (SIGMA) antibodies. Secondary HRP-conjugated anti-rabbit/anti-mouse IgG were used and visualized with ECL (GE Healthcare).

One cell embryo injection

All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at MIT. B6D2F1 (C57BL/6 X DBA2) female mice and ICR mouse strains were used as embryo donors and foster mothers, respectively. Super-ovulated female B6D2F1 mice (7-8 weeks old) were mated to B6D2F1 stud males, and fertilized embryos were collected from oviducts. Cas9VP48 plasmid (200 ng/µl), Nanog::EGFP construct (200 ng/µl), and sgRNAs (50 ng/µl for each) were mixed and injected into the cytoplasm of fertilized eggs with well-recognized pronuclei in M2 medium (Sigma). Injected oocytes were cultured in KSOM medium for 96 h to examine their development in vitro. Images
of resulting embryos were acquired with an inverted microscope under the same exposure parameters.

[00382] Bioinformatics analysis of gene expression and CRISPR off-target analysis

[00383] Affymetrix U133A 2.0 array was used for microarray gene expression analysis. Gene expression values were processed and normalized using affy package for R [Gautier, 2004 #27].

[00384] qRT-PCR expression analysis

[00385] Total RNA was isolated using the Rneasy Kit (QIAGEN) and reversed transcribed using the Superscript III First Strand Synthesis kit (Invitrogen). Quantitative RT-PCR analysis was performed in triplicate using the ABI 7900 HT system with FAST SYBR Green Master Mix (Applied Biosystems). Gene expression was normalized to GAPDH. Error bars represent the standard deviation (SD) of the mean of triplicate reactions. Primer sequences are included in Table 13.

[00386] Table 13 qRT-PCR primers.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOX2</td>
<td>CACCTACAGCATGTCTACTCG (SEQ ID NO: 377)</td>
<td>GGGTTTCTCCATGCTTTTCT (SEQ ID NO: 378)</td>
</tr>
<tr>
<td>IL1RN</td>
<td>GACCCCTCTGGGAGAAAATCC (SEQ ID NO: 379)</td>
<td>GTCCCTTGCAAGTATCCAGCA (SEQ ID NO: 380)</td>
</tr>
<tr>
<td>OCT4</td>
<td>GCTCGAGAAGGATGTGGTCC (SEQ ID NO: 381)</td>
<td>CTGGTGCATAGTGCTGCT (SEQ ID NO: 382)</td>
</tr>
<tr>
<td>GAPDH</td>
<td>CGAGATCCCTCCAAAATCAA (SEQ ID NO: 383)</td>
<td>ATCCACAGTCTTCTGGGTGG (SEQ ID NO: 384)</td>
</tr>
</tbody>
</table>
### Table 14: sgRNA designs, DNA targets, oligos for cloning.

<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Target sequences</th>
<th>Forward oligo</th>
<th>Reverse oligo</th>
</tr>
</thead>
<tbody>
<tr>
<td>sgTetO</td>
<td>Tet binding site</td>
<td>gCTTTTCTCTATC</td>
<td>caccGCTTTTCTCTATT</td>
<td>aacTATCAGTGAT AGAGAAAAAGC (SEQ ID NO: 385)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACTGATAGGG</td>
<td>CACTGATA</td>
<td>(SEQ ID NO: 386)</td>
</tr>
<tr>
<td>sgTetOMut</td>
<td>Mutant version of sgTetO</td>
<td>gCTTTTCTTaATC</td>
<td>caccGCTTTTCTTTAT</td>
<td>aacTACCAATGAT AGAGAAAAGC (SEQ ID NO: 387)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGgTAGGG</td>
<td>CATTTGTA</td>
<td>(SEQ ID NO: 389)</td>
</tr>
<tr>
<td>sgNanol-1</td>
<td>mouse Nanog</td>
<td>GTAAcGAAAAAG</td>
<td>caccGTAAcGCAAA</td>
<td>aacTACgAGCTTCT TTTGcATTAc (SEQ ID NO: 391)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAGcTGAAAGGG</td>
<td>GAAGcTGAAGA</td>
<td>(SEQ ID NO: 392)</td>
</tr>
<tr>
<td>sgNanol-2</td>
<td>mouse Nanog</td>
<td>GACTcTCTcAGGc</td>
<td>caccGACTcTcTAGTG</td>
<td>aacGAAcACTcTCC AcTaGAGAc (SEQ ID NO: 393)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAAGTTTcCAGG</td>
<td>GcAgGTTTcCc</td>
<td>(SEQ ID NO: 395)</td>
</tr>
<tr>
<td>sgNanol-3</td>
<td>mouse Nanog</td>
<td>GcTcTTcACATtGc</td>
<td>caccGCtTcACATc</td>
<td>aacGcTTcCATTtCc TcTcAcAGAc (SEQ ID NO: 397)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGAacCCTGGc</td>
<td>GCcGAacCTGc</td>
<td>(SEQ ID NO: 400)</td>
</tr>
<tr>
<td>sgNanol-4</td>
<td>mouse Nanog</td>
<td>GcTcTTcACATGc</td>
<td>caccGCtTcACATc</td>
<td>aacGcTTcCATTtCc TcTcAcAGAc (SEQ ID NO: 399)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGAacCCTGGc</td>
<td>GCcGAacCTGc</td>
<td>(SEQ ID NO: 401)</td>
</tr>
<tr>
<td>sgNanol-5</td>
<td>mouse Nanog</td>
<td>GCgTTaaaacGc</td>
<td>caccGCgTTaaacAG</td>
<td>aacGcAAcTGcGcGc TtTcAaCcGc (SEQ ID NO: 403)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCGcACTTTcGc</td>
<td>CcGCcACTTTc</td>
<td>(SEQ ID NO: 404)</td>
</tr>
<tr>
<td>sgNanol-6</td>
<td>mouse Nanog</td>
<td>GAGcGTtTTaaaTc</td>
<td>caccGAGcGTtTTaaA</td>
<td>aacCTcAcATtaAT TtAaAcATCc (SEQ ID NO: 406)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AATcTGATGAGGc</td>
<td>TcTcAGATGc</td>
<td>(SEQ ID NO: 407)</td>
</tr>
<tr>
<td>sgNanol-7</td>
<td>mouse Nanog</td>
<td>GAGcGTtTTcAGGc</td>
<td>caccGAGcGTtTTcAGG</td>
<td>aacCTcAcATtaAT TtAaAcATCc (SEQ ID NO: 408)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCGcAGACTGc</td>
<td>CcCCcAGACTc</td>
<td>(SEQ ID NO: 409)</td>
</tr>
<tr>
<td>sgIL1RN-1</td>
<td>IL1RN</td>
<td>GCACCCACTGAGG</td>
<td>caccGCACCCACTGAGG</td>
<td>aacGcTcCGGcGTA CGTGAAcActC (SEQ ID NO: 410)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTACAAGGAGG</td>
<td>AGTACAAGG</td>
<td>(SEQ ID NO: 411)</td>
</tr>
<tr>
<td>sgIL1RN-2</td>
<td>IL1RN</td>
<td>GCAGCCACTGAGG</td>
<td>caccGCACCCACTGAGG</td>
<td>aacGcTcCGGcGTA CGTGAAcActC (SEQ ID NO: 412)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCGCAGGcAGG</td>
<td>ATGCcAGGcAGG</td>
<td>(SEQ ID NO: 413)</td>
</tr>
<tr>
<td>sgIL1RN-3</td>
<td>IL1RN</td>
<td>gGGcGTGACTTTGAGG</td>
<td>caccGcGGcGTGACTTTG</td>
<td>aacGcTTcGcGcTAC TcTGcAGGtc (SEQ ID NO: 414)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GCGcAGGcAGG</td>
<td>ATGCcAGGcAGG</td>
<td>(SEQ ID NO: 415)</td>
</tr>
<tr>
<td>sgIL1RN-4</td>
<td>IL1RN</td>
<td>gGTtTTcGAGGAGG</td>
<td>caccGGTtTTcGAGGAGG</td>
<td>aacGcGTCACGcC CcTGcAGGtc (SEQ ID NO: 416)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTACACTGAGG</td>
<td>GGGTcAGCTc</td>
<td>(SEQ ID NO: 417)</td>
</tr>
<tr>
<td>sgIL1RN-5</td>
<td>IL1RN</td>
<td>gGGcGTtTTcTCTG</td>
<td>caccGGTtTTcTCT</td>
<td>aacGcTCTTcACGcGcA (SEQ ID NO: 418)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(SEQ ID NO: 419)</td>
</tr>
<tr>
<td>sgILlRN-5</td>
<td>IL1RN</td>
<td>CGTAAGATGG (SEQ ID NO: 421)</td>
<td>GCGTAAGA (SEQ ID NO: 422)</td>
<td>ATAAGAACCC (SEQ ID NO: 423)</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>sgILlRN-6</td>
<td>IL1RN</td>
<td>GATTTGGAACAA (SEQ ID NO: 424)</td>
<td>caccGATTGGGAACA (SEQ ID NO: 425)</td>
<td>aaacGTCTGGCTT (SEQ ID NO: 426)</td>
</tr>
<tr>
<td>sgSOX2-1</td>
<td>SOX2</td>
<td>GGGGAGAGGAGG (SEQ ID NO: 430)</td>
<td>caccGGGGAGAGGAGG (SEQ ID NO: 431)</td>
<td>aaacCTCCCCCTCC (SEQ ID NO: 432)</td>
</tr>
<tr>
<td>sgSOX2-2</td>
<td>SOX2</td>
<td>GAGAGAGGCAAA (SEQ ID NO: 433)</td>
<td>caccGAGAGAGGCAA (SEQ ID NO: 434)</td>
<td>aaacGATTCCAGTT (SEQ ID NO: 435)</td>
</tr>
<tr>
<td>sgSOX2-3</td>
<td>SOX2</td>
<td>GCTGCCGGGTTTT (SEQ ID NO: 436)</td>
<td>caccGCTGCCGGGTTTT (SEQ ID NO: 437)</td>
<td>aaacTTCATGCAAA (SEQ ID NO: 440)</td>
</tr>
<tr>
<td>sgSOX2-4</td>
<td>SOX2</td>
<td>GCCGGCCGCGCG (SEQ ID NO: 441)</td>
<td>caccGCCGGCCGCGCG (SEQ ID NO: 442)</td>
<td>aaacGCCTCCCCCTCC (SEQ ID NO: 443)</td>
</tr>
<tr>
<td>sgSOX2-5</td>
<td>SOX2</td>
<td>GCCGGCCGCGCG (SEQ ID NO: 442)</td>
<td>caccGCCGGCCGCGCG (SEQ ID NO: 443)</td>
<td>aaacGCCTCCCCCTCC (SEQ ID NO: 444)</td>
</tr>
<tr>
<td>sgSOX2-6</td>
<td>SOX2</td>
<td>GCCAGGCGAGGAGG (SEQ ID NO: 445)</td>
<td>caccGCCAGGCGAGGAGG (SEQ ID NO: 446)</td>
<td>aaacCCTCCCCCTCC (SEQ ID NO: 447)</td>
</tr>
<tr>
<td>sgSOX2-7</td>
<td>SOX2</td>
<td>GTATCCCCCTCAGGAACACAGG (SEQ ID NO: 448)</td>
<td>caccGTATCCCCCTCAGGAACACAGG (SEQ ID NO: 449)</td>
<td>aaacGTGGTGGCGG (SEQ ID NO: 450)</td>
</tr>
<tr>
<td>sgSOX2-8</td>
<td>SOX2</td>
<td>GCAGGGTACTTA (SEQ ID NO: 451)</td>
<td>caccGCAGGGTACTTA (SEQ ID NO: 452)</td>
<td>aaacTCCTCATTGAA (SEQ ID NO: 453)</td>
</tr>
<tr>
<td>sgSOX2-9</td>
<td>SOX2</td>
<td>GCAGCTAAGGTTGCGGGG (SEQ ID NO: 454)</td>
<td>caccGCAGCTAAGGTTGCGGGG (SEQ ID NO: 455)</td>
<td>aaacCACCESSC (SEQ ID NO: 456)</td>
</tr>
<tr>
<td>sgSOX2-10</td>
<td>SOX2</td>
<td>GGCCTGCACTACCCTG (SEQ ID NO: 457)</td>
<td>caccGGCTGCACTACCCTG (SEQ ID NO: 458)</td>
<td>aaacGAAATACGA (SEQ ID NO: 459)</td>
</tr>
<tr>
<td>sgOCT4-1</td>
<td>OCT4</td>
<td>GAAGGAAGGCGC (SEQ ID NO: 460)</td>
<td>caccGAAGGAAGGCGC (SEQ ID NO: 461)</td>
<td>aaacGGCTTGGGCGC (SEQ ID NO: 462)</td>
</tr>
<tr>
<td>sgOCT4-2</td>
<td>OCT4</td>
<td>OCT4</td>
<td>GGTGAAATGAGG</td>
<td>caccGGTGAAATGAGG</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>---------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGCTTGCGA</td>
<td>GTGCTTCCTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGTGAAATGAGG</td>
<td>caccGGTGAAATGAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGCTTGCGA</td>
<td>CTGCTTCCTTC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sgOCT4-3</th>
<th>OCT4</th>
<th>OCT4</th>
<th>GGGGGGAGAAAC</th>
<th>caccGGGGGAGAAAC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGAGGCAGAGAG</td>
<td>CTGAGGCAGAGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGGGGGAGAAAC</td>
<td>caccGGGGGAGAAAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGAGGCAGAGAG</td>
<td>CTGAGGCAGAGAG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sgOCT4-4</th>
<th>OCT4</th>
<th>OCT4</th>
<th>GGTGGTGGAAT</th>
<th>caccGGTGGTGGAAT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGGGGGAGAAAC</td>
<td>caccGGGGGAGAAAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGAGGCAGAGAG</td>
<td>CTGAGGCAGAGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGTGGTGGAAT</td>
<td>caccGGTGGTGGAAT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sgOCT4-5</th>
<th>OCT4</th>
<th>OCT4</th>
<th>GACACAACTGGG</th>
<th>caccGACACAACTGGG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>GCCCTCCAGG</td>
<td>CGGCCCTCCAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GACACAACTGGG</td>
<td>caccGACACAACTGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GCCCTCCAGG</td>
<td>CGGCCCTCCAGG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sgOCT4-6</th>
<th>OCT4</th>
<th>OCT4</th>
<th>GGATGGGTGG</th>
<th>TGGATGGGTGG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>GACACAACTGGG</td>
<td>caccGACACAACTGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GCCCTCCAGG</td>
<td>CGGCCCTCCAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGATGGGTGG</td>
<td>TGGATGGGTGG</td>
</tr>
</tbody>
</table>

Last three bases are PAM (5'–NGG-3') motif. Lowercase italic letters in the target sequences indicate mismatch (first g to allow efficient U6 transcription) or as mutant control (other changes; as in sgTetO-mut). Lowercase letters in the oligo sequences indicate overhang compatible to the BbsI-digested vectors.

[00387] Results

[00388] Fusion of nuclease-deficient Cas9 to transactivation domain generated an RNA-programmable transcription factor

[00389] To generate a CPJSPR/Cas-based transcription activator, the H840A mutation was introduced in the human codon-optimized Cas9(D10A) nickase (Cong et al., Science; 339 (6121):819-823(2013)) to create a nuclease-deficient dCas9 (H840A; D10A) and fused a 3x minimal VP16 transcriptional activation domain (VP48) to its C-terminus (Figure 23A). We first tested dCas9VP48 in human HeLa cells carrying integrated tdTomato reporter transgene under the control of a Tetracycline-inducible promoter composed of seven copies of rtTA binding sites and a CMV minimal promoter (TetO::tdTomato). As a positive control, these cells constitutively express the rtTA transactivator that induced tdTomato expression upon doxycycline treatment (Figure 23B panel ii). Transient transfection of dCas9VP48 with sgRNA complementary to rtTA binding site (sgTetO) activated the
TetO::tdTomato reporter in the absence of doxycycline at almost the same efficiency as the positive control (Figure 23B panel iv). Transfection of dCas9VP48 without sgRNA did not activate tdTomato expression. (Figure 23B panel iii). Activation of a TetO::tdTomato reporter lasted for one week but became weak afterwards (Figure 27). Similarly, co-expression of dCas9VP48 with sgTetO activated tdTomato transgene in mouse NIH3T3 cells carrying an integrated TetO::tdTomato reporter (Figure 28B, panel iv), while expression of dCas9VP48 alone did not activate tdTomato expression (Figure 28B, panel iii). These results indicate that CRISPR-on activates a transgene reporter robustly in human and mouse cells to a similar level as rtTA in the presence of doxycycline and that the binding of dCas9VP48 to the tetO promoter is strictly dependent on sgTetO. The higher fraction of fluorescent HeLa cells as compared to that in NIH3T3 cells is likely due to higher transfection efficiency.

CRISPR-on was tested whether it could activate a single-copy transgene in embryonic stem cells (ESC). For this dCas9VP48 was co-transfected with sgTetO into ESC cells carrying a Tet-inducible Musashil (MSI1) transgene at the Col1a locus and the rtTA-M2 in the Rosa26 locus (Kharas et al., Nature medicine; 16 (8):903-908 (2010)) (Figure 29). Transient transfection of dCas9VP48 alone did not activate MSI1 expression (Figure 29, Lane 1) while co-transfection of dCas9VP48 with sgTetO or addition of doxycycline (positive control) activated MSI1 expression (Figure 29, Lanes 2 and 7). Neither expression of dCas9VP48 with a mutant TetO sgRNA (sgTetO-mut) carrying mismatches to the TetO binding sites (Figure 29, Lane 3) nor expression of sgTetO with dCas9 lacking an activation domain activated MSI1 expression (Figure 29, Lane 4).

To further characterize the system, HEK293T/TetO::tdTomato cells were transfected with dCas9 activator and a serial titration of sgRNAs (Figure 30). A near-linear relationship was observed between the amount of sgTetO transfected with the mean fluorescence by FACS (Figure 30B), indicating that the level of gene activation could be controlled precisely using CRISPR-on.

To test whether CRISPR-on can activate genes in vivo, dCas9VP48 plasmid, seven different sgRNAs (sgNanog-l~7) targeting the mouse Nanog promoter and a Nanog::EGFP construct containing 1kb promoter and 5' UTR of Nanog were co-injected into mouse zygotes (Figures 23C and 23D). Two days after
injection, a GFP signal was detected in 4-cell embryos by fluorescence microscopy and higher GFP expression was observed in morula and blastocyst on day 3 and day 4, whereas no GFP signal was observed in control embryos without the sgRNAs. This indicates that dCas9VP48/sgNanogs can specifically activate the GFP transgene in mouse embryos.

Activation of endogenous genes

Having established that the CRISPR-on system can activate reporter transgenes, sgRNAs targeting the endogenous human IL1RN gene were designed and tested their transactivation activity in HEK293T cells. To identify the binding sites most efficient for gene induction, six sgRNAs were designed to span the lkb IL1RN promoter (Figures 31A-31B). Initially, dCas9VP48 was transfected with all 6 sgRNAs, but failed to induce IL1RN gene expression (Figures 31A-31B). To test whether a stronger activation domain can activate IL1RN, a VP16 domain containing 10 tandem copies of VP16 motifs was fused with dCas9 to generate dCas9VP160 (Figure 24A). When co-transfected with multiple but not single sgRNAs, dCas9VP160 readily activated IL1RN (Figures 24B and 24C).

Transduction of three proximal sgRNAs (sgIL1RN1-3) activated IL1RN approximately 6 fold, whereas the three distal sgRNAs (sgIL1RN4-6) did not induce robust induction. Addition of sgRNA4-6 to the proximal sgRNAs (sgIL1RN1-3) did not significantly augment the expression (Figure 24C). These data suggest that gene activation is synergistically promoted by multiple dCas9VP160/sgRNA binding events at the proximal region of the IL1RN promoter.

A similar result was obtained with 10 sgRNAs spanning the SOX2 promoter (Figures 24D and 24E). As for IL1RN, expression of single sgRNAs did not yield strong activation of SOX2, while the triple sgRNAs (3-5, 4-6, 5-7, 8-10) activated SOX2 more than 4 fold. Seven fold activation was achieved with sgSOX2-4-6 and sgSOX2-5-7, while further distal sgRNAs (sgSOX2-8-10) or those downstream of TSS (sgSOX2-1-2) were less potent. Quintuple sgSOX2-1-5 had a lower activity than triple sgSOX2 3-5 suggesting that sgRNAs downstream of TSS (sgSOX2-1-2) may be detrimental to activation. Binding of dCas9VP160 downstream of transcriptional start sites may sterically hinder transcription by blocking polymerase, consistent with a previously report on CRISPRi (Qi et al, Cell; 152 (5):1 173-1 183 (2013)). To further confirm this observation, six sgRNAs
were designed spanning Oct4 promoter, including two targeting downstream of TSS (sgOCT4-l~2) (Figure 24F). An eight fold activation was achieved with sgOCT4-3~6, albeit all six sgOCT4-l~6 had a much lower activity than sgOCT4-3~6 confirming that sgRNAs downstream of TSS (sgSOX2-l~2) have a negative effect on gene activation (Figure 24G). Therefore, in IL1RN, SOX2, and OCT4 promoters, three to five dCas9VP160/sgRNAs binding within 300bp upstream TSS induced the most efficient gene activation.

Multiple exogenous and endogenous genes can be simultaneously activated by CRISPR-on

Single, double and triple activation of a TetO::tdTomato transgene and the endogenous SOX2 and IL1RN genes (Figure 3A) were tested in HEK293T cells carrying the stably integrated TetO::tdTomato transgene (HEK293T/TetO::tdTomato). Transfection of sgRNAs targeting the individual promoters (sgTetO for TetO::tdTomato, sgSOX2-l~10 for SOX2 or sgIL1RNl~6 for IL1RN) activated the respective genes (TetO: 6.6x; SOX2: 3.5x; IL1RN: 10.7x) while not affecting expression of the other two genes (Figure 25A). Simultaneous transfection of sgRNAs targeting two or three promoters activated the corresponding sets of genes (Figure 25A).

To test whether the system allows the activation of three different endogenous genes in a dose dependent manner, HEK293T cells were co-transfected with dCas9VP160 and the most efficient sgRNAs targeting all three genes (sgIL1RNl~3 for IL1RN, sgSOX2-5~7 for targeting SOX2, and sgOCT4-l~3 for OCT4) in different ratios (Figure 25B). When sgRNAs targeting one gene, or two genes were used, only the respective genes were activated. When all sgRNAs targeting three genes were transfected, albeit in different ratios, robust activation of all three genes was observed (Figure 25B). More significantly, when different ratios of sgRNAs were used targeting SOX2 and IL1RN while maintaining the OCT4 sgRNAs constant, the predicted change of the ratio of SOX2 and IL1RN expression levels, and the OCT4 expression remained stable (Figure 25B). These results demonstrate that the CRISPR-on system can be robustly used for multiplexed activation of endogenous genes.

CRISPR-on is highly specific
To test the specificity of CRISPR-on-mediated gene activation, microarray experiments were conducted to compare genome-wide gene expression profiles of cells transfected with dCas9VP160 and specific sgRNAs to cells transfected with dCas9VP160 and sgTetO-mut control sgRNA (Figures 26A-26D). While efficiently activating target genes, CRISPR-on did not cause major perturbations in the transcriptome (Figures 26A and 26B) as only three genes showed an over two fold up regulation upon transduction of dCas9VP160/sgTetO (Figure 26C). While CRISPR-on mediated activation of IL1RN induced the IL1RN target gene 13 fold, only 16 other genes showed an about twofold increase in expression (Figure 26D). The minor upregulation of these genes may be secondary due to the over-expression of tdTomato or IL1RN.

Discussion

Artificial transcription factors (ATFs) are valuable tools for studying gene functions and transcriptional networks. Zinc-fingers and TALE transcription factors have been developed over the recent decades and show promises in both bioengineering and therapeutic applications (Sera T., Adv Drug Deliv Rev; 61 (7-8):513-526 (2009); Perez-Pinera et al, Nat Methods; 10 (3):239-242 (2013); Maeder et al, Nature methods 2013; 10 (3):243-245 (2013)). Here, CRISPR-on was established as a novel class of artificial transcription factors based on the CRISPR/Cas system. A major advantage of this system is that only one Cas9 protein is required to activate multiple genes individually or simultaneously and that its DNA binding specificity is determined by sgRNAs, which are designed based on simple RNA/DNA complementarity.

Using CRISPR-on, robust activation was demonstrated of exogenous reporter genes in both human and mouse transformed cells as well as in ES cells. When the system was introduced into one-cell mouse embryos, efficient reporter gene activation occurred. This system can be used to manipulate transcriptional networks in early embryos.

Robust endogenous gene activation was achieved using the stronger activation domain VP160. Further optimization of activation domains, such as using different linker sequences, may improve the CRISPR-on activation efficiency even further. The promoter scanning experiments demonstrated that efficient activation of endogenous genes could be achieved by three to five sgRNAs binding within 300bp
region upstream of transcription start sites. Using additional sgRNAs targeting further upstream or downstream regions did not significantly improve the level of induction. This data suggest that only a small number of sgRNAs targeting the proximal promoter are sufficient to activate endogenous genes.

It is shown here that the CRISPR-on system can be used for the simultaneous induction of at least three different endogenous genes. More significantly, the stoichiometry of gene induction of multiple genes can be tuned by adjusting the relative amount of their cognate sgRNAs. Simultaneous activation of multiple endogenous genes with defined stoichiometry opens up novel opportunities for systems biology as it allows for the predictable manipulation of transcriptional networks.

Finally, with the ease of design and synthesis, a library of sgRNAs could be generated. When introduced into a cell line constitutively expressing dCas9 protein, gene activation screens mediated by RNA (RNAa) could be achieved. Since the specificity components (sgRNA) can be separately designed and constructed from the effector component (Cas protein), the same library of sgRNAs could be used with different dCas9 fusions (e.g., VP160 domain for transactivation, KRAB domain for transcriptional repression, chromatin modifier domains for specific histone modification) to exert different functions at particular genomic loci.

CRISPR dCas9 fusion peptides sequences

dCas9VP160-2A-puro

MYPYDVPDYASPKKKKRKEASDKKYSIGLAIGTNWSVGAVITDEYKVSIFKFKVLGNTIDRHSLIKKLIGALLDGETAEATRLKRTARRYTRRKNRI CYLQEIFSNEAKVDSSFHRLEESFLVEEDKHERHPIFNGNIVDEYHERKPYTIYHLRKRLVDSTDGLRLYLIYALAHMIKFRGHFLIEGLDPDNVDKLFIQLOQYNOLFEENPINASVDAKALSLRKSURLENHNQALPGEJKNLIFGLNLALSLGLTPFNSNFLDAEKALSQLSDYDDDLNLLAQIGDQYADLEAAKLNSDALLDLRILVRNTETKAPLRSAMIKRDHEHDQDILLKALVRQQPAQYKEIFDDQSKNGASYIDGASQEEFYKFIPILEKMDGTEELVLKINREDLLR KORTFDSNGISPHQHGLHELHAIRQRQEDFYFPLKDHREIKELFRIPYYVGPLARGNSRFAWMTKREEETTPWNEEVVDKGSAQSFIERMTFNDKNLNPENKVPKSLLELYFTVKYELTKYVRGMKQKLRQUGRLRSLRKLINGIRDQSKSGKTLDLFSGDFANRNFMQLHDDSLTFTKEDIQKAYQSSQGDSLHEHIANLAPAIKAGILQTVKVDELVKVMRHPKENV IEMAREQTTQK贤NSRERMKREEGIQELGSQILKEHVPVENTQLANEKLLYLYQNGR
dCas9p65

MYPYDVPDYASPKKKVEASDSSYSLAIAGTNSVQGWAVITDEYKVPSKKFKVLGNTD

PSIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLATDLIIQSI 479

[SEQ ID NO: 479]

[SEQ ID NO: 480]
dCas9KRAB

\[00414\]

MYPYDVPDYASPKKRKVEASKKYSIGLAIGTNSVGAVITDEYKVPSSFKFKVLGNTD

RHSIKKNIAGALLFDSETAEATRLKRTARRYTRRRKNI CYLQE IFSNEMAKVDSFFHR

LEESFLVEDDKHERRIPFIGNIDVEAVYHEKYTIYHLRRKLVSTDKADILRILYALAH

MIKFRGHFLIEGDLNPNDSDVCKLFQLQVQYQNFILLENPRONASVGDAKAILSARLKSXR

RLENLIAOLPGEKKNLGNLISGLLTGTPFNSFDFAESDAKLQLSKDITYDDLNLNA

QIGDQYADLFLAANLSDAILLSDIDRNVTENITKAPLSAMIKRYDEHQLDLTILLKALVR

QOLPEKYKEIFPSKGNYNGYIDGASQEEFYKFKPIFLKIDMGTEELVKNREDLRL

KQRTFDNGSIPHQIHLGELHAIIRQQEDYFYPKLDNREIKEKILTFRIPYYVPGAPRNS

RFAMTRKSKSEITTPNWFEDEVKDGASASSFIERMTFDSNKLPNEKVKFLSHLLYEFYTV

YNELTKVYVTGMRKPAFLSAGEQKKAIVDLDFTKRKNKTVQLKDEYFKKIECFDSVEI

SGVEDRFNASLGTYHDILKIKDKFDLDNEENDEDIDIVLTLLFDREMIIEERLKYTA

HLFDKVKMKQLKRRTYGTWGRLSRKLINGRIDDRQSGKTLIDFLKSDGFANRNFMQLIHD

SLTFKEDIQKAQVSQGDSLHEHIANLAPSKAIKIGILQTQKVVDLESVFKMGRHFPENIV

IEMARENQTQKISQKNSRMRKRIEQIKELGSDLKHEPVENTQLQNEKLKYLYLQNGR

DMYVQYNQLINRLSDYDVAIPQPSLKDSIDNKSLOTNSRGNKSNDMVPEEYVMXMK

NYWRQQLNKLITQRKDNLTKAERGGLSELDKAGFKRQLVETRQIKHQAQILDSRMN

TKYDENDKLIREVKVIKTLKSVDSRKFDFQYKVRKIVEINNYHAHAHDAYLNAVGTALIKK

YPKLESEFVYGDKYVDRMKIAKSEQEIQGKATAKYFFYSNIMNFKTEITLANGEIRKR

PIETNGETGEIVWDKGREDATVRFKXSMPQVINVKKTEQGFSKESILPKRNSDKLIA

ARKKDWDPKYGFFDSTPVAVLVAVKVEKGSKKLSVKELLGITIMERSSSFKNPID

FLEAKGYKVEKDLIKLFLKPKYSLVELENGRKRMLASAGELQKGNELALPSKYNFLYLAS

HYEKLKSPDENQEQKFQVFQKHKYLDIEI EQIIFSKRVILADANKVLSAYNKRHDKP

PIREQAGNEI HLFTLNLGAPAAMKYAYDFTIDDRKRTSTKEVLDATLHQSITGLEYTRI

DLSQLGDSPKKRKVEASGPAAASPKKRKVEASMDAKSLTAWSRITLTFKDFVDFPTE

EWKLTLDTAQQIVYRNMLENKLNVLGQLQGTKPDVILRLEEGEPWLVSRSID

(SEQ ID NO: 481)
[00415]  dCas9PCP

[00416]  MYPYDVPDYASPKKKRKEASEDKKYSIGLAIIGNVSQGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNGALLFDGSETAAEATRLKRATKARRRTYRKRKKRI CYLQEIFSNEMAKVDDSFTHR
LEESFLVEEKHHERHPFIGNIVDEVAYHEKPYTIYHLRKLVLVDSTKDMLRLIYLALAH
MIKFGHFLIEGLDPNDSSVDFKLIQLVQTQYNLFEENPINASGDVAAILSLAROKSR
RLENLIAPLPGEKKNLGFNLASLGLTQPFINSNLDEVAKLQLSKDTYDDLDDLNLALL
QIGDQYADFLIAAKNLSDAILSLDILRVNTEIKApLASMISIKRYDEHQDLTQLCLVLVR
QQPEKYKEIFDDQSKNGAYGIDGASQEEFYFKIPKelMDTVEILVVKNLDLLL
KRTRFDNGSIPHIILGELHALIRQSFEDFYPLKDNREIKEIILFTRPIYPYVPPLARGNS
RFAMTRKSEETITPWNFEEVVDKGAASQFSIERMTNFLNPEKVLPEHSLYYEFVTV
YNELETKVYTEGMRKPAFLSEPQGDSQDLHEHIANLAGSPIKGIGQTVKVDLVEKVRHKPN
IEMARENQTQKQGKSNREKMKREEIKELGSQILKEHPEVENTQLNEKLYLYQNGR
DVMYQDELNRLSDFDVSQFSLKDDSNKVTRESKKNRGSKDNEPVSEEVVKKMK
NYWRQLNKLITQRKFDNLTKAERGLSLEDKAFGIRQLVETRQITKHVAILDSRMN
TKYENDKLIREVKVTILSKLSDFKRDFQFQYKREVINYHHAHAHYDAYLANVGTALK
YPKLESEFYVGYDKYDVRMKIAKSEQEIGKATKYFFSSNIMFKEITEILANGEIRK
PLIETNGTGEYWDGGDFATVRKNSPQVNIKVKEVQTFGGFSKELSIRPKRNKLI
ARKKDWDPPKYGFFSDTVPAYSVLVAVKEKGSKKLSKVSSEELTIMERSSEFEKNPID
FLEAKGYKEVKKDLI IKLKPYSFELRNGRKLMSALAGEQQGKRNELALPSKYYTVNLYAS
HYEKLKGSPEEDNEQKQLFVEQKHYYLEI IEIQIEFSKRVLADANLKLVSAYNKRHDK
PIREQAIENHLFTILNLGAPAFAKYFDTIDDIDKRYSTKEVDAITALIGQSTLCFVETRI
DSLQLGGDSPPKKKRKVEASEGPAGIDMSKTIVLSVEATRTLIEIQSTADRQIFEKEGVPLV
GRLRLASLRQNGAKTAYRVNLKDQADVVDCTSCVGEIPLKVRMTQWSHDVITVANST
EASRKSLYDTLSKVLATSQVEDLVNVLPLGR  (SEQ ID NO: 482)

[00417]  dCas9MS2

[00418]  MYPYDVPDYASPKKKRKEASEDKKYSIGLAIIGNVSQGWAVITDEYKVPSKKFKVLGNTD
RHSIKKNGALLFDGSETAAEATRLKRATKARRRTYRKRKKRI CYLQEIFSNEMAKVDDSFTHR
LEESFLVEEKHHERHPFIGNIVDEVAYHEKPYTIYHLRKLVLVDSTKDMLRLIYLALAH
MIKFGHFLIEGLDPNDSSVDFKLIQLVQTQYNLFEENPINASGDVAAILSLAROKSR
RLENLIAPLPGEKKNLGFNLASLGLTQPFINSNLDEVAKLQLSKDTYDDLDDLNLALL
QIGDQYADFLIAAKNLSDAILSLDILRVNTEIKApLASMISIKRYDEHQDLTQLCLVLVR
QQPEKYKEIFDDQSKNGAYGIDGASQEEFYFKIPKelMDTVEILVVKNLDLLL
KRTRFDNGSIPHIILGELHALIRQSFEDFYPLKDNREIKEIILFTRPIYPYVPPLARGNS
RFAMTRKSEETITPWNFEEVVDKGAASQFSIERMTNFLNPEKVLPEHSLYYEFVTV
YNELETKVYTEGMRKPAFLSEPQGDSQDLHEHIANLAGSPIKGIGQTVKVDLVEKVRHKPN
IEMARENQTQKQGKSNREKMKREEIKELGSQILKEHPEVENTQLNEKLYLYQNGR
DVMYQDELNRLSDFDVSQFSLKDDSNKVTRESKKNRGSKDNEPVSEEVVKKMK
NYWRQLNKLITQRKFDNLTKAERGLSLEDKAFGIRQLVETRQITKHVAILDSRMN
TKYENDKLIREVKVTILSKLSDFKRDFQFQYKREVINYHHAHAHYDAYLANVGTALK
YPKLESEFYVGYDKYDVRMKIAKSEQEIGKATKYFFSSNIMFKEITEILANGEIRK
PLIETNGTGEYWDGGDFATVRKNSPQVNIKVKEVQTFGGFSKELSIRPKRNKLI
ARKKDWDPPKYGFFSDTVPAYSVLVAVKEKGSKKLSKVSSEELTIMERSSEFEKNPID
FLEAKGYKEVKKDLI IKLKPYSFELRNGRKLMSALAGEQQGKRNELALPSKYYTVNLYAS
HYEKLKGSPEEDNEQKQLFVEQKHYYLEI IEIQIEFSKRVLADANLKLVSAYNKRHDK
PIREQAIENHLFTILNLGAPAFAKYFDTIDDIDKRYSTKEVDAITALIGQSTLCFVETRI
DSLQLGGDSPPKKKRKVEASEGPAGIDMSKTIVLSVEATRTLIEIQSTADRQIFEKEGVPLV
GRLRLASLRQNGAKTAYRVNLKDQADVVDCTSCVGEIPLKVRMTQWSHDVITVANST
EASRKSLYDTLSKVLATSQVEDLVNVLPLGR  (SEQ ID NO: 482)
[00419] dCas9VP160ER

[00420] MYPYDVPDYAPKKKRKEASDGYKSLGIGAILAGINSVGSWAIVITEKVPSKFIKVLGLNIGHTR
RHSKKNLIGALLFDSGETEAATRLKRTARRRTYTRRNRI CYLQEIFSNEMAKVDSSFFHR
LEESLIVEEDKHERPIFIGNVEDAYHEKYPIHYLRKIKVSTDKADLRIYLAH
MIFRGRHHFIELGDNLPSDVSVDKLFLQVLQTYNLFEENINASVCAKILARLSKSR
RENIAQILPGKEKGNLFLIALSLGLTPNFSKNDLAEADKQLSKDYDDLLNLQL
QIGDQYADLFLAAKNSDAILLSDLVRNEITKAPLSAMKKYDEH7QDLTLALKVALVR
QOLPERYKEIFFDSQKNG VyGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMFFKETEELDLR
KQRTFINDGSHIPQHLHGLHELARLRQEDFYPFGLDNREKIEIKLTRIFYVGPLARGNS
RFAMTRKRESETITPNWVFVGDASQASFFERIMTFDKNLPEKIKSHLLLYEFTV
YNELTKYVTEGMXKPAFLSGQKKAIVDLFIKTVKTVKLQKILQYF KKFieCFDSVEI
SGVDELNASGTVHYDILKK I KDDFDN ENEINED ILEDIVLTLTLEDREMIEREELRTYA
HLFDDKMQLKRRRRTYGWRGLSRKINGIRDQOSGTKTILDLFKSDGFANRFMQLIHHDD
SITFKEDIQKAVGQGDSLHELHIANLAGSPAIKMGILQTVKVDILVKVRGKHKPENIV
IEMARENQTQKQKNSRERMMKRIEESIKELGQSILKEHVQVQTNQNEKLYLYLYLQNGR
DMYQVELDINLRSLDYDVAIFQSPFLKLOSIDNVKLRSDKNRMGSKDNVPSXGVKMK
NYWQRLILAKLTQFRDFNLTAEGRGSELLDKAGTIRQLVETRQILTKVQAIILDRSMN
T IHttpKVRKILREVKTVLKLVLSDVSDFRDFQFYKVRLEINVNYHHAHAYLNAGVTALMK
YPKLESEFYGKDYKVQVRMIAKSEQEIGKATAKYFFYSNIMFFKETEELDLR
PILQEDGETGEIVWDGKRDGAFTRKVLSPMQVNIKKEVQTGFFKESILPKRNSDKLI
ARKKWDPPKYGGFDSPTVAYSVLVAKEVGKSSLKSVKELGITTERRSFEFKNID
FLEAKGYKVEKDLKI I KLPKYSLFLEUENGRMLASAGELQKVNLPSKYNVNYLAILAS
HYEKLKGSQDENQKLFVFQHEKKHYLDEI EQISEFSKRVLADANLDKLVSAYNHKHDR
PIREQAENI IHLFTLTNLGAPAAFKYFDTTIDDRKRYSTKEVLDAATLHQSITGLEYTR

[SEQ ID NO: 483]
The teachings of all patents, published applications and references cited herein are incorporated by reference in their entirety. While this invention has been particularly shown and described with references to example embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.
What is claimed is:

1. A method of mutating one or more target nucleic acid sequences in a stem cell or zygote comprising:
   (a) introducing into the stem cell or zygote
      (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associated (Cas) protein; and
      ii) a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity; and
   (b) maintaining the cell or zygote under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences, and the Cas protein cleaves each of the one or more target nucleic acid sequences upon hybridization of the one or more RNA sequences to the portion of the target nucleic acid sequence; thereby mutating one or more target nucleic acid sequences in the stem cell or zygote.

2. The method of claim 1 wherein the Cas protein is Cas9.

3. The method of any one of the preceding claims wherein the Cas protein cleaves one strand or both strands of one or more of the target nucleic acid sequences.

4. The method of any one of the preceding claims wherein the Cas protein nicks one strand or both strands of one or more of the target nucleic acid sequences.

5. The method of any one of the preceding claims wherein the RNA sequence is from about 10 base pairs to about 150 base pairs in length.
6. The method of any one of the preceding claims wherein the stem cell or zygote is a rodent stem cell or zygote, a primate stem cell or zygote, a canine stem cell or zygote, a feline stem cell or zygote, a bovine stem cell or zygote, an equine stem cell or zygote, a caprine stem cell or zygote, a porcine stem cell or zygote, or an avian stem cell or zygote.

7. The method of claim 6 wherein the rodent stem cell or zygote is a mouse stem cell or zygote.

8. The method of claim 6 wherein the primate stem cell or zygote is a human stem cell or zygote.

9. The method of any one of the preceding claims wherein the stem cell is an embryonic stem cell or an induced pluripotent stem cell.

10. The method of any one of the preceding claims wherein one or more of the target nucleic acid sequences are associated with a disease or condition.

11. The method of any one of the preceding claims wherein one or more of the target nucleic sequences are a gene.

12. The method of any one of the preceding claims wherein both copies of one or more of the target nucleic acid sequences in the stem cell or zygote are mutated.

13. The method of any one of the preceding claims wherein the one or more target nucleic acid sequences are endogenous to the stem cell or zygote.

14. The method of any one of the preceding claims wherein one or more of the target nucleic acid sequences encode a transcription factor, a transcriptional co-activator or co-repressor, an enzyme, a chaperone, a heat shock factor, a heat shock protein, a receptor, a secreted protein, a transmembrane protein, a histone, a peripheral membrane protein, a soluble protein, a nuclear protein,
a mitochondrial protein, a growth factor, a cytokine, an interferon, a chemokine, a hormone, an extracellular matrix protein, a motor protein, a cell adhesion molecule, a major or minor histocompatibility (MHC) gene, a transporter, a channel an immunoglobulin (Ig) superfamily (IgSF) gene, a tumor necrosis factor, an NF-kappaB protein, an integrin, a cadherin superfamily member, a selectin, a clotting factor, a complement factor, a plasminogen, plasminogen activating factor, a proto-oncogene, oncogene, or tumor suppressor gene.

15. The method of any one of the preceding claims further comprising introducing one or more nucleic acid sequences that are complementary to a portion of the one or more target nucleic acid sequences cleaved by the Cas protein.

16. The method of claim 15 wherein the one or more nucleic acid sequences are a single stranded oligonucleotide, a double stranded oligonucleotide, a plasmid, a cDNA, a gene block or a PCR product.

17. The method of claim 15 or 16 wherein the one or more nucleic acid sequences replace one or more nucleotides, introduce one or more additional nucleotides, delete one or more nucleotides or a combination thereof in the one or more target nucleic acid sequences.

18. The method of claim 15, 16, or 17 wherein the one or more nucleic acid sequences introduce a point mutation in one or more of the target sequences.

19. The method of claim 15, 16, or 17 wherein the one or more nucleic acid sequences replace one or more mutant nucleotides with one or more wild type nucleotides in one or more of the target sequences.

20. The method of claim 15, 16, 17, 18 or 19 wherein the nucleic acid sequences is from about 10 nucleotides to about 1000 nucleotides.
21. The method of any one of the preceding claims further comprising introducing the stem cell or zygote into a nonhuman mammal.

22. The method of claim 21 wherein the nonhuman mammal is of the same species as the stem cell or zygote.

23. The method of claim 21 wherein the nonhuman mammal is a mouse.

24. The method of any one of the preceding claims further comprising assessing whether the one or more of the target nucleic acid sequences have been mutated.

25. The method of any one of the preceding claims further comprising isolating the stem cell or zygote.

26. The method of any one of the preceding claims wherein 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 target nucleic acids sequences are mutated in the stem cell or zygote.

27. The method of any one of the preceding claims wherein at least two of the target nucleic acid sequences are endogenous nucleic acid sequences.

28. The method of any one of the preceding claims wherein at least two of the target nucleic acid sequences are endogenous genes.

29. The method of any one of the preceding claims wherein at least two of the target nucleic acid sequences are at least 1kB apart.

30. The method of any one of the preceding claims wherein at least two of the target nucleic acid sequences are on different chromosomes.
31. The method of any one of the preceding claims wherein at least one mutation comprises an insertion of a tag, a transgene, or an insertion of a site recognized by a recombinase.

32. The method of any one of the preceding claims wherein at least one mutation renders expression of an endogenous gene conditional.

33. The method of any one of the preceding claims wherein at least one mutation renders expression of an endogenous gene inducible, repressible, or tissue-specific.

34. The method of any one of the preceding claims wherein the mutations comprise inserting recombination sites (e.g., loxP sites or FRT sites) flanking a selected genomic region, wherein the selected genomic region is optionally within a gene.

35. The method of any one of the preceding claims wherein the mutations comprise inserting a recombination-site-STOP-recombination site cassette (e.g., a loxP-STOP-loxP or FRT-STOP-FRT cassette) in a gene, between a promoter and a coding region of a gene, or in a regulatory region of a gene.

36. The method of claim 35, wherein the recombination-site-STOP-recombination site cassette is positioned so as to disrupt expression of the gene and wherein excision of the cassette by a recombinase renders the gene expressible.

37. An isolated mutated stem cell or zygote produced by the method of any one of the preceding claims.

38. A method of producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences comprising:
   (a) introducing into a zygote or an embryo
(i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to a portion of each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR-associated (Cas) protein; and

ii) a Cas nucleic acid sequence or a variant thereof that encodes a Cas protein having nuclease activity; and

(b) maintaining the zygote or the embryo under conditions in which RNA hybridizes to the portion of each of the one or more target nucleic acid sequences, and the Cas protein cleaves each of the one or more target nucleic acid sequences upon hybridization of the RNA to the portion of the target nucleic acid sequence, thereby producing an embryo having one or more mutated nucleic acid sequences;

(c) introducing the embryo having one or more mutated nucleic acid sequences into a foster nonhuman mammalian mother; and

(d) maintaining the foster nonhuman mammalian mother under conditions in which one or more offspring carrying the one or more mutated nucleic acid sequences are produced, thereby producing a nonhuman mammal carrying mutations in one or more target nucleic acid sequences.

39. The method of claim 38 wherein the Cas protein is Cas9.

40. The method of claim 38 or 39 wherein the Cas protein cleaves one strand or both strands of one or more of the target nucleic acid sequences.

41. The method of any one of claims 38-40 wherein the Cas protein nicks one strand or both strands of one or more of the target nucleic acid sequences.

42. The method of any one of claims 38-41 wherein the RNA sequence is from about 10 base pairs to about 150 base pairs in length.
43. The method of any one of claims 38-42 wherein nonhuman mammal is a rodent, a nonhuman primate, a canine, a feline, a bovine, a porcine, an equine, or a caprine.

44. The method of claim 43 wherein the rodent is a mouse.

45. The method of any one of claims 38-44 wherein one or more of the target nucleic acid sequences are associated with a disease or condition.

46. The method of any one of claims 38-45 wherein one or more of the target nucleic sequences are a gene.

47. The method of any one of claims 38-46 wherein both copies of one or more of the target nucleic acid sequences in the zygote or the embryo are mutated.

48. The method of any one of claims 38-47 wherein the one or more target nucleic acid sequences are endogenous to the zygote or the embryo.

49. The method of any one of claims 38-48 wherein one or more of the target nucleic acid sequences encode a transcription factor, a transcriptional co-activator or co-repressor, an enzyme, a chaperone, a heat shock factor, a heat shock protein, a receptor, a secreted protein, a transmembrane protein, a histone, a peripheral membrane protein, a soluble protein, a nuclear protein, a mitochondrial protein, a growth factor, a cytokine, an interferon, a chemokine, a hormone, an extracellular matrix protein, a motor protein, a cell adhesion molecule, a major or minor histocompatibility (MHC) gene, a transporter, a channel an immunoglobulin (Ig) superfamily (IgSF) gene, a tumor necrosis factor, an NF-kappaB protein, an integrin, a cadherin superfamily member, a selectin, a clotting factor, a complement factor, a plasminogen, plasminogen activating factor, a proto-oncogene, oncogene, or tumor suppressor gene.
50. The method of any one of claims 38-49 further comprising introducing into the zygote or the embryo one or more nucleic acid sequences that are complementary to a portion of the one or more target nucleic acid sequences cleaved by the Cas protein.

51. The method of claim 50 wherein the one or more nucleic acid sequences are a single stranded DNA oligonucleotide, a double stranded DNA oligonucleotide, a plasmid, a cDNA, a gene block or a PCR product.

52. The method of claim 50 or 51 wherein the one or more nucleic acid sequences replace one or more nucleotides, introduce one or more additional nucleotides, delete one or more nucleotides or a combination thereof in the one or more target nucleic acid sequences.

53. The method of claim 50, 51, or 52 wherein the one or more nucleic acid sequences introduce a point mutation in the target nucleic acid sequence.

54. The method of claim 50, 51, or 52 wherein the one or more nucleic acid sequences replace one or more mutant nucleotides with one or more wild type nucleotides in the target nucleic acid sequence.

55. The method of claim 50, 51, 52, 53, or 54 wherein the one or more nucleic acid sequences is from about 10 nucleotides to about 1000 nucleotides.

56. The method of any one of claims 38-55 wherein the nonhuman mammal is of the same species as the embryo or zygote.

57. The method of any one of claims 38-56 wherein the nonhuman mammal is a mouse.

58. The method of any one of claims 38-57 further comprising assessing whether the one or more target nucleic acid sequences have been mutated.
59. The method of any one of claims 31-50 wherein 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 target nucleic acids sequences are mutated in the nonhuman mammal.

60. The method of any one of the claims 38-59 wherein at least two of the target nucleic acid sequences are endogenous nucleic acid sequences.

61. The method of any one of claims 38-60 wherein at least two of the target nucleic acid sequences are endogenous genes.

62. The method of any one of claims 38-61 wherein at least two of the target nucleic acid sequences are at least 1kB apart.

63. The method of any one of claims 38-62 wherein at least two of the target nucleic acid sequences are on different chromosomes.

64. The method of any one of the preceding claims wherein at least one mutation comprises an insertion of a tag, a transgene, or an insertion of a site recognized by a recombinase.

65. The method of any one of the preceding claims wherein at least one mutation renders expression of an endogenous gene conditional.

66. The method of any one of the preceding claims wherein at least one mutation renders expression of an endogenous gene inducible, repressible, or tissue-specific.

67. The method of any one of the preceding claims wherein the mutations comprise inserting recombination sites (e.g., loxP sites or FRT sites) flanking a selected genomic region, wherein the selected genomic region is optionally within a gene.
68. The method of any one of the preceding claims wherein the mutations comprise inserting a recombination-site-STOP-recombination site cassette (e.g., a loxP-STOP-loxP or FRT-STOP-FRT cassette) in a gene, between a promoter and a coding region of a gene, or in a regulatory region of a gene.

69. The method of claim 35, wherein the recombination-site-STOP-recombination site cassette is positioned so as to disrupt expression of the gene and wherein excision of the cassette by a recombinase renders the gene expressible.

70. A non-human mammal produced by the method of any one of claims 31-55.

71. A method of modulating the expression and/or activity of one or more target nucleic acid sequences in a cell comprising:
   (a) introducing into the cell
      (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associate (Cas) protein;
      (ii) a Cas nucleic acid sequence or a variant thereof that encodes the Cas protein that targets but does not cleave the target nucleic acid sequence; and
      (iii) an effector domain;
   (b) maintaining the cell under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences, the Cas protein binds to each of the one or more RNA sequences and the effector domain modulates the expression and/or activity of the target nucleic acid,

   thereby modulating the expression and/or activity of one or more target nucleic acid sequences in the cell.

72. The method of claim 71, wherein the Cas protein and effector domain are fused creating a chimeric protein.
73. The method of any one of claims 71-72, wherein the one or more RNA sequences are complementary to all or a portion of a regulatory region of the one or more target nucleic acid sequences.

74. The method of claim 73, wherein the regulatory region of the one or more target nucleic acid sequences is a promoter, enhancer, and/or operator.

75. The method of any one of claims 71-74, wherein the Cas protein is Cas9.

76. The method of claim 75, wherein the Cas9 protein comprises one or more mutations.

77. The method of claim 76, wherein the Cas9 protein comprises a mutation at amino acid position 10, 840, or a combination thereof.

78. The method of claim 77, wherein the amino acid at position 10 is mutated from aspartate (D) to alanine (A) and the mutation at amino acid position 840 is mutated from histidine (H) to alanine (A).

79. The method of any one of claims 71-78, wherein the effector domain modulates the expression and/or activity of the one or more target nucleic acid sequences by activating, coactivating, regulating, repressing, organizing, remodeling, modifying, and/or fusing the expression and/or activity of one or more target nucleic acid sequences.

80. The method of any one of claims 71-79, wherein the effector domain is a transcription activating domain, a coactivator domain, a transcriptional pause release factor domain, a negative regulator of transcriptional elongation domain, a transcriptional repressor domain, a chromatin organizer domain, a remodeler domain, a histone modifier domain, a DNA modification domain, a RNA binding domain, a protein interaction input devices domain, or a protein interaction output device domain.
81. The method of any one of claims 71-80, wherein the effector domain is VP16, VP48, VP64, VP96, or VP160.

82. The method of any one of claims 71-81, further comprising introducing to the cell an effector molecule that activates the effector domain.

83. The method of claim 82, wherein the effector molecule binds to the effector domain.

84. The method of any one of claims 82-83, wherein the effector molecule is an antibiotic.

85. The method of any one of claims 71-84, wherein the RNA sequence is from about 10 base pairs to about 150 base pairs in length.

86. The method of any one of claims 71-85 wherein the cell is a rodent cell, a primate cell, a canine cell, a feline cell, a bovine cell, an equine cell, a caprine cell, a porcine cell, or an avian cell.

87. The method of claim 86 wherein the rodent cell is a mouse cell.

88. The method of claim 86 wherein the primate cell is a human cell.

89. The method of any one of claims 71-88 wherein the cell is a stem cell or zygote.

90. The method of any one of claims 71-89 wherein the stem cell is an embryonic stem cell or an induced pluripotent stem cell.

91. The method of any one of claims 71-90 wherein one or more of the target nucleic acid sequences are associated with a disease or condition.
92. The method of any one of claims 71-91 wherein one or more of the target nucleic acid sequences are a gene.

93. The method of any one of claims 71-92 wherein both copies of one or more of the target nucleic acid sequences in the cell are mutated.

94. The method of any one of claims 71-93 wherein the one or more target nucleic acid sequences are endogenous or exogenous to the cell.

95. The method of any one of claims 71-94 wherein one or more of the target nucleic acid sequences encode a mutated protein, a reprogramming factor, a transcription factor, a transcriptional co-activator or co-repressor, an enzyme, a chaperone, a heat shock factor, a heat shock protein, a receptor, a secreted protein, a transmembrane protein, a histone, a peripheral membrane protein, a soluble protein, a nuclear protein, a mitochondrial protein, a growth factor, a cytokine, an interferon, a chemokine, a hormone, an extracellular matrix protein, a motor protein, a cell adhesion molecule, a major or minor histocompatibility (MHC) gene, a transporter, a channel an immunoglobulin (Ig) superfamily (IgSF) gene, a tumor necrosis factor, an NF-kappaB protein, an integrin, a cadherin superfamily member, a selectin, a clotting factor, a complement factor, a plasminogen, plasminogen activating factor, a proto-oncogene, oncogene, or tumor suppressor gene.

96. The method of claim 95, wherein the transcription factor is Oct4, Sox2, Klf4, or c-Myc.

97. The method of any one of claims 71-96 further comprising introducing the cell into a nonhuman mammal.

98. The method of claim 97 wherein the nonhuman mammal is of the same species as the cell.
99. The method of any one of claims 97 or 98 wherein the nonhuman mammal is a mouse.

100. The method of any one of claims 71-99 further comprising isolating the cell.

101. The method of any one of claims 71-100 wherein at least two of the target nucleic acid sequences are endogenous or exogenous nucleic acid sequences.

102. The method of any one of claims 71-101 wherein at least two of the target nucleic acid sequences are endogenous or exogenous genes.

103. The method of any one of claims 71-102 wherein at least two of the target nucleic acid sequences are at least 1kB apart.

104. The method of any one of claims 71-103 wherein at least two of the target nucleic acid sequences are on different chromosomes.

105. The method of any one of claims 71-104, wherein the modulation of one or more target nucleic acid sequences comprises simultaneous activation of the one or more target nucleic acid sequences.

106. The method of any one of claims 71-105, further comprising adjusting the level of modulation of one or more target nucleic acid sequences by adjusting the amount of the one or more ribonucleic acid sequences introduced into the cell or zygote.

107. The method of any one of claims 71-106, wherein 2, 3, 4, or 5 RNA sequences which bind within 300 bases upstream to a transcription start site to each of the one or more target nucleic acid sequences are introduced into the cell or zygote.
108. The method of any one of claims 71-107, wherein modulation of the expression and/or activity of the one or more target nucleic acid sequence reprograms the cell's potency.

109. The method of claim 108, wherein a differentiated cell is reprogrammed to a pluripotent cell.

110. An isolated nucleic acid sequence that encodes a fusion protein comprising all or a portion of a Cas protein fused to all or a portion of an effector domain.

111. The isolated nucleic acid sequence of claim 110, wherein the Cas protein is Cas9.

112. The isolated nucleic acid sequence of claim 111, wherein the Cas9 protein comprises one or more mutations.

113. The isolated nucleic acid sequence of claim 112, wherein the Cas9 protein comprises a mutation at amino acid position 10, 840, or a combination thereof.

114. The isolated nucleic acid sequence of claim 113, wherein the amino acid at position 10 is mutated from aspartate (D) to alanine (A) and the mutation at amino acid position 840 is mutated from histidine (H) to alanine (A).

115. The isolated nucleic acid sequence of any one of claims 110-114, wherein the effector domain modulates the expression and/or activity of the one or more target nucleic acid sequences by activating, coactivating, regulating, repressing, organizing, remodeling, modifying, and/or fusing the expression and/or activity of one or more target nucleic acid sequences.

116. The isolated nucleic acid sequence of any one of claims 108-114, wherein the effector domain is a transcription activating domain, a coactivator
domain, a transcriptional pause release factor domain, a negative regulator of transcriptional elongation domain, a transcriptional repressor domain, a chromatin organizer domain, a remodeler domain, a histone modifier domain, a DNA modification domain, a RNA binding domain, a protein interaction input devices domain, and a protein interaction output device domain.

117. The isolated nucleic acid sequence of any one of claims 108-115, wherein the effector domain is VP16, VP48, VP64, VP96, or VP160.

118. An isolated fusion protein comprising all or a portion of a Cas protein fused to all or a portion of an effector domain.

119. The isolated fusion protein of claim 118, wherein the Cas protein is Cas9.

120. The isolated fusion protein of claim 119, wherein the Cas9 protein comprises one or more mutations.

121. The isolated fusion protein of claim 120, wherein the Cas9 protein comprises a mutation at amino acid position 10, 840, or a combination thereof.

122. The isolated fusion protein of claim 121, wherein the amino acid at position 10 is mutated from aspartate (D) to alanine (A) and the mutation at amino acid position 840 is mutated from histidine (H) to alanine (A).

123. The isolated fusion protein of any one of claims 118-122, wherein the effector domain modulates the expression and/or activity of the one or more target nucleic acid sequences by activating, coactivating, regulating, repressing, organizing, remodeling, modifying, and/or fusing the expression and/or activity of one or more target nucleic acid sequences.

124. The isolated fusion protein of any one of claims 118-123, wherein the effector domain is a transcription activating domain, a coactivator domain, a
transcriptional pause release factor domain, a negative regulator of transcriptional elongation domain, a transcriptional repressor domain, a chromatin organizer domain, a remodeler domain, a histone modifier domain, a DNA modification domain, a RNA binding domain, a protein interaction input devices domain, and a protein interaction output device domain.

125. The isolated fusion protein of any one of claims 118-124, wherein the effector domain is VP16, VP48, VP64, VP96, or VP160.

126. An isolated polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 14 (dCas9VP64).

127. An isolated polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 15 (dCas9VP96).

128. An isolated polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 16 (dCas9VP160).

129. An isolated fusion protein comprising at least six VP16 transactivation domains fused to each other.

130. The fusion protein of claim 129, wherein at least one VP16 transactivation domain is a minimal transactivation domain.

131. The fusion protein of claim 130, wherein the VP16 minimal transactivation domain comprises the amino acid sequence DALDDFDLDML.

132. The fusion protein of any one of claims 129-131, comprising 6, 7, 8, 9, or 10 tandem copies of the VP16 transactivation domains.

133. The fusion protein of any one of claims 129-132, further comprising one or more linkers.
134. A method of modulating the expression and/or activity of one or more target nucleic acid sequences that cause a disease in an individual in need thereof comprising:
   (a) introducing into the individual
      (i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associate (Cas) protein;
      (ii) a Cas nucleic acid sequence or a variant thereof that encodes the Cas protein that targets but does not cleave the target nucleic acid sequence; and
      (iii) an effector domain;
   (b) maintaining the cell under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences, the Cas protein binds to each of the one or more RNA sequences and the effector domain modulates the expression and/or activity of the target nucleic acid,
   thereby modulating the expression and/or activity of one or more target nucleic acid sequences that cause a disease in the individual.

135. The method of Claim 134 wherein the one or more target nucleic acid sequences are one or more genes.

136. The method of Claims 134 or 135 wherein the effector domain represses the activity and/or expression of the one or more target nucleic acid sequences.

137. A method of detecting a sequence variation, chromatin state or a combination thereof at a defined loci of one or more target sequences of a cell and exerting an effect on the cell upon detection of the sequence variation, chromatin state or a combination thereof at the defined loci comprising:
   (a) introducing into the cell
(i) one or more ribonucleic acid (RNA) sequences that comprise a portion that is complementary to each of the one or more target nucleic acid sequences and comprise a binding site for a CRISPR associate (Cas) protein;

(ii) a Cas nucleic acid sequence or a variant thereof that encodes the Cas protein that targets but does not cleave the target nucleic acid sequence; and

(iii) an effector domain;

(b) maintaining the cell under conditions in which the one or more RNA sequences hybridize to the portion of each of the one or more target nucleic acid sequences, the Cas protein binds to each of the one or more RNA sequences and the effector domain exerts an effect on the cell upon detection of the sequence variation, chromatin state or a combination thereof at the defined loci,

thereby detecting a sequence variation, chromatin state or a combination thereof at a defined loci of one or more target sequences of the cell and exerting an effect on the cell upon detection of the sequence variation, chromatin state or a combination thereof at the defined loci in the cell.

138. The method of Claim 137 wherein the cell is in an individual in need thereof.

139. The method of Claim 137 or 138 wherein the effector domain is one or more caspases.

140. The method of any one of Claims 137-139 wherein the methylation, chromosomal translocation, or a combination thereof is detected at the defined loci.
SEQ ID NO: 27
agtgtggctgctgctaggGAGCTCattgagactaggtgagga

SEQ ID NO: 28
cacgtgaaagtgcctcaacaGATATCuggctgcagaatgagga

SEQ ID NO: 29
gccaagggagaaagttCTCGAGgctgtcctagcagaatgagga
FIG. 1D-2

Triple mutant clone #41

5' GACCAAGTGCTGGCTGCTGCAGGGAGCTCATGAGACTAGTGAGGAACCTCCTTT 3' WT (SEQ ID NO: 30)  
Tet1 GACCAAGTGCTGGCTGCTGCAGGGAGCT-----AGAGACTAGTGAGGAACCTCCTTT -3 bp (SEQ ID NO: 39)  
GACCAAGTGCTGGCTGCTGCAG----AGAGACTAGTGAGGAACCTCCTTT -9 bp (SEQ ID NO: 40)  

5' AGAAACACGTGAAAGTGCCACACAGATATCCAGGGCTGCAAGATCGGAGAACCAACGCC 3' WT (SEQ ID NO: 33)  
Tet2 AGAAACACGTGAAAGTGCC--------TCCAGGGCTGCAAGATCGGAGAACCAACGCC -9 bp (SEQ ID NO: 41)  
AGAAACACGTGAAAGTGCCACACAGATATCCAGGGCTGCAAGATCGGAGAACCAACGCC +1 bp (SEQ ID NO: 42)  

5' ATCTACACGGCAAGGGAAGAGGGAAGTTCTCGAGGGCTGCTGCCATCGGCAATTGGGT 3' WT (SEQ ID NO: 43)  
Tet3 ATCTACACGGCAAGGGAAGAGGGAAGCTGCTGCCATCGGCAATTGGGT -24 bp (SEQ ID NO: 44)  
ATCTACACGGCAAGGGAAGAGGGAAGCTGCTGCCATCGGCAATTGGGT -8 bp (SEQ ID NO: 45)  

5' ATCTACACGGCAAGGGAAGAGGGAAGCTGCTGCCATCGGCAATTGGGT
FIG. 2B

Tet2 mutants

#1 #2 #3 #4* #5 WT M

6.5kb

4.4kb

TM 5.6kb

WT 4.3kb

Tet2 probe
FIG. 2C

5'GACCAAGTGGCTGCTGCAGGACGCTCA\textbf{TGGAGA}}CTAGGTGAGGAACTCTGCTT 3' WT (SEQ ID NO: 30)

-9 bp (SEQ ID NO: 46)

+1 bp (SEQ ID NO: 32)

\textbf{Tet1 #2}
GACCAAGTGGCTGCTGCAGGACGCTCA---------TGGAGA}CTAGGTGAGGAACTCTGCTT
GACCAAGTGGCTGCTGCAGGACGCTCA\textbf{TGGAGA}}CTAGGTGAGGAACTCTGCTT

5'AGAAACACGTGAAAAGTGCCAACAGATATCCAGGCTGCAGAATCGGAGAACCACGCGC 3' WT (SEQ ID NO: 33)

-15 bp (SEQ ID NO: 34)

-150 bp (SEQ ID NO: 47)

\textbf{Tet2 #4}
AGAAACACGTGAAAAGTGCCAACAGATATCCAGGCTGCAGAATCGGAGAACCACGCGC
AGAAACACGTGAAAAGTGCCAACAGATATCCAGGCTGCAGAATCGGAGAACCACGCGC

8/110
**FIG. 2D**

**Tet1** probe

<table>
<thead>
<tr>
<th>RFLP</th>
<th>Southern</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td>9.4kb</td>
</tr>
<tr>
<td>#1</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#2</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#3</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#4</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#5</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#6</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#7</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#8</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#9</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#10</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#11</td>
<td>9.4kb</td>
</tr>
<tr>
<td>#12</td>
<td>9.4kb</td>
</tr>
<tr>
<td>WT</td>
<td>5.8kb</td>
</tr>
</tbody>
</table>

**Tet2** probe

<table>
<thead>
<tr>
<th>RFLP</th>
<th>Southern</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td>6.5kb</td>
</tr>
<tr>
<td>#1</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#2</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#3</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#4</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#5</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#6</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#7</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#8</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#9</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#10</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#11</td>
<td>6.5kb</td>
</tr>
<tr>
<td>#12</td>
<td>6.5kb</td>
</tr>
<tr>
<td>WT</td>
<td>4.3kb</td>
</tr>
</tbody>
</table>

**Double mutants**

- **Tet1**
  - TM 6.4kb
  - WT 5.8kb
- **Tet2**
  - TM 5.6kb
  - WT 4.3kb
**FIG. 2E-1**

**Double mutant #9**

5'GACCAAGTGTGGCTGCTGTCAGGGAGCTCA**TGGAGACTAGGTGAGGAACCTCTGCTT** 3'  WT (SEQ ID NO: 30)

-147 bp (SEQ ID NO: 48)

3' **TGGAGACTAGGTGAGGAACCTCTGCTT** -9 bp (SEQ ID NO: 46)

5'AGAAACACGTTGAAAGTGCCAAACAGATAATCC**AGGTGCGAGAATCGGAGAAACCACGC**C 3'  WT (SEQ ID NO: 33)

+1 bp, m1

(SEQ ID NO: 42)

3' **AGGTGCGAGAATCGGAGAAACCACGC** +1 bp (SEQ ID NO: 49)

Tet1

Tet2
FIG. 2F

Double mutants

bp
1. m1
0
FIG. 3C

Mice #5 from Double Oligo Injection

5' GACCAAGTGTGCTGCTGTCAGAAGCTCTCATGGAGACTAGGTTGAGGAACTCTGCTT 3'  WT (SEQ ID NO: 30)
Tet1
GACCAAGTGTGCTGCTGTCAG GTA TA TGGAGA CTA GGTGAGGA ACTCTGCTT
GACCAAGTGTGCTGCTGTCAG----------------GGAGACTAGGTTGAGGAACTCTGCTT

3' HDR (SEQ ID NO: 58)
-9 bp (SEQ ID NO: 40)

5' AGAAACACGTAAGGATGCAACAGATATTCGAGCTGCAAGATCGGAGAACCACGCC 3'  WT (SEQ ID NO: 33)
Tet2
AGAAACACGTAAGGATGCAACAGA AT TCCAGGGTGCAGAATCGGAGAACCACGCC
AGAAACACGTAAGGATGCAACAGA AT TCCAGGGTGCAGAATCGGAGAACCACGCC

HDR (SEQ ID NO: 59)

Mice #7 from Double Oligo Injection

5' GACCAAGTGTGCTGCTGTCAGAAGCTCTCATGGAGACTAGGTTGAGGAACTCTGCTT 3'  WT (SEQ ID NO: 30)
Tet1
GACCAAGTGTGCTGCTGTCAGAAGCTCTCATGGAGACTAGGTTGAGGAACTCTGCTT
GAC-------------------------TAGGGTGAAGGAACTCTGCTT

-34 bp (SEQ ID NO: 60)

5' AGAAACACGTAAGGATGCAACAGATATTCGAGCTGCAAGATCGGAGAACCACGCC 3'  WT (SEQ ID NO: 33)
Tet2
AGAAACACGTAAGGATGCAACAGA AT TCCAGGGTGCAGAATCGGAGAACCACGCC
--------------------------------GCTGCAGAATCGGAGAACCACGCC

HDR (SEQ ID NO: 59)
-72 bp (SEQ ID NO: 61)
FIG. 4B
One Step Generation of Mice With Multiple Mutations

Targeted Mutations (Deletion / Insertion)

Mutant

Embryo Transfer

Blastocyst

Mutant

Embryo Transfer

Blastocyst

Microinjection

NHEJ

Zygote

Microinjection

Predefined Precise Mutations

Zygote
FIG. 6B

\textbf{Tet1}  
\[ 5' \text{GACCAAGTGTTGGCTGTGCTAGAGAGCTCACTAGGTGAGGAACCTCTGCTT} \]  
\[ \text{GACCAAGTGTTGGCTGTGCTAG-------------GGAGACTAGGTGAGGAACCTCTGCTT} \]  
\[ 3' \text{WT (SEQ ID NO: 30)} \]  
\[ -9 \text{bp allele 1 (SEQ ID NO: 40)} \]

\textbf{Tet1}  
\[ 5' \text{GACCAAGTGTTGGCTGTGCTGTAAGGAGCTCAAGGAGACTAGGTGAGGAACCTCTGCTT} \]  
\[ \text{GACCAAGTGTTGGCTGTGCTGTA-------------TGGAGACTAGGTGAGGAACCTCTGCTT} \]  
\[ 3' \text{WT (SEQ ID NO: 30)} \]  
\[ -9 \text{bp allele 2 (SEQ ID NO: 46)} \]

\textbf{Tet2}  
\[ 5' \text{AGAAACACGTGAAAAGTGCCAAACAGATATCCAGGCTGCAGAATCGGAGAAACCACGCC} \]  
\[ \text{AGAAACACGTGAAAAGTGCCAA-------------AGGCTGCAGAATCGGAGAAACCACGCC} \]  
\[ 3' \text{WT (SEQ ID NO: 33)} \]  
\[ -8 \text{bp allele (SEQ ID NO: 52)} \]

\textbf{Tet2}  
\[ 5' \text{AGAAACACGTGAAAAGTGCCAAACAGATATCCAGGCTGCAGAATCGGAGAAACCACGCC} \]  
\[ \text{AGAAACACGTGAAAAGTGCCAAACAGA-------------ATCGGAGAAACCACGCC} \]  
\[ 3' \text{WT (SEQ ID NO: 33)} \]  
\[ -15 \text{bp allele (SEQ ID NO: 34)} \]
FIG. 6F

5'TCATCTACACGGCAAGGAGGGGAAGTAGTTCTCGAGGCTGTCCCACATGCGCAAGTTGG 3' WT (SEQ ID NO: 36)

Tet3 #14
TCATCTACACGGCAAGGAGGGGAAGTAGTTCTCGAGGCTGTCCCACATGCGCAAGTTGG
TCATCTACACGGCAAGGAGGGGAAGTAGTTCTCGAGGCTGTCCCACATGCGCAAGTTGG

-13 bp (SEQ ID NO: 65)

WT (SEQ ID NO: 36)
FIG. 7A

\[
\begin{align*}
ggctgctGTCAGGGAGCTCATGG & \quad \text{Tet1 (SEQ ID NO: 66)} \\
ctgttgtGTCAGGGAGCTCAAGG & \quad 1810013L24Rik (SEQ ID NO: 67) \\
gggcgaAGTCAGGGAGCTCAAGG & \quad \text{Lipg (SEQ ID NO: 68)} \\
gtttagtGTCAGGGAGCTCAAGG & \quad \text{Olfr339 (SEQ ID NO: 69)} \\
gaaagtgCCAACAGATATCCAGG & \quad \text{Tet2 (SEQ ID NO: 70)} \\
gcaagaagaCCAACAGATATCCAGG & \quad \text{Ubr1 (SEQ ID NO: 71)} \\
aggaacCCAACAGATATCCCGG & \quad \text{AK169506 (SEQ ID NO: 72)} \\
cacacccCCAACAGATATCCGGG & \quad \text{Cyp2c55 (SEQ ID NO: 73)} \\
gagataaCCAACAGATATCCGGG & \quad \text{E430025E21Rik (SEQ ID NO: 74)}
\end{align*}
\]
FIG. 8C

Embryo #2 from Double Oligo Injection

5'GACCAAGTGTTGCTGCTGTCAGGGAGCTCATTGGAGACTAGGTGAGGAACCTCTGCTT 3' WT (SEQ ID NO: 30)

Tet1
GACCAAGTGTTGCTGCTGTCAGGGATTCATGGAGACTAGGTGAGGAACCTCTGCTT HDR (SEQ ID NO: 58)
GACCAAGTGTTGCTGCTGTCAGGGACATGGAGACTAGGTGAGGAACCTCTGCTT -10bp (SEQ ID NO: 75)

5'AGAAACACGTGAAAGTGCCACACGATACCTCGAGCTGAGAACTCGGAGAACCACGCC 3' WT (SEQ ID NO: 33)

Tet2
AGAAACACGTGAAAGTGCCACACGATACCTCGAGCTGAGAACTCGGAGAACCACGCC HDR (SEQ ID NO: 59)
AGAAACACGTGAAAGTGCCACACGATACCTCGAGCTGAGAACTCGGAGAACCACGCC +1bp (SEQ ID NO: 53)
FIG. 9

Unique 20bp DNA sequences that encode sgRNAs

Tet1: ggctgctgtcagggagctca (SEQ ID NO: 76)

Tet2: GAAAGTGCCAAACAGATATCC (SEQ ID NO: 77)

Tet3: AAGGAGGGGAAGAGTTCTCG (SEQ ID NO: 78)

Sry: gcatttatgttgtggtcgcg (SEQ ID NO: 79)

Uty: ttttttttttctattaccta (SEQ ID NO: 80)

Full length sgRNA sequence:

Tet1:

ggcugcugucaggacucaGUUUUAGAGCUAGAAUAGCAAGUUAAUAAAGGCUA
GUCCGUUAUCAACUUGAAAGUGGCACCGAGUCGGUGCUUUUU (SEQ ID NO: 81)

Tet2:
gaaagugccaacagauuccGUUUUAGACCUAGAAUAGCAAGUUAAUAAAGGCUA
GUCCGUUAUCAACUUGAAAGUGGCACCGAGUCGGUGCUUUUU (SEQ ID NO: 82)

Tet3:
gaaggagggaagagucucgGUUUUAGAGCUAGAAUAGCAAGUUAAUAAAGGCUA
AGUCCGUUAUCAACUUGAAAGUGGCACCGAGUCGGUGCUUUUU (SEQ ID NO: 83)

Sry:
gcauuauagggugggaggccgGUUUUAGAGCUAGAAUAGCAAGUUAAUAAAGGCUA
GUCCGUUAUCAACUUGAAAGUGGCACCGAGUCGGUGCUUUUU (SEQ ID NO: 84)

Uty:

quuuucuuuuuccuauaccuagUUUUAGAGCUAGAAUAGCAAGUUAAUAAAGGCU
AGUCCGUUAUCAACUUGAAAGUGGCACCGAGUCGGUGCUUUUU (SEQ ID NO: 85)
FIG. 10C-2
+ 2 Days Dox

81%
FIG. 14

Wt Cas9 (S. pyogenes)

ATGGGATAAGAACATCTCAATAGGCTTTAGATATCGGCACAATAATAGCGTC
GGATGCGGCGGTATCAGTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGG
TTCGAGAAATACAGACGCCCAAGTGATACAAAAAAAATCTTTATAGGGCTCT
TTTATTGACAGTTGGAGAAGACACAAGCGGAAGCGACTCGTCTCAAAACGGACAGCT
CGTAGAGAGGTATACACGTGGGAAAGATCTGTATTTGTATCTATACAGGAGATT
TTCAAAATGAGATGGCGAAAGATGATGTAGATGTATTTCTTCATCGACTTGAAGAG
TCTTTTTGGTGGAAGAGACAAAGACAGATGAACGTCATCTTATTTTGGAAA
TATAGTAGATAGAAGTTGCTTATCATGAGAAAATATCCAACATCTATCATCATCTGC
GAAAAAATTGGTAGATTCTACTGATAAAAAGCGGATTGGCCTTAATCTATTTG
GCCCTAGCGCATATGATTTAGTTCTGCTGGCTATTTTTGATTTGAGGAGATT
AAATCCTGATAATAGTGATGGACAAAACTATTTATCCAGTTGGCTACAACACT
ACAATCAATTATTTGAAGAAAAACCTATTAACGCAAGTGAGTAGATGATGCTAA
AGCGATTCTTTCTGCACAGTTGATATACAAAGACAGTTAGAAGAAATCTCATTG
CTCAGCTCCCCGTTGAGAAAGAAAAATGCTTTATTTGGGATATCTCATTTGTTTG
TCATTGGGTTTGAACCCCTAAATTTAATCAAATTTTGGATTTTGAGCAAGAGATGC
TAAATTACAGCTTTTCAAAAGATACCTTACGATGATGGTTAGATAATTTTATTTG
CGCAAATTGGAGATCAATATGCTAGTTTTGGTAGCTAAAAGACAGTTATATCAC
GATGCTATTTTACTTTCCAGATATCTCAAAGAAGTAATACTGAATAACTAAGGC
TCCCCTATCAGCCTCATAATGATTTAATACGCTAGGAACATCATCAAGACTTGA
CTCTTTTAAAGCTTTTAGTCTCGACAAAACTTTCCAGAAAGATATAAAGAAT
CTTTTTGGTATCAATCAAAACAGGATATGCAAGTTATATTGATGTTGGGAGCT
AGCCAAGAAGAATTATTAAAAATTATCAAACAAATTTTGAAGAAAAATGAGATG
GTACTGAGGAATTATGGTGAAACTAATACGTAAGATTGTGCCGCAAGACA
ACGGACCTTTGACAACGGCTCTATTTCCCATCAAAATTTACACTTTGGGTAGAGTGC
ATGCTATTTTGAGAAGACAAGAGACTTTTATCCATTATTTAAAAGACAAAATCGT
GAAAGATGAAATATACTTTGTACCTTTCCGAATTCTATTTATATTGTTGGTCATT
GGCGCGTGGCAATAGTCTGTGGTGGCAGAATCGAAGCTCAGGAAAGACACGAT
ATTCCCATGGAATTTTGAAAGATTGTCTACAGCTCAATC
ATTATTAGAAGCAGATAAATTTTGGATATAAATCTTTCCAAATGAAAAAGTA
FIG. 14 (cont.)

CTACCAAAACATAGTTGGTCTTTATGAGTATTTTACGTTTTATAACGAATTGAC
AAAGGTCATATATGTTACTGAAAGGATGCGAAAACCAGCATTTTTCTGAGT
GAACAGAAAGAAAGCCTAGGTTGATTTACTCTTTAAAAAAACAGATATGTTTTGATAG
TGTTGAAATTTTCCCAGGGTTGAGATAGATTAAAAATGCTTTATGTTACTGATACC
ATGATTTGCTAAAAATTATTAAGATAAAGATTTTTCTGATAATTGAGAAA
TGAAGATATCTCTAGAGGATATTGTTTTAATACATGACCTTATTTGAGAAGGAGG
AGATGATTGGAGAAGACTTAAACATATCTCGACCTCTTTGATGATAAAGGT
GATGAAACAGCTTTAACGTCGCTCCTTTATACCTGGTGGGAGGACCTTTGTCTCGA
AAATTGATTAATGGAATTTAGGGATAGCAATCGCATAAAAAACATATTAGATT
TTTGAATATCGATGTTTTTGCCAATCGCAATTATATGCAAGCTGATCATATGAT
GATAGTTGGACATTAAAGAAAGACATTTCAAAAAAGCACAAGTGTCGGACAAG
GCGATAGTTTACATGAACATATTTGCAAATTTAATGGTGTAGCCCTGTATTTAA
AAAGGTATTTTACAGACTGAAAGGATTTGTTGATGAATTGTCAAAAAGTGG
GGCGGCATAAGCCAGAAAAATATCGTTATGGAAATGCGCAGTGAAAATCAGAC
AATCTAAAAAGGGCCAGAAAAATTTCGCAGAGCGTGATGAAACGAATCGAAGA
AGGTATCAAAGAATAGGAAGTCAGATTCTTTAAAAGAAGCATCTCTGTGAAAAT
ACTCAATTGCAAATGAAAAAGCCTCATATCTCATTATCTCCAAAAATGAGAGG
ACATGATGTTGGACAAGAATGATATAATATGTTTAAAGTGATATTGATGTC
GATCACATTTGGTCCACAAAAAGTCTTTCTTTAAAGACGATTCATAGACAATAAGG
TCTTTACCCGTTCTGATAAAAATCTGTTGTAATACGGAATACGCTTCAAGTGA
AGAAGTGATCATAAAAAAGTGAAAACATATTTGAGAACAATCTCTAAACGCAA
GTTAATCTACTAAGCTAGTTGGTAATAATTTAACGAAAGCTGACGTCGGAGGT
TTGAGTGAACTTGGTAAAAAGCTGGTTTTATCAACGCATAATTGGGAAAATCTCG
CCAAAATCATAAGCATTGCGAACAATTTTTGAGATAGTCGATGAAATACTAAA
TACGATGAAAATGATAAAACTTATTCCGAGAGTTAAGGATGATCTACCTAAAAAT
CTAAATAGTTTCTGACTTTCGAAAAGATTTTCAACTTATATAAGTACGTTAG
ATTAAAATTTACATCAGTCCCATGATGCTATCTAAATGCGCTGTTGGAAC
FIG. 14 (cont.)

TGCTTTGATTAAGAAATATCCAAAAACTTGAAATCGGAGTTGTCTATGGTGATT
ATAAAGTTTATGATGTTGCTAAATGATATGCTAGTCTGAGCAAGAAATTAGG
CAAAGCAACGCIAATATTTTCTTTACTCTATATCATGAAACTTTCTTCAAA
CAGAAATTACACTTGGCAATAATGGGAGATCTCGCAACCGCTCTATAATCGAAAC
TATGAGGGAAAATCGGAGAAATTTGTCTGGGATAAAGGGCGAGATTTTGCCACA
GTGCGCAAGTAGTTGTCATGCCCCAACAGTCATATTGTCAGAAAGCAGAAG
TACAGACAGCCGATTCTCACAAGGAATTTATCAGCAATTTTACCAAAAAAGAAATTTCGGA
CAAGCTTATTGCTCGTAAAAGAGCCTGGGATCCCAAAAAATATGTTGTTTT
GATAGTCTCACAAGGTTAGCTTTATTCAGTCTCTAGTGTGGTTGCTAAGGTTGAAAAG
GGAAATCGAAGAAGTAAATCCCGTTAAAGAGTTACTAGGGAICAACATTTAT
GGAAAGAAAGTTCTCTTGGAAAAAATCCGATTTGACCTTTTGAAGCTAAGGGA
TATAAGGAAGTTAAAAAACAGCTTAACTATTAAACTACCAAAATATAGTCTTTT
TGAGTTAGAAACCGTGTAACCGGAGTGGCTGCGTATGGCGCCGGAGAATTCAAA
AAAGGAAATGAGCCTGGCTCTGCAAAGCAATATGTAATTAAAAATTACG
TAGTCAATATGAAAAGTTGAAGGTAAGTCAGAAAGATACAGAACAAAAACA
ATTGTTTGTGAGCAGCATAAGCATTATTAGATGAGATTATTAGCACAATCATCA
GTGAATTTCCTAAAGCCTGTTATAATTAGCAGATGCCAATTAGATAAAGTTCTT
AGTGCATATAAACAACATAGAGACAAAAACCAATACGTAAGCAAAGCAGAATA
ATTATTCTTATTATCTGGTGACAAATCTTGGGAGCTCCCAGCTCTTTAAATAT
TTTGGATAACAACATTGATCGAAACGATATACGTCTACAAAAAGAAGTTTTAG
ATGCCACTCTTTATCCATCAATATATCTGTTTTATTGAAACAGCATTGAT
TTGAGTCAGCTAGGAGGTGACTGA

(SEQ ID NO: 485)
FIG. 15A

hSRY  DRVKRPMAFIVWSDRQRRKALENPRLR NSISKQLGLQWKMLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAK (SEQ ID NO: 86)

mSRY  GHVKRPMAFIVWSDRGERHKLAAQQNPMQNTISKQLQ CRWKLTEAEKRPFQEEAQRLKILHREKYPNYKYPHRRAV (SEQ ID NO: 87)

cSRY  DRVKRPMAFIVWSDRQRRKALENPRLRNSISKQLGLQWKMLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAN (SEQ ID NO: 88)

pcSRY DRVKRPMAFIVWSDRQRRKALENPRLRNSISKQLGLQWKMLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAN (SEQ ID NO: 88)

gSRY  DRVKRPMAFIVWSDRQRRKALENPRLRNSISKQLGLQWKMLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAK (SEQ ID NO: 86)

pySRY DRVKRPMAFIVWSDRQRRKALENPRLRNSISKQLGLQWKMLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAK (SEQ ID NO: 86)

hISRY DRVKRPMAFIVWSDRQRRKALENPMLRNSISKQLGRLQWKMLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAK (SEQ ID NO: 89)

bSRY  DRVKRPMAFIVWSDRQRRKALENPMLRNSISKQLGRLQWKMLTEADKPFQEEAQKLOQMHRKYPNYKRYPRRRKAK (SEQ ID NO: 90)

cjSRY NRVKRPMAFIVWSDRQRRMAVENPDNLNSKRLQYWKLLTEAEKPFFQEEAQKLOQMHRKYPNYKRYPRRRKAN (SEQ ID NO: 91)

Helix 1  Helix 2  Helix 3
FIG. 16B

<table>
<thead>
<tr>
<th>M</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>WT</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>V5F + SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SF + SR
FIG. 17B

<table>
<thead>
<tr>
<th>Ncol</th>
<th>WT #1</th>
<th>#2</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WT 11.5kb
T 5.6kb

3' external probe

Internal probe

FIG. 17C

Nanog-cherry blastocysts
Nanog-cherry mES clone
Nanog-cherry mES clone differentiated
FIG. 17E

Oct4-eGFP blastocysts

Oct4-eGFP mES clone

FIG. 17F

<table>
<thead>
<tr>
<th>Hind III</th>
<th>Hind III</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>#1</td>
</tr>
<tr>
<td>WT</td>
<td>WT</td>
</tr>
</tbody>
</table>

9.4  WT 9kb
6.5  T 7.2kb
4.4

3' external probe

Internal probe
FIG. 18A

5' ... tacagtat GCTAGC ataact... aqttat cctagggaagtt... 3'  L2 oligo donor (SEQ ID NO: 102)

5' gcttgagCCAAGGATAACGTATCTCTAaggtt 3' (SEQ ID NO: 103)

Mecp2

EcoRI

Nhel

loxp

R1

5' gaccccccTTCAATGATGTGAGGTAGGAatcagtt 3' (SEQ ID NO: 104)

3' ctgggggTTCATGATCTGAGTGAGGAagttca 5' (SEQ ID NO: 105)

3' ...ctgggggttcga CTTAAG tatgga... tcaata tgaatctgg... 5'  R1 oligo donor (SEQ ID NO: 106)

EcoRI

Nhel

loxp

Exon3 probe

3.2kb

Exon4 probe

1.3kb

0.7kb
FIG. 18B

EcoRI & Nhel

♀ ♂ ♂ ♂ ♀ ♂
#1 #2 #3 WT #4 #5

22 — 9.4 — 6.5 — 4.4 — 2.3 — 2

WT 5.2kb
L2-loxP 3.9kb
R1-loxP 2kb
2loxP 0.7kb

Exon 3 probe

EcoRI & Nhel

♀ ♂ ♂ ♂ ♀ ♂
#1 #2 #3 WT #4 #5

22 — 9.4 — 6.5 — 4.4 — 2.3 — 2

WJ 5.2kb
L2-loxP 3.9kb
2loxP 3.2kb
R1-loxP 3.2kb

Exon 4 probe
FIG. 18C

Deletion (DF&DR)

<table>
<thead>
<tr>
<th></th>
<th>#1</th>
<th>#3</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cre:</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Circle (CF&CR)

<table>
<thead>
<tr>
<th></th>
<th>#1</th>
<th>#3</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cre:</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
FIG. 18D

Targeted fragment

FIG. 18E

WT 5' gcttgagcccaggatatcataaggt... 3' cgaactGGTCTATGTCATAGGATcccttca...

Del 1 5' gcttgagcccaggatat... 3' (SEQ ID NO: 109)

Del 2 5' gcttgagcccaggata... 3' (SEQ ID NO: 110)

Del 3 5' gcttgagcccaggat... 3' (SEQ ID NO: 111)
FIG. 19A

**Tet1**

```
agtgt**GGCTGCTGTCAAGGAGCTCA**tggagactaggtgaggaa (SEQ ID NO: 112)
5'...gtcagggag GAATTCCaaaact...agtttatctcatggagac...
EcoRI   loxP
3' Tet1 oligo donor (SEQ ID NO: 113)
```

**Tet2**

```
cacgt**GAAAGTGCAACAGATATCC**aggctgcagaatcgaggaa (SEQ ID NO: 114)
5'...ccaacagat GAATTCCaaaact...agtttatatccaggtgagc...
EcoRI   loxP
3' Tet2 oligo donor (SEQ ID NO: 115)
```
FIG. 21A

5'...tacagtat GCTAGC ataaact...agttatat cctagggagtt... 3' L1 oligo donor (SEQ ID NO: 116)
5' gcttga CCAAAGGATACAGTACCTAagttt... 3' (SEQ ID NO: 117)

Mecp2

5' aatct TGGGGCCCCAGCCTGACCAAgtaca 3' (SEQ ID NO: 118)
5'...cccagcttg a GCTAGC ataaact...agttatat cccaaaggtacag... 3' L2 oligo donor (SEQ ID NO: 119)

5'...gttgtggattag GAATTCCataact...agttatat gtttgtggtaccaac... 3' R2 oligo donor (SEQ ID NO: 120)
5' gaaatct TTTGCTTTGATTATTCTGggtacca 3' (SEQ ID NO: 121)

Mecp2

5' gacccccAAGTACTAGACCTCCTCCTcgattt 3' (SEQ ID NO: 122)
3' ctggggggCTCATGATCTGAGTGAGGA 5' (SEQ ID NO: 123)
3'...ctgggggtcga CTTAAGtatcatt...tcgtatatgtctgg... 5' R1 oligo donor (SEQ ID NO: 124)
FIG. 22A

Mice from oligos double injection

Mecp2-L2
LF + LR

Mecp2-R1
RF + RR

+ Nhel

+ EcoRI
FIG. 22B

Partial chromatograph from single sequence file of a floxed allele

---

Intron 2

Intron 3

loxp

Nhel

EcoRI
FIG. 22C  Partial chromatograph from Cre-loxP mediated recombination PCR products
Mecp2 exon3 deletion

Mecp2 exon3 circle
FIG. 23A

dCas9VP48

D10A  H840A

rTta-M2  +Dox  sgTetO  TetO  tdTomato

HeLa/TetO::Tomato;EF1a::rtTA-M2
FIG. 23B-3

iii. dCas9VP48

77/110
iv. dCas9VP48 +sgTetO
<table>
<thead>
<tr>
<th>Nanog::EGFP:dCas9VP48+sgNanog1-7</th>
<th>GFP</th>
<th>HMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Day2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Day3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Day4</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 23D**
FIG. 24E

SOX2

Fold

dCas9VP160 + sgSOX2
FIG. 25A
FIG. 26A

Gene Count

Log₂ fold change
(sgTetO/sgTetO-mut)

FIG. 26B

Gene Count

Log₂ fold change
(sgIL1RN1-3/sgTetO-mut)

IL1RN
FIG. 27

Fold (log₁₀)

- No transfection
- dCas9VP48
- dCas9VP48 + sgTetO

2 days 12 days 18 days

FIG. 28A

D10A H840A

* *

dCas9

16 16 16

VP48

rTta-M2 VP48

+Dox

dCas9

sgTetO

TetO tdTomato
FIG. 28B-1

NIH3T3/TetO::tdTomato; EF1a::rtTA-M2

i. GFP control

Phase

tdTomato

FACS

P6
FIG. 28B-3

iii. dCas9VP48

Histogram showing a distribution of values with a peak around $10^4$.
iv. dCas9VP48 + sgTetO
FIG. 29

KH2MSI mESC line
CAGGS::rtTA-M2 at Rosa26 Locus
TetO::Msi1 at Col1A Locus (flip-in)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSI1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- dCas9VP48: + + + - - - + +
- dCas9: - - - + + +
- sgTetO: - + - + -
- sgTetO-mut: - - + - -

GFP ctl + Dox - Dox
No Transfection
pAC91: dCas9VP64 on pmax expression vector

```
1  tcaatatttg ccattagcca tattattcat tggttatatga gcataaatca atattggcta
61  ttggccatttg catacgttgt atcttatata taatatagtc attttatattg gctcatgtcc
121  atatatgcgg ccatgtttgcc atattattat taatagataat cattactggg
181  tgttattgtt atagacccct atatgaggtt ccgcctttaca taacttacgg taaatggcc
241  gccctgctga ccgcgaacgc acctccggccc atgcagctca ataagctacg atgttcccat
301  aagtaacgcca atagggacttt ttccattgcgc tcaatgggtg gagtagttac sgtaaatgc
361  ccactttgga gcacactgca aagctatcatt gtcgactatgt gccaagtgcc ccctaccttg acgtcaatga
421  cggtaaatgg ccgccttgcc aatattggca gtaatagctc ttacctagct ttcctacttg
481  gccagttctac tgatttttag gagcaagttg ttaacagttttg cacgttacac ggcgaagctc
541  caatgggcggc ggatagcgcgt ttgaacctcc gggatttccca aqtctccacc ccttaagcgt
601  caatggaggtt tggtaattggg acaacccgct acgaaacctc aagctctctcc aaaaatatgt ctaataacc
661  cgccctgcgtg aagccagtgg ggcctggccttg tgtacgtggtt gaggtctata taagcagcggg
721  tctgtttttag aaccgcggaga tccaagttcg cacctcagg tttatgtcgc tgcgatgggtt
781  tggcataagc atggattgtca ggtcttcgtta gacagagaag attctttgtgc tctgtatagg
841  aagggaggcc aataaaacct ggccttggtcg gacagagaat aagcttttgtct tgtttcttgg
901  cacctatttg tctttactgac atccacattg cctttctctc cacagggaag aaaCAATTGA
961  caagttttagta caaaaaagca ggtctccgaat tcacccgggct caccatgtac ccatacgatg
1021  tgtcctgctgc ggcttggact gcggacctgc gcggaaacgc ggcttcgag cggctgagg
1081  gcatcggcct ggccttcgca acaactcttc tgggctgggc cgctagtacc gcacagtaac
1141  aggtgccccag cagaaatatt aggtgcttcgg caaacaaccc gagcggacgc atcaagaga
1201  acggtatcgg acgctctgtg cgggacagcc cggggaaagc cggagcagcg cggcagagga
1261  gaaaccgcgg aagagacatg accaggccag aagaccgctg acaaacctgt ctcgtactcg caagagatct
1321  tcagcaacgca gatggcctgcag tggagcagaca gcttctcctca cagactggaa gatcttccttc
1381  tggaggagaa ggaataggag cagcggaggg accccatacc ggcaacagtc gtaggacgagg
1441  tggccctacca cggagagttacc cccaccctct accacactgag aagagaaaactg gtaggacagca
1501  cggacaaggg cgcacctccgg ctgctctctc tggcccttgc cccatgtac aagttccggg
```
FIG. 32A (cont.)

3241 ATCTGGCGCG CAGC CCCGCC ATTAAAGAGG GCATCTGAGA GACAGTAGA GTGGTAGGACG
3301 AGCTCTGAGA AGTGAGGAGG C CGACAGACGC CCGAAGACT CGTGCTGAGA ATGAGGACAG
3361 AGACCCAGAC CACCGAGAAG GACAGAGAAG ACAAGCGGGA GAGAATGAAG CGGATCGAAG
3421 AGGGCATCAA AGAAGCTGCG AGCCGATACC T GAAAGAAAAG CCCCTGGAGA AACACCCAGC
3481 TGCAAGAAGAC GAAGCTGATCG TGCTGATACC TGGAGAATGG GGGGATATGG TACGTAGGACC
3541 AGGAATCGTG GTACACAGAG GCAGGTCGAG CTGCTGATGG AGGAGCTGGG CCGCTAGCTT
3601 TTCTGAGAAGG AGACTCCACT GAACAAAAGG TGGCTGACAG AGCCGACAGG AACCGGGAGA
3661 AGAGCGACAA CGTGCCCCTCC GAAGAGGTGC TGAAGAGATG GAAGAGAAT C TGGCGCGACA
3721 TGCTGAAACGC CAAGCTGATT ACCAGAGAGA AGTGCACGAA TGTGAGCAAG CTGAGCAAGG
3781 GCGGCTGAGA C GAACTGGAAT AAGCCCGCCT CTATCAAGAG ACAGCTGAGT GAAACCGAGC
3841 AGATCAACAA GCAGTGAGCA CAGATCTGCT GACTCCCAGAT GAACACTAA C TACAGAGAGA
3901 ATGAACACCT GTACGAGGAA GTGAGAGTCA TGCACGCTGA GCTGAGCACT GTGCGAGATTT
3961 TCCGGAAGGA TTTCCAGTTC TACAAAGTGC GCGAGATCAA CACGAGCAAC ACCCCAGAGC
4021 ACAGCTACCT GAAAGCCTGT GAGGAAAGGC C CCGTATACAA AAGTACCCCT AAGCTGAGAA
4081 GCGAGTGGTT GTACGAGTGA TACAGCTTGC AGCGAAGATT CGAGTGATGC TCCAGAGCGC
4141 AGCAGGAAAT CGCGAAAGGT ACGCCCAAGT ATCTCTTCTA CGACATACAT ATGACCTTTTT
4201 TCAAGGCCGA GATTACCTTG GCCAACCAGG AGATCCGGAA GCGCCCTCTG ATCAGACAAA
4261 AGGGCGCAAG CGGGAGAAGT GGTGAGGATA AGGCGCGGAA TTTTCCGACC GTCGAGAACA
4321 TGCTGAGCAT GCGCAAGT GTATGAGGAT AAGAGAGCT GGTGAGCGAA GCCAGCTTCA
4381 GCAAAGAGTC CAGCTGGCCT AAGGAAAGTG CAGATAAGCT GCTCCGAGCA AAGAGACTTC
4441 GGGACCCTAA GAAGTACGAG GGCCTGCCGA GCCCACCCTG GGCCTATTCT GTCGTTGGTGT
4501 TGCGAGAAGT GAAAGAGGGA AAGTTCCAGA ACGTGAAGAG TGTGAAAGAG CTGCTGGGGA
4561 TCACACCATC GAAGAAAGAC AGCTGCGACT GAAAGAGCAT CAGCTTTCTG AGTCAGAAAG
4621 GCTAAGAGA AGTGAAGAC AGCCGAGTCTA TACAGGCTAC TAAATCTTCC GTTCTGAGGC
4681 TGCGAAAAGGA CGGGAAGAGA ATGCTGCTCT TGTACGAGGA AGTGAAGAAG AGGAAACAGAC
4741 TGCGCGCTGG CCCTGAATAT GTTCTGCTCC TGTACGCTGC AGGCAGCTAT GAGAAGCCTGA
4801 AGGGCTCCCC CGAGGATAAT GACAGGAAAG AGCTGTTTGT GGAACAGCAC AAGCTACCTAC
4861 TGGACGAGAT CATCGACGAG ATCAGCGAGT CTGCAAGAG TACGTATTTC GCCAGCGCTA
FIG. 32A (cont.)

6601 gcggtcagcc catctggccgc caagcttttc agcaatatca cgggtagcca acgtcatgtc
6661 ctgatagcgg tccgccacac ccagccgggc acagtcgatg aatccagaaa acggcctatt
6721 ttccaccatag atatctggga acgagccgct gcctagggat akgacagatg cctcggtgctc
6781 ggccccagcc gctcttggagcc tggcgaacag ttggtctggt gccggcctct gatgtccttc
6841 gtccccagatca tcctgtagcga caagaccggc ttccatccga gtaacgtgctc gctccagatcg
6901 atgtttcctgt ctggtggtcga atgggcaagtt acgacgcagta acggtatgcga ggccgcccgt
6961 tgtcgctagcgc atgatggata ctttcctggc agagacgaag tagatgcaga ggagactctg
7021 cccgctgaact tcgcccacatt ccgacaggtc cctccctgct tccagtgaacca cgtccagacac
7081 agctgctgcaaa ggaacgcgccc tgcgtgctgct ccaagatagc agcgctgctct cgtctttcag
7141 ttcatttcaag gcaccgcgaca ggtccggcttt cacaacacag aacgagcgcgc cctgcgctgta
7201 cagcgcgcaac acgagcggcat cagagcgacc gattgtcctgt tgtgcccaagt cattacgcca
7261 tagctctctcc atcccaagcgg ccgggaaccc tgcgctgcaat cccatctttg tcaatccataat
7321 attatctggaag catttatcacg gtttcgtcttc ttccctgtcgc ccctccatgc atg

(SEQ ID NO:486)

FIG. 32B

pAC92: dCas9VP96 on pmax expression vector

1  tcaataattgg ccattagcga tattatctat tggttatata gcataaatca atattggcga
61  ttggcctattg cataagtttgt atatatataa taatatgtac attatatattt gctcatgtcc
121  aatatagcccg ccagttgagg attgaatttt gactatgatt taatagtaat caattacgag
181  gtcattaggta catagccccc atatttgctta ccgctgatca taatttgacgg taaatgccc
241  gccttggtgta ccgccccacgc acccccgggg attgacgtca ataatgacgt atgttcccct
301  aagtaagggga atagggagtt tccattggacg tcaatggtggt gagattttcg ggtaaactgc
361  ccaattgttgc gcatacagca agttcgatctg ttcgatagc ggccccatgg cagcttactg
421  cggtaaatgg cccgctctgag attatatgca gtaacgtagga ttaaggagact ttcctacttg
481  gcagtagcatac tacagttattg tcatcgcgtat taccatgcct gttcgggtttt ggcagtaac
Fig. 32B (cont.)

2221 CCGAGGAACT GCTCGTGAGG CTGAACAGAG AGGACCTGCT GCGGAAACAG CCGACCTCGA
2281 ACAACGCGAC CATCCCCCAC CAGATCCACC TGGGAGAGCT GACAGCATTG CTGCAGCGCC
2341 AGGAAGATTT TTACCCTTCT CGTAAAGGACA ACCGGGAAAA GATCGAGAAG ATCCCTGACTT
2401 TCCGACTCCG CTAATCTCCTG GGGCTCTGGG CCAAGGAAAC CAGAGATTTC GCTGGATGTA
2461 CCAAGAAAGG CGAGGAACAC ATCAACACCTT GGAACCTCAG GGAAGTGGTG GACAGGGCGG
2521 CTTCCCGCCA GAGCTTACCT GACGGCTGTA CCAACTTCCG TAAAGACCTG CCAACACCGA
2581 AGGAGCTGTCG CAAGGACACG CTGCTCTGAC AGTACTTCTC GGTGATAAC GAGCTGACCA
2641 AAGGTGAAATA CGTGAGCCAG GGAATGAGAA AGCCGCGCTT CCTGAGCCGG GACAGGAAAC
2701 AGGAGATCTG TGAACCTGCT TTTACACCCA ACCGGAAGGT GACGCTGAA GAGCTGGAAG
2761 AGGACTACTT CAAGAAATAC GAGTGGCTC TTTATCGTGA AATCTCCGCG GTGAAAGACT
2821 GGTCTCAAGGC CTTCCTGAGG AATCCACAGC ATCTGCTGAA AATTATCAAG GCAAGGACT
2881 TCCTGACGAA TGAAGCCTCG TTTGAGATAT TGTGCTGACC GTAGAGCACC CTGACAGCTG
2941 TGGAGGACAG AGAGATGATC GAGGAACGGC TGAATACCTA TGCCACCTTG TCCGACGCA
3001 AAGGTGATGAA CGCGCTGAAA GGCAGGATAT CGGCCGCTTG GGCAGGCTG AGCCGAAGAC
3061 TGACTCAAGGC ATTCGGCGAC CAGAAGCTCC ACAAGACTCC CCTGAGTTCCT CTAAGTCCG
3121 AGGAGCCTTCG CAACACAAAG TTCTATGACG TGCTAGAGCG GTGAGCAGTC ACCCTTAAAG
3181 AGGACATCCA GAAAGCAGAC ATGAGCTGTC TGGAGGTCAG CAGAGGCTG AATTGACGCA
3241 ATCTGGCAGG CAGCCCGGAA ATTACAGAGC GCTACTTCA GACAGTAAG GGTGGAGCG
3301 AGCTCTGTGA AAGTATGGGC CGGCAACAGG CGGGAAATGT CTGATGCAAC ATGCCAGAGC
3361 AGGACACAGAC CCAAGAGAAC TACAGCTTGG GCAGTGAAAG AGAATCGAAG TGGAGCGAAG
3421 AGGGCATCAA AGAGCTTGGC AGCGAATGTC TGAAGAAACA CGGCGTGAA AACACACGAC
3481 TGCAAGAACGA GAGCTGTGGC CTGACTACC GCGGATATGG GGGAGTATCG TACGTTGACC
3541 AGGAACTGCGA CATCAGACCG CGGTAGCCTG GGCATCTGCG ACCATGCTG CCTACAGACT
3601 TTCTGAAAGG CAGACTTATC GACACAAAGG TGCTAGACCG AGCGACAAG AACGCCGGCA
3661 AGAGGCAGACA GGTGGGCGCC GAGAGGATGC TGAAGAGATG GAAAGACTAC TGGCGACGCC
3721 TGCTGACGCA CAAGATGATT ACCCGAGAAA AGTTCCGCAA CTCTGAAAGG GCCGAGAGAG
3781 CGGGCGGTGAG CGAAGCTGTT AAGCGCGCTG TTAATCAAGG ACGCTCCTGG CAAACCCGGC
3841 AGATCACAAGA GCACGTTGCCA CAGATCTCGG ACTCCCGGAT GAAACACACTG TACGAGAAG
FIG. 32B (cont.)

3901 ATGACAAAGCT GATCCGGGAA GTGAAAGTGA TCAACCCTGA GTCCAAGCTG GTGTCGAGAT
3961 TCCGGAAGGA TTTCCAGTTT TACAAGTGTG GCGAGATCAA CAACTACCAC CACGCCACGG
4021 ACGCCTACCT GAACGGCGGC GTGGAAGACG CCTGTAGCAA AAGTACCTCT AAGCTGGAAGA
4081 GCGAGTTCGT GTACGGCGAC TACAAAGGTT ACGACGTGGC GAAAGATGAC GCAAGAGACG
4141 AGCACGGAAT ATGGAAGGGT ACCGACAGTG ACTCTCTCTTA CAGCAACATC ATGAACTTTT
4201 TCAAGAGCAGA TATACCCCTG GCCAACCGGC ATAGCCGGAA GCCGCTCTTG ATCGAGACA
4261 ACGGCGAAAC CGGAGGAGTG GTGTTGGATA AGGGGCAGGA TTTTGCCACC GTGCGGGAAG
4321 TGCTGAGCAT GCCCCAGTGA AAATATCTGA AAAAGACCCA GGTGCAAGAA GCGGCTTCA
4381 GCAAAGAGTC TATCCGCCC AGAGGGAACA GCGATAAGCT ATCGCCAGA AAGAAGACT
4441 GGGACCCTAA GAAGTACGGC GCTTCCGACA GCCCCACCGT GGCCTATTTT GTGCTGGTGG
4501 TGCCCAAAGT GGGAAAGGCG AGTCTCAAGA ACTGGAAGAG TGTGAAGGAG CTGCTGGGGA
4561 TCACCCTCAT GGGAAAGGCG AGTCTCAAGA AGAATCCCTG CGACTCTCTG TGAAGCAAGG
4621 GCTACAAAGA AGTGAAGAAC GACCTGTGAG CTAAAGTGCC TAAATCTCCC CTGGTGACGC
4681 TGGGAAAACCC CGGAGGAGAG ATGCTGGGCC CTGCGCGGCG ACTGCGAAGAG GAAAGCAAG
4741 TGCCCTCTCC CTCCAAATAT GTGCAAGCCTG TTGGACCTGC CAGCCACTAT GAGAAGCTGA
4801 AGGGCCTCCC CCAGGAATAT GACAGAGAAC ACGTGTTTGT GGAACAGCAC AAGCCTACC
4861 TGAGAGGAGG CATCGAGCAG ATCCAGAGAT TCTCCAGTCC ATGATGCTGC GCCGAGCTAA
4921 ATCTGGACAA AGTGCTGTCC GCCGTACAAG AGCAGCGGGA TAAGCCTCATC AGAGAGCAAG
4981 CCGAGAATAT CATCCACCTG TTTACCCTGA CCAATCTGGA AGCCTCCTCC GCTCTCAAGT
5041 ATCTGATCAC CCATACAGCC CGGAGATATC AGACAGACAC CAAAGAGGGA AACTGGAAGA
5101 CCCTGATCCA CGAGACGAC TACGGCCTGT AGCAGACAG CAGTGACCTG TCTCAAGTGG
5161 GAGGGCGACAG CCCAAGAAGA AAGAAGAAGG TGGAGCGCGA Cgggccggcc GAAATCCGGGC
5221 GCCGCCCGGC GCTTGAGCTG TTTGGATCTG CACATGGTG GGTCTGAGCT GCCGAGACTA
5281 TTGAACCTGGG ATATGTTGGGA AGCAGCGCAT TGGATGACTT TGATCTGGAC ATGCTGGGCT
5341 CCGATCTCTCT GGACGATTTCC GATCTCGATA TGGATTggtc agacgactcg gatgatttcg
5401 accttgATAT gttgGGAAGC GATGCCCTTG ATGATTTCCA TTTGGATATG TTGATATCCG
5461 AATgATTAAT TAAAGATTTG ACCAGCTTTT CTGGTACAAA GTGGTGATGA TCCAGACAG
5521 TGGCGGGCAGC TCGATGCTAG AGGGCCCCGC GTTCAAGAGT AAGCCTATCC CTAACCCCTT
FIG. 32B (cont.)

7261 cggcgccccc tgcgctgaca gccggaacag ccagccgctca gacgacccga tgtgttgttg
7321 tgcccaagtc aagccggaat gaacctccac ccaagccggc ggaaacccgt cgtgcaatcc
7381 atcttggtca atcatataat tattgaagca tttaaaaaggg ttcttcctgt cccggtctcc
7441 tcccatgcat g (SEQ ID NO:487)

FIG. 32C

pAC93: dCas9VP160 on pmax expression vector

1 tcaatatttg ccattagcca tattattcat tgttatata tgcgttcatca atattggtctga
61 ttggccattg cattgcatgt atcttatata taatatgttac atttatattg gctcattgtcc
121 aatataagcc ccagttttggc atttgattatt gactagttat taatagtaat caattacggg
181 gcgtcggctga ccggcacaac gcacgctgcc atgtgagcgt caataagcag gtttcccat
241 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
301 atcctgacgg attaggcaatt tttctggaag ggttattgct caattacgtcc gtttacgcgt
361 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
421 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
481 gcgtcggctga ccggcacaac gcacgctgcc atgtgagcgt caataagcag gtttcccat
541 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
601 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
661 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
721 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
781 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
841 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
901 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
961 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
1021 ccaccttgga gcggcggtgc attttgctga tccggttaaa tctacttacgg ctttaaagccc
| 4441 | GGGACCCCTAA GAACTACGGC | GGGCTTGACA GCCCCACTCGT GGGCTATTCT GTGCTGGTGG |
| 4501 | TGGCCAAAGT GAGAAGGACG | AAGTCAAGAG AACTAGAGAG TGGTAAAGAG CTGCTGGGGA |
| 4561 | TCACCATCAT GGAAGAAGAAG | AGCCTCGAAG AAGATACCAT CGACTCTCTTGG GAAGGCAAGG |
| 4621 | GCTACAAGAAG AGTGAAAGAAAG AGCCTGATCA TAAGACTGCC TAACTGGTAC CTGGTCGAGC |
| 4681 | TGGAAGACCGG CCGAAAGAGA AGTCTGGCTTG CTCGCCGGCGA AAGGACAGAG GAAACGAAGAC |
| 4741 | TGCCCTGCCC CTCCAATATG GTGACTATGCT GGACTGGGCC CAGCCACTAT GAGAACGGTG |
| 4801 | AGGGCTCCCT CGAGAGGATAAG ACAGCAAGAAG AGCTGTGTCGAG TGGGCAGAACCA AAGCACACTA |
| 4861 | TGGACAGAGAT CATAGAGCTCT AGTACGCTGAG TCTCAGACTAGG CAGCACTAAGG CCTAGCTAGG |
| 4921 | ATCTGGACAA AGTGGCTGGCT CGCCTACAACAGG AAGCAGGCGA AAGCCATGCTT CAGAGCAGG |
| 4981 | CCGAGAATAT CACCACACTTT TTTACCTGAG CGAATGTGGG AGCCCCCTGCC CCTTTCAAGTG |
| 5041 | ACTCTTGACAC CACCAGACAG GCGAGAAGAGT ACACCAAGAC CAAAGAGGCTG CTGGACGCCCA |
| 5101 | CCCTGATCCA CCAGAGACTC ACCGGCTTGG AGAGGACGACG GACGACGCTTGG TCTGAGCAGT |
| 5161 | GGGCGGACAG CCCCCAGAGA AAGGAAGAACG TGGAGGGGAG Cgggcccgggc GGATCCGGGC |
| 5221 | GCGCGGACCGC GTCTGGAGCTT TGCGATCTCCTG ACCAGCTGGG TCTGATGTTCC CTCGATGACT |
| 5281 | TTGACTGTTGA TATGTTGGAA AGCCGACGATG TGGATGACTT TGATCTGGAC ATGTCGCGCTG |
| 5341 | CGCATGCTTGG GGAGCTTTCT GATCTAGATA TGTTACgttc agacgcacctgt gatgttttcg |
| 5401 | accttgatat gttgGGAAGC GATGCGCTTG ATGATTCTAG CAGGACATCTT CTCggcagcg |
| 5461 | acgacctggtg cgatttccatgct ctggagactgc tcggGTCGCGA TGGCTTGGAGG ATTTTCACTT |
| 5521 | TGGAATATGCTG Cgggagtgtat gccccgtgacct acctttgacct gacacagctg GGCTCCGGATG |
| 5581 | CGCTGATGTA CTTGGATTGG TATGATGTTTG ATATCGATCG aTTAAATTAAGG AATTCGACCC |
| 5641 | AGCTTTCTTG TCAAAGTTGG TTGATATCCA GCACAGTGCC GCAGCGCTGCA GTCTAGAGGG |
| 5701 | CCCGGGTCCTC GAAGGTAAGC CTATCTCTAA CCTCTCCTCTG GCTCTGCTGATC TACGCGGTAC |
| 5761 | CGGTAGGCTCA TGGTTTTTGGG GATAGGTTGACG CAGAAACACC aatcagagaatg ccagtaaaag |
| 5821 | aatgtcttttatt ttgtagaaatt ttggtccttttta ttgcatcatt gtaaaccatt ataagctgca |
| 5881 | ataaacaactt tcaacacacc atatggctacct atttatattt tcaggttcag gggagggttg |
| 5941 | gggaggttttt ttaaacagcaag taaacacctt acaaatgtgg tacttaagacg gggagacca |
| 6001 | aagggccgaga cgttaaggcc tccagtgcagc tgtgagcaca aggccaccaag aaggccagga |
| 6061 | accgtaaaaa ggccgcgttg ctggctggcttt tccataggtctt ccgccccccct gacgagcatc |
FIG. 32C (cont.)

6121 aaaaaactcgc acgcctcaagtt cagaggtgagc gaaacccggac aggaactataaa agatacagg 
6131 cggttcccctcc ctcggctgcct ctcctgctcc gaccctgcggc cttaccggagct 
6241 acctgctgcgac ttcacctccctc tcgggaagcg ttgctgcttc tcataatgtca gcgtgtagcg 
6301 atctcagtttc ggtgtagttc gcctcgctca agctgggtgctct tgtgcaccagaa ccccccccggc 
6361 acgccgacccg ctcgcctctta ccgggttaacct ctgctctgtaa gtcacaccccg qtaagacagcg 
6421 acttatcgcc actggcagaca gcacactggta acag gtaccatgag cagctgcagcct tagtgtaggc 
6481 gttgcctacaga gttctttgaag tgtgtgctgct aactacgcgta cactagaagctg acagtattgg 
6541 gtatctgctgcgt ctcgcctgaaag ccagttactcag tcggaaagagc agttggttacg tcttgctccg 
6601 gcacacaacac cagctggtga gcggttggcttt ttttgggggctc aagcagcagata ttaagcgcagc 
6661 aaagaaggaactctcaagaa ctccctctctct ccggtaaggaat aacgtgtgaac gctcgtgagaa 
6721 cggaaactctca gtttaagggactttggtctact cgggtctcag aaggaactcgta cagaagaggcxg 
6781 atagagggcag atgcgtctgcgt agtgggagctc ggctgatccgc tasagctacg ggaagcttcgc 
6841 aggccatttcgc ccgcacagct ctctcaagagtc atcagcggat gcgaagcctca tgtcctgtgtaa 
6901 gcgcctgcgcct caccagcgcc ggcaccagtc gtagaatcata gaaagcggcc catttttcac 
6961 catgatatgtt gcagagcaggg gtagggcctctgcc ggtcagcagc agattcctccg cgtcgggca 
7021 ggtcgcccttg gaagctgccgca acaggtcggcc tggcgtgatcg cccctgtactc ctctgcgtgc 
7081 atctctccgca tcgacaaagac gcgcctcccttc cggagattctg gctgcgctgag cgcattgtgctt 
7141 cgctttgctgg gcagatgtcgc atcaagctgcga atcagccgcct gctgacgctcgc gcaatgcagt 
7201 agcacattcctgc gatctccctct gcagagggagct aagtggtagatt gcacaggagatt cctgcccccgg 
7261 cacttcgcctgc aacctcagccg atctccagcttc cggccttcagc aacacgctcgc gcacagctgcc 
7321 gcagaggaagcg cccgcgcggtgc cacgcagcagct gcagcgcgcct ggcctcgtctgt gcagtcgattg 
7381 caggggacgagc cagagtctcgct tttgacaaaa aagagccggc gcgccccgctc gctgacagccg 
7441 gacacagcgcc gcatacgaggc gcgctgattgtg ctgctgctgcct cagtcgatagc cgaatagcct 
7501 ccctcgcacca gcgcccgggag aacctgcgttc atctccatctgc tgttcaatcctc taatatattt 
7561 gaagcattttta tcagggcttgc ttcgcgctccgc gtctctcctcc atgcatg (SEQ ID NO: 488)