

MODIFIED NUCLEIC ACID EDITING SYSTEMS FOR TETHERING DONOR DNA

Technical Field

[001] The technology relates to compositions for tethering donor DNA to a nuclease the use of those compositions for improving the efficiency of *in vivo* gene editing.

Related Application

[002] This application claims priority to Australian provisional patent application No 2018900990 filed 25 March 2018 which is herein incorporated by reference in its entirety.

Background

[003] A number of genome editing technologies are known including Transcription Activator-Like Effector Nucleases (TALENs), the CRISPR-Cas9 system (Clustered, Regularly Interspaced, Short Palindromic Repeats and CRISPR associated protein 9), and Zinc finger nucleases (ZFNs). TALENs, CRISPR-Cas9 protein and ZFNs use endonucleases to initiate double-strand breaks (DSBs) at almost any target sequence in genomic DNA and can be used for gene knockouts, gene knock-ins, gene tagging, and correction of genetic defects.

[004] The type II bacterial clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein 9 (Cas9) system is a tool for the targeted introduction of mutations into cellular DNA. Well-designed single guide (sg) RNAs induce Cas9-mediated double stranded breaks (DSBs) at desired target sites in cellular DNA while minimizing effects at other locations. Double stranded breaks stimulate DNA repair by at least two distinct mechanisms non-homologous end joining (NHEJ) and homology directed repair (HDR). Cas9-mediated modification of cellular DNA by NHEJ can reach efficiencies of 20–60% but because NHEJ is error-prone and introduces unpredictable patterns of insertions and deletions it is only suitable for introducing small random mutations. Co-application CRISPR-Cas9 with a single-stranded or double-stranded DNA template homologous to the sequences flanking the cleavage site on the cellular DNA enables precise genome editing by HDR-mediated incorporation of an exogenous, or donor DNA fragment. However, the frequency of HDR is inherently low and the efficiency of insertion of a donor DNA using this strategy is only 0.5–20%.

[005] TALENs are fusions of transcription activator-like (TAL) proteins and a Fok I nuclease. TAL proteins are typically composed of 33-34 amino acid repeating motifs with two variable positions that have a strong recognition for specific nucleotide sequences. By assembling arrays of TALs and fusing them to a Fok I nuclease, specific cutting of the

genome can be achieved. When two TALENs bind and meet, the Fok I domains induce a double-strand break which can inactivate a gene, or can be used to insert DNA of interest. TALENs are able to modify chromosomes with efficiency of up to about 33%.

[006] ZFNs are a class of engineered DNA-binding proteins that allow targeted genome editing of the genome by creating double-strand breaks in DNA at desired locations. ZFNs consist of two functional domains, a DNA-binding domain comprised of a chain of Zinc-finger modules, each recognizing a unique DNA hexamer. Multiple Zinc-fingers can be assembled to form ZFN with specificity for a sequence of 24 bases or more. The second functional domain is the nuclease domain of Fok I. Using ZFNs can result in single or biallelic edits occurring at an efficiency of 1–20% of clone population.

[007] The present inventor has developed compositions and methods that utilise the target specificity of gene editing systems such as CRISPR-Cas9, TALENs and ZFNs to tether a donor DNA to a desired target DNA sequence.

Summary

[008] In a first aspect, there is provided a composition comprising a composition for tethering donor DNA to a nuclease, the composition comprising a nucleic acid comprising donor DNA and a consensus sequence for a DNA binding domain; and at least one of:

a fusion protein comprising a nuclease coupled to a DNA binding domain for binding the consensus sequence; and

a nucleic acid encoding the fusion protein.

[009] The nuclease may be a Cas, a Transcription activator-like effector nuclease (TALEN), a meganuclease, or a Zinc Finger. In one embodiment there is a Cas protein, for example Cas9

[010] The fusion protein may further comprise a nuclear localization sequence.

[011] The composition may further comprise a guide RNA that interacts with the Cas protein and a target DNA sequence.

[012] The consensus sequence may comprise the Lac operator (SEQ ID NO: 66), the TRP operator (SEQ ID NO: 68), the TET operator (SEQ ID NO: 67), the GAL-4 binding site (SEQ ID NO: 1), or the IHF binding site (SEQ ID NO 2).

[013] The consensus sequence may comprise a sequence with at least 80%, 85%, 90%, 95% or at least 99% identity to the Lac operator, the TRP operator, the TET operator, the GAL-4 binding site, or the IHF binding site

[014] The DNA binding domain may comprise the LAC repressor, TET repressor, TRP-repressor, GAL-4, or IHF, or a portion thereof sufficient to bind the consensus sequence.

[015] The DNA binding domain is the LAC repressor, preferably amino acids 43-403 of SEQ ID NO 9.

[016] The nuclease may be coupled to the DNA binding domain via a linker. The linker may comprise a sequence selected from any one of SEQ ID Nos: 3 to 7, a GGS linker, or amino acids 404-419 of SEQ ID NO 9.

[017] In one embodiment the fusion protein comprises the LAC repressor and Cas9.

[018] The composition may comprise a vector, wherein the vector comprises either or both of:

- a. the nucleic acid comprising the donor DNA and the consensus sequence for a DNA binding domain; and
- b. the nucleic acid encoding the fusion protein.

[019] The vector may further comprise a nucleic acid sequence encoding a guide RNA that interacts with the Cas9 and a target DNA sequence.

[020] In some embodiments the fusion protein is nuclease deficient.

[021] In a second aspect there is provided an isolated host cell comprising the composition of the first aspect.

[022] In a third aspect there is provided a method for editing DNA in a cell, the method comprising;

contacting the cell with the composition of the first aspect under conditions suitable for the interaction of the fusion protein with a target DNA sequence.

[023] In a third aspect there is provided a method for editing DNA in a cell, the method comprising

- a) contacting the cell with the composition of claims 15 under conditions suitable for the interaction of the fusion protein with a first target DNA sequence; and
- b) contacting the cell with a nucleic acid editing system adapted to edit the genomic DNA at a second target DNA sequence, under conditions suitable for nucleic acid editing.

[024] The target DNA sequence may be selected from genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

[025] In one embodiment the efficiency of editing is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% at least 95%, or at least 99%.

[026] The cell may be in a subject, preferably a human subject.

[027] In a fourth aspect there is provided a kit comprising:

a nucleic acid comprising donor DNA and a consensus sequence for a DNA binding domain; and at least one of

a fusion protein comprising a nuclease coupled to a DNA binding domain for binding the consensus sequence; and
a nucleic acid encoding the fusion protein.

[028] In one embodiment the fusion protein in the kit comprises a Cas protein, a Transcription activator-like effector nuclease (TALEN), a meganuclease, a Zinc Finger or a MADzyme. In one embodiment the Cas is Cas9.

[029] Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

[030] Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed before the priority date of each claim of this specification.

[031] In order that the present invention may be more clearly understood, preferred embodiments will be described with reference to the following drawings and examples.

Brief Description of the Drawings

[032] **Figure 1** is a depiction of two embodiments of the composition disclosed herein comprising a modified nucleic acid editing system.

[033] **Figure 2** is a map of a plasmid vector containing an active Cas 9 gene sequence and its guide RNA capable of directing binding of the Crispr/Cas9 complex to a genomic target DNA sequence, donor DNA sequence that can be used for repair of Crispr/Cas9 induced double strand breaks.

[034] **Figures 3 and 4** are maps of plasmids used in a binary system where the Crispr/Cas9 gene sequence containing the lac repressor fusion and guide RNA are present on a tether expression vector and the donor DNA sequence is present on a second tethered gene targeting vector.

[035] **Figures 5 and 6** are maps of plasmids for generating an *in vitro* system whereby purified Crispr/Cas9 would be used to bind a gene targeting vector prior to transfection into cells.

[036] **Figure 7** is a workflow diagram for construction of an *in vivo* tethered gene-targeting plasmid, where A represents the synthesis of U6 expression cassette containing Ug

promoter with a central Sspl recognition sequence and flanking Sspl compatible overhangs that do not reconstruct the plasmid Sspl recognition sequence. B represents the synthesis of CMV promoter-SV40-polyA cassette containing central NotI cloning site and Eco109I compatible overhangs. C represents Synthesis of gene encoding Cas9-lac repressor fusion protein with NotI compatible overhangs. D represents synthesis of gene encoding RNA complementary to genomic target DNA sequence with Sspl compatible overhangs. E represent PCR amplification of donor DNA sequence with oligonucleotide containing flanking Sall restriction endonuclease recognition sequences and digestions with Sall.

[037] **Figure 8** is a workflow diagram for the construction of pGT1. A. Construction of lactose repressor gene with inactivated BbsI restriction endonuclease sites. B. Three overlapping DNA fragments comprising Flag/SV40 NLS, lacI and Cas9 were made, and these were cloned into pSpCas9 BB-2A-GFP(px458), (C) to create pGT1. The vector map of pGT1 is shown in D.

[038] **Figure 9** is a vector map of the ptet repressor plasmid used in the construction of pGT9. **Figure 10** is a workflow diagram for the construction of pGT9. A: cloning of a tet repressor cassette into pUC57 to create ptet repressor. B: three polymerase chain reaction products, Flag/SV40 NLS, lacI and Cas9, were generated and these were cloned into pSpCas9 BB-2A-GFP(px458), (C) to create pGT9. The vector map of pGT9 is shown in D.

[039] **Figure 11** is an overview of a donor fragment, showing a 500 base pair donor fragment with delF508 Mutation and lacO or tetO Recognition Sequences

[040] **Figure 12** is an overview of the use of the compositions described herein for homology directed repair by homologous recombination between the genomic CFTR target Cas9-induced double strand break and donor fragment to catalyze transfer of the delF508 DNA sequence to the genomic target.

[041] **Figure 13** illustrates that combination of pGT1 vector (lactose repressor-Cas9 fusion) with the 130117 guide (genomic target forward strand) and lactose operator on the 3' end of donor DNA (Figure 13, lane 7) demonstrates higher gene editing efficiency as compared to the px458 vector with unmodified Cas9 and same donor DNA fragment

[042] **Figure 14** illustrates that gene editing with the pGT9 tetracycline repressor-Cas9 fusion vector, guide sequences, and donor fragments did not yield appreciably better gene editing frequencies than the control px458 vector

Description of Embodiments

[043] The present disclosure provides compositions, methods, and kits for tethering donor DNA to a DNA target sequence with a modified nucleic acid editing system. The disclosure provides for improved efficiency of *in vivo* cellular DNA modification (such as gene editing)

using a modified nucleic acid editing system, such as CRISPR-Cas9. One embodiment of this process is illustrated in Figure 1. The modified gene editing system is adapted to bind to a nucleic acid comprising donor DNA and tether the nucleic acid to a specific site on the cellular DNA at or near to the nucleotide sequence to be edited.

Modified Nucleic Acid Editing Systems (nuclease-DNA binding domain fusions)

[044] The compositions and methods described herein can include a nuclease of any nucleic acid editing system capable of site specific binding. For example, useful nucleases include Cas nucleases, TALENs, meganucleases, ZFNs and MADzymes. The nuclease is modified by combining it with a DNA binding domain.

[045] The nuclease and the DNA binding domain may be joined via linker. Suitable linkers include for example, linkers comprising the sequences (Gly-Gly-Gly-Gly-Ser)_n, (Gly)_n, or (Gly)_nS, (EAAAK)_n, (AP)_n, (XP)_n, or A(EAAAK)_n where n is any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 and X is any amino acid. Other suitable linkers comprise the sequences KESGSVSSEQLAQFRSLD, EGKSSGSGSESKST, GSAGSAAGSGEF, KESGSVSSEQLAQFRSLE, or GGSAGGSGSGSSGSSGASGTGTAGGTGSGSGTGSG.

[046] Cas systems are divided into three major types (type I, type II, and type III) and twelve subtypes, which are based on their genetic content and structural differences. However, the core defining features of all CRISPR-Cas systems are the Cas genes and their proteins: cas1 and cas2 are universal across types and subtypes, while Cas3, Cas9, and Cas10 are signature genes for type I, type II, and type III, respectively.

[047] Any Cas may be used. For example the Cas nuclease may be Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10. In some embodiments the Cas is Cas9.

[048] Cas9 (CRISPR associated protein 9) is an RNA-guided DNA nuclease used to induce site-directed double strand breaks in DNA for the gene inactivation or the introduction of heterologous genes through non-homologous end joining and homologous recombination respectively. Systems and nucleic acid sequences for expressing Cas9 are commercially available. The Cas9 may be codon optimized for a human or other mammalian system. The Cas9 protein may contain a nuclear localization signal at the C-terminus. The RNA encoding Cas9 may be capped and polyadenylated to support expression in mammalian cells, and may contain modifications to reduce immune stimulation. The amino acid sequence and encoding nucleic acid sequence for Cas9 and functional derivatives and homologs which can be used in the compositions and methods are known in the art.

[049] The Cas9 may be delivered in conjunction with a guide RNA (gRNA) that directs the editing system to the nucleotide sequence recognized by the gRNA. In general, a gRNA can be designed to target any nucleotide sequence. The gRNA structure is disclosed in, for example, Ran FA, Genome editing using the CRISPR-Cas9 System. PNAS 8(1 1):2281-308 (2013); and Pyzocha et al., RNA-guided genome editing of mammalian cells. Methods Mol. Biol. 1 1 14:269- 77 (2014), which are hereby incorporated by reference in their entirety. Generally for Cas9, gRNAs guide the Cas9 to the complementary 20 nucleotide sequences with a downstream NGG protospacer-adjacent motif (PAM).

[050] The CRISPR-Cas9 system including the construction of guide sequences is further disclosed in US 8,697,359, which is hereby incorporated by reference in its entirety.

[051] In place of a CRISPR-Cas9 system, alternate nucleic acid editing systems may be used. For example, suitable systems include any CRISPR/cas system (e.g., any Cascade-like CRISPR/cas, Type I CRISPR cas, Type II CRISPR/cas, and type III CRISPR/cas), zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), meganucleases, homing endonucleases, and MADzymes (for examples the Human or *E. coli* MADzymes encoded by SEQ ID Nos: 69 and 70, respectively).

[052] The DNA binding domain component of the modified nucleic acid editing system can be any sequence specific high affinity DNA binding domain. The term 'DNA binding domain' as used herein refers to any complete protein or fragment thereof which can bind DNA. Accordingly the term 'DNA binding domain' includes complete proteins such as the LAC repressor and fragments of the protein which retain the DNA binding

[053] In some embodiments the DNA binding domain is not mammalian in order to reduce off target effects. For example the DNA binding domain may be from a bacteria or yeast. DNA binding domains useful in the compositions and methods described herein can be selected from the group consisting of be the LAC repressor, TET repressor, TRP-repressor, GAL-4, or IHF. Each of these bind specific consensus sequences that are known in the art.

[054] The consensus sequence may be the Lac operator (e.g. SEQ ID NO: 66), the TRP operator (e.g. SEQ ID NO: 68), the TET operator (e.g. SEQ ID NO: 67), the GAL-4 binding site (5'-CGG-N₁₁-CCG-3') or the IHF binding site (5' – WATCAANNNTTTR - 3'). W is A or T, and R is A or G, and N is any nucleotide.

[055] The nucleic acid editing systems can be present in the compositions disclosed herein in the form of purified proteins or in the form of nucleic acids encoding the nucleic acid editing system. The nucleic acid may be a vector, for example a plasmid vector comprising sequences encoding the nucleic acid editing system operably linked to a constitutive or inducible promoter.

[056] The term "vector" includes a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors used to deliver the nucleic acids to cells as described herein include vectors known to those of skill in the art and used for such purposes. Certain exemplary vectors may be plasmids, lentiviruses or adeno-associated viruses. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, lentiviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are expression vectors capable of directing the expression of genes to which they are operatively linked. Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. Recombinant expression vectors can comprise one or more nucleic acids encoding a modified nucleic acid editing system in a form suitable for expression of the nucleic acid in a cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression. That is the regulatory elements are operatively-linked to the nucleic acid sequence to be expressed. 'Operably linked' means that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[057] 'Regulatory element' includes promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Regulatory elements are known in the art. Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide

sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes).

[058] Regulatory elements may also direct expression in a temporal- dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector may comprise one or more pol III promoter (e.g. 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g. 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g. 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer), the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter and Pol II promoters described herein. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I; SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β -globin. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts (e.g. guide RNA), proteins, or peptides, including fusion proteins (such as the modified nucleic acid editing systems) or peptides.

[059] The vector may include one or more terminator sequences. A terminator sequence includes a section of nucleic acid sequence that marks the end of a coding sequence during transcription.

[060] The vector may include one or more sequences encoding an epitope tag or reporter gene sequences. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, betaglucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and auto-fluorescent proteins including blue fluorescent protein (BFP). The epitope tag, reporter gene

or both may be expressed from the vector as a fusion with the modified nucleic acid editing system (nuclease-DNA binding domain fusion).

[061] Alternatively or in addition the compositions may comprise a mRNA encoding the fusion protein.

[062] The mRNA can be modified, and the modification selected from one or more of modifications of the phosphate backbone (e.g., phosphorothioate linkages or boranophosphate linkages), ribose ring modifications such as 2'-O-methyl and/or 2'- fluoro and/or 4'-thio modifications, and locked or unlocked nucleic acids. Other modifications may include pseudouridine, 2-thiouridine, 4-thiouridine, 5-azauridine, 5-hydroxyuridine, 5-aminouridine, 5-methyluridine, 2-thiopseudouridine, 4-thiopseudouridine, 5-hydroxypseudouridine, 5-methylpseudouridine, 5-aminopseudouridine, pseudoisocytidine, 5- methylcytidine, N4-methylcytidine, 2-thiocytidine, 5-azacytidine, 5-hydroxycytidine, 5-aminocytidine, N4-methylpseudoisocytidine, 2-thiopseudoisocytidine, 5-hydroxypseudoisocytidine, 5-aminopseudoisocytidine, 5-methylpseudoisocytidine, N6-methyladenosine, 7-deazaadenosine, 6-thioguanosine, 7-deazaguanosine, 8-azaguanosine, 6-thio- 7-deazaguanosine, 6-thio-8-azaguanosine, 7-deaza-8-azaguanosine, and 6-thio-7-deaza-8- azaguanosine.

[063] In some embodiments the modifications are selected for one or more of the following: reduce immune stimulation, RNA stabilization, improve expression of the encoded protein. For example, the RNA may have a combination of 2-thiouridine and 5-methylcytidine to reduce immune stimulation through pattern recognition receptors such as TLR3, TLR7 and TLR8. In some embodiments, the mRNA has one or more pseudouridine to stabilize the mRNA against cleavage, and improve expression rates.

[064] The modified nucleic acid editing systems also comprise a DNA binding domain. The DNA binding domain facilitates the localisation of the modified nucleic acid editing system to cellular DNA.

[065] Construction of the vectors disclosed herein is by standard methods known in the art such as ligation of synthetic nucleic acids, or nucleic acids produced by, for example PCR, into a plasmid that has been cut by one or more site-specific nucleases.

[066] There are alternative strategies for producing the vectors disclosed herein. In one approach, the entire vector(s) are synthesized *de novo* using a commercially available service, for example by a company that specialises in the synthesis of large DNA molecules.

[067] Alternatively, a combination of classical cloning techniques and synthetic biology to modify a standard laboratory plasmid such as pUC19. In this approach, gene cassettes encoding the Cas9-lac operon fusion gene, guide RNA, and donor DNA sequences would

be synthesized *de novo* and individually cloned into unique restriction endonuclease sites present in the vector backbone using T4 DNA ligase. Donor DNA sequences are amplified by polymerase chain reaction to generate products that are cloned into the vector backbone. Modification of Cas9-lac repressor fusion proteins to add other DNA binding domains or epitopes for affinity purification of modified Cas9 proteins can be performed by recombinant PCR followed by ligation into restriction endonuclease sites in the vector backbone. A workflow illustrating this process for preparing the plasmid vectors disclosed herein is shown in Figure 7.

Nuclease deficiency

[068] In some embodiments the nucleic acid editing system is nuclease deficient. In these embodiments the nuclease deficient nucleic acid editing system is used to tether the donor DNA near to a target nucleic acid sequence so that donor DNA is available for use by an additional nucleic acid editing system.

[069] For example the nuclease deficient gene editing system has reduced or eliminated nuclease activity, alternatively the nuclease activity is absent or substantially absent within levels of detection. In some embodiments the nuclease activity of the gene editing system may be undetectable using known assays, i.e. below the level of detection of known assays.

[070] Nuclease deficient gene editing systems can be prepared by those skilled in the art using standard molecular biology techniques. Typically this involves deleting or altering one or more amino acids crucial for nuclease activity to substantially eliminate or eliminate nuclease activity.

[071] In some embodiments the Cas9 is a nuclease-deficient Cas9. A nuclease-deficient Cas9 may be one in which one or more amino acids in Cas9 are altered or removed. For example a nuclease deficient Cas9 may be generated by removing or mutating one or more of the amino acids D10, H840, D839 and N863 (See Jinke et al., *Science* 337, 816-821 (2012). For example one or more of these amino acids may be deleted or substituted with alanine or glycine to substantially eliminates or eliminates nuclease activity.

Donor DNA

[072] The term "donor DNA" includes a nucleic acid sequence which is to be inserted into cellular DNA (such as of genomic DNA, mitochondrial DNA, or viral DNA). The donor nucleic acid sequence may be expressed by the cell. The donor nucleic can be exogenous, foreign to the cell or non-naturally occurring within the cell.

[073] The donor DNA is associated with a sequence that can be bound by a DNA binding domain. For example the donor DNA may be contiguous with a consensus sequence for a

DNA binding domain or may be present on the same vector as the a consensus sequence, or may be present on the same polynucleotide as the consensus sequence.

Target Nucleic Acid Sequence

[074] A target nucleic acid sequence includes any nucleic acid sequence, such as a genomic nucleic acid sequence or a gene to which a nuclease of a nucleic acid editing system can co-localise. Target nucleic acids include nucleic acid sequences capable of being expressed into proteins. According to one aspect, the target nucleic acid is genomic DNA, mitochondrial DNA, plastid DNA, viral DNA, or exogenous DNA.

[075] One of skill in the art will readily be able to identify or design guide RNAs, TALENs, ZFNs, meganucleases and homing endonucleases which co-localize to a target nucleic acid sequence.

Methods

[076] The compositions and vectors described herein can be used in for editing DNA in a cell. The methods comprise contacting the cell with a composition comprising a modified gene editing system or vector encoding a modified gene editing system under conditions suitable for the interaction of the modified nucleic acid editing system with a target DNA sequence. The methods also comprise use of a nuclease-deficient modified nucleic acid editing system that interacts with a first target DNA sequence and a conventional nucleic acid editing system to edit the DNA at a second target DNA sequence.

[077] In order to increase the efficiency of DNA editing the donor DNA is positioned close to the target DNA sequence so that it is readily available for nucleic acid editing.

[078] In embodiments using a nuclease-deficient modified nucleic acid editing system this is achieved by spacing the first and second target sequences so that the donor DNA is closed to the conventional nucleic acid editing system when it is colocalised with its target sequence.

[079] For example the first and second target sequences are about 75 to 150 base pairs apart, about 150-250, 250-350, 450-550, 550-650, 650-750, 750-850, 850-950, 950-1050 base pairs apart or about 1-1.5 kb apart.

[080] The methods require delivery of the compositions or vectors to the cell. Methods of non-viral delivery of nucleic acid vectors, RNA or proteins include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, agent-enhanced uptake of DNA, nanoparticles, and electroporation/nucleofection. . Lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides are known. Delivery can be to cells (e.g.

in vitro or *ex vivo* administration) or target tissues (e.g. *in vivo* administration). The term native includes the protein, enzyme, RNA, or guide RNA species itself as well as the corresponding nucleic acid encoding the same.

Delivery Vehicles

[081] Delivery vehicles for the compositions and vectors disclosed herein provided herein may be viral vectors or non-viral vectors. In some embodiments, the modified nucleic acid editing system is provided in a viral vector or a non-viral vector. In other embodiments, the guide RNA is provided in a viral vector, and the modified nucleic acid editing system is provided in a non-viral vector. In still other embodiments, the guide RNA is provided in a non-viral vector and the modified nucleic acid editing system is provided in a viral vector

[082] In some embodiments, the viral vector is selected from an adeno-associated virus (AAV), adenovirus, retrovirus, and lentivirus vector. While the viral vector may deliver any component of the system described herein so long as it provides the desired profile for tissue presence or expression, in some embodiments the viral vector provides for expression of one or more of the modified nucleic acid editing system, guide RNA and optionally the delivers the donor DNA. In some embodiments, the viral delivery system is adeno- associated virus (AAV) 2/8. However, in various embodiments other AAV serotypes are used, such as AAV1 , AAV2, AAV4, AAV5, AAV6, and AAV8. In some embodiments, AAV6 is used when targeting airway epithelial cells, AAV7 is used when targeting skeletal muscle cells (similarly for AAV1 and AAV5), and AAV8 is used for hepatocytes. In some embodiments, AAV1 and 5 can be used for delivery to vascular endothelial cells. Further, most AAV serotypes show neuronal tropism, while AAV5 also transduces astrocytes. In some embodiments, hybrid AAV vectors are employed. In some embodiments, each serotype is administered only once to avoid immunogenicity. Thus, subsequent administrations employ different AAV serotypes.

[083] In some embodiments, the delivery system comprises a non-viral delivery vehicle. In some aspects, the non-viral delivery vehicle is lipid-based. In other aspects, the non-viral delivery vehicle is a polymer. In some embodiments, the non-viral delivery vehicle is biodegradable. In embodiments, the non-viral delivery vehicle is a lipid encapsulation system and/or polymeric particle.

[084] In certain embodiments, the delivery system comprises lipid particles. In some embodiments, the lipid-based vector is a lipid nanoparticle, which is a lipid particle between about 1 and about 100 nanometers in size. In some embodiments, the lipid-based vector is a lipid or liposome. Liposomes are artificial spherical vesicles comprising a lipid bilayer.

[085] The lipid-based vector can be a small nucleic acid-lipid particle (SNALP). SNALPs comprise small (less than 200nm in diameter) lipid-based nanoparticles that encapsulate a nucleic acid. In some embodiments, the SNALP is useful for delivery of an RNA molecule. In some embodiments, SNALP formulations deliver nucleic acids to a particular tissue in a subject, such as the liver.

[086] In some embodiments, the guide RNA, the modified nucleic acid editing system (or the RNA encoding the same) is delivered via polymeric vectors. In some embodiments, the polymeric vector is a polymer or polymerosome. Polymers encompass any long repeating chain of monomers and include, for example, linear polymers, branched polymers, dendrimers, and polysaccharides. Linear polymers comprise a single line of monomers, whereas branched polymers include side chains of monomers. Dendrimers are also branched molecules, which are arranged symmetrically around the core of the molecule. Polysaccharides are polymeric carbohydrate molecules, and are made up of long monosaccharide units linked together. Polymersomes are artificial vesicles made up of synthetic amphiphilic copolymers that form a vesicle membrane, and may have a hollow or aqueous core within the vesicle membrane.

[087] [0058] Various polymer-based systems can be used for administering RNA encoding modified nucleic acid editing system. Exemplary polymeric materials include poly(D,L- lactic acid-co-glycolic acid) (PLGA), poly(caprolactone) (PCL), ethylene vinyl acetate polymer (EVA), poly(lactic acid) (PLA), poly(L-lactic acid) (PLLA), poly(glycolic acid) (PGA), poly(L-lactic acid-co-glycolic acid) (PLLGA), poly(D,L-lactide) (PDLA), poly(L- lactide) (PLLA), PLGA-b-poly(ethylene glycol)-PLGA (PLGA-bPEG-PLGA), PLLA-bPEG-PLLA, PLGA-PEG-maleimide (PLGA-PEG-mal), poly(D,L-lactide-co-caprolactone), poly(D,L-lactide-co-caprolactone-co- glycolide), poly(D,L-lactide-co-PEO-co-D,L-lactide), poly(D,L-lactide-co-PPO-co-D,L-lactide), polyalkyl cyanoacrylate, polyurethane, poly-L-lysine (PLL), hydroxypropyl methacrylate (HPMA), polyethyleneglycol, poly-L-glutamic acid, poly(hydroxy acids), polyanhydrides, polyorthoesters, poly(ester amides), polyamides, poly(ester ethers), polycarbonates, polyalkylenes such as polyethylene and polypropylene, polyalkylene glycols such as poly(ethylene glycol) (PEG), polyalkylene oxides (PEO), polyalkylene terephthalates such as poly(ethylene terephthalate), polyvinyl alcohols (PVA), polyvinyl ethers, polyvinyl esters such as poly(vinyl acetate), polyvinyl halides such as poly(vinyl chloride) (PVC), polyvinylpyrrolidone, polysiloxanes, polystyrene (PS), polyurethanes, derivatized celluloses such as alkyl celluloses, hydroxyalkyl celluloses, cellulose ethers, cellulose esters, nitro celluloses, hydroxypropylcellulose, carboxymethylcellulose, polymers of acrylic acids, such as poly(methyl(meth)acrylate) (PMMA), poly(ethyl(meth)acrylate),

poly(butyl(meth)acrylate), poly(isobutyl(meth)acrylate), poly(hexyl(meth)acrylate), poly(isodecyl(meth)acrylate), poly(lauryl(meth)acrylate), poly(phenyl(meth)acrylate), poly(methyl acrylate), poly(isopropyl acrylate), poly(isobutyl acrylate), poly(octadecyl acrylate) (polyacrylic acids), and copolymers and mixtures thereof, polydioxanone and its copolymers, polyhydroxyalkanoates, polypropylene fumarate), polyoxymethylene, poloxamers, poly(ortho)esters, poly(butyric acid), poly(valeric acid), poly(lactide-co-caprolactone), trimethylene carbonate, polyvinylpyrrolidone, polyortho esters, polyphosphazenes, Poly([beta]-amino esters (PBAE), and polyphosphoesters, and blends and/or block copolymers of two or more such polymers. Polymer-based systems may also include Cyclodextrin polymer (CDP)-based nanoparticles such as, for example, CDP-admantane (AD)-PEG conjugates and CDP-AD-PEG-transferrin conjugates.

[088] In one embodiment, nanoparticles are formulated with Cas9 mRNA chemically modified to reduce TLR responses, as disclosed in Kormann et al. Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. *Nat. Biotechnol.* 29: 154-157 (2011). In a further embodiment, the nanoparticles are formulated using controlled microfluidic mixing systems, as disclosed in, for example, Chen et al. Rapid discovery of potent siRNA-containing lipid nanoparticles enabled by controlled microfluidic formulation. *J. Amer. Chem. Soc.* 134:6948-6951 (2012).

[089] In some embodiments, the lipid-based delivery system comprises a lipid encapsulation system. The lipid encapsulation system can be designed to drive the desired tissue distribution and cellular entry properties, as well as to provide the requisite circulation time and biodegrading character. The lipid encapsulation may involve reverse micelles and/or further comprise polymeric matrices. In some embodiments, the particle includes a lipophilic delivery compound to enhance delivery of the particle to tissues, including in a preferential manner. Such compounds may generally include lipophilic groups and conjugated amino acids or peptides, including linear or cyclic peptides, and including isomers thereof.

[090] The lipid or polymeric particles may have a size (e.g., an average size) in the range of about 50 nm to about 5 μ m. In some embodiments, the particles are in the range of about 10 nm to about 100 μ m, or about 20 nm to about 50 μ m, or about 50 nm to about 5 μ m, or about 70 nm to about 500 nm, or about 70 nm to about 200 nm, or about 50 nm to about 100 nm. Particles may be selected so as to avoid rapid clearance by the immune system. Particles may be spherical, or non-spherical.

[091] In some embodiments, the non-viral delivery vehicle may be a peptide, such as a cell-penetrating peptides or cellular internalization sequences. Cell penetrating peptides are

small peptides that are capable of translocating across plasma membranes. Exemplary cell-penetrating peptides include, but are not limited to, Antennapedia sequences, TAT, HIV-Tat, Penetratin, Antp- 3A (Antp mutant), Buforin II, Transportan, MAP (model amphipathic peptide), K-FGF, Ku70, Prion, pVEC, Pep-1, SynB I, Pep-7, I-IN-1, BGSC (Bis-Guanidinium-Spermidine-Cholesterol, and BGTC (B is-Guanidinium-Tren-Cholesterol) .

[0072] In some embodiments, the guide RNA or RNA encoding the modified nucleic acid editing system is modified at the 5' end or the 3' end. In a preferred embodiment, the modification is made at the 3' end of the RNA. The RNA may be modified by conjugating to cholesterol, other lipophilic molecules, polymers, peptides, antibodies, aptamers, and/or small molecules. In some embodiments, the RNA is conjugated to a N- acetylgalactosamine (GalNAc). GalNAc binds the asialoglycoprotein receptor (ASGPR) on hepatocytes, and therefore can be used to target an RNA to the liver. In some embodiments, the RNA is conjugated to a trivalent targeting ligand, e.g., triantennary GalNAc. Such conjugates comprise an RNA conjugated at the 3' terminus to three GalNAc molecules.

[092] The delivery vehicles (e.g. conjugates, viral or non-viral vectors, or any combination thereof) may be administered by any method known in the art, including injection, optionally by direct injection to target tissues. In some embodiments, the guide RNA, modified nucleic acid editing system, and, optionally, donor DNA are administered simultaneously in the same or in different delivery vehicles. In other embodiments, the guide RNA and modified nucleic acid editing system and, optionally, donor DNA are administered sequentially via the same or separate delivery vehicles. In some embodiments, the guide RNA and/or donor DNA is administered 1, 3, 5, 7, 10, 14, or 30 days prior to administration of the modified nucleic acid editing system, such that the guide RNA and/or donor DNA accumulates in the target cell or tissue prior to administration of the modified nucleic acid editing system. In some embodiments, the guide RNA, donor DNA and/or nucleic acid editing system is administered in a plurality of doses, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more doses. In various embodiments, the gRNA, donor DNA and/or nucleic acid editing system is administered over a time period of from one day week to about a month.

[093] In one embodiment, one or both of the guide RNA and donor DNA, are provided in an AAV vector that is administered to the tissue or cell prior to administration of the modified nucleic acid editing system. In a further embodiment, the AAV vector comprising the gRNA is administered 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 days prior to the administration of the nanoparticle modified nucleic acid editing system, to allow expression of the guide RNA from the AAV vector. In a yet further embodiment, the modified nucleic acid editing system

is administered multiple times, for example, once every 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 days.

[094] In another embodiment, the donor DNA is delivered via an AAV vector, and is injected 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 days prior to the administration of either or both of the modified nucleic acid editing system and the guide RNA.

[095] In particular embodiments, either or both of guide RNA and donor DNA are provided in an AAV vector that is administered first, and the modified nucleic acid editing system is administered subsequently in a lipid-based delivery vehicle in one or more doses.

[096] In another embodiment, each component of the compositions described herein (e.g., the modified nucleic acid editing system, guide RNA and donor DNA) are each delivery using a different vehicles, alternatively one or more components may be used with the same deliver vehicle. In a further embodiment, the modified nucleic acid editing system, guide RNA, and donor DNA, are administered at multiple time points, for example, every 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 days. In another embodiment, the administration of the modified nucleic acid editing system, guide RNA and donor DNA are administered at different time points.

[097] In some embodiments, expression of the modified nucleic acid editing system is transient. In some embodiments, such transient expression of the modified nucleic acid editing system minimizes off-target effects. For example, expression of the modified nucleic acid editing system is controlled via selection of the delivery vehicles and/or promoters.

[098] In some embodiments, the present disclosure provides compositions and methods that allow for increased safety and/or efficacy of conventional nucleic acid editing systems. Advantageously, the methods disclosed herein provide for repeated dosing with conventional and modified nucleic acid editing systems such that the efficiency of gene editing increases with each dose. For example, in some embodiments, the methods disclosed herein result in an increase in efficiency of gene editing by conventional nucleic acid editing systems when used in conjunction with the modified nucleic acid editing systems disclosed herein. For example the percentage efficiency of gene editing increases by about 1%, about 2%, about 3%, about 4%, about 5%, about 6%, about 7%, about 8%, about 9%, about 10%, about 11%, about 12%, about 13%, about 14%, about 15%, about 16%, about 17%, about 18%, about 19%, about 20%, about 21%, about 22%, about 23%, about 24%, about 25%, about 26%, about 27%, about 28%, about 29%, about 30%, about 31%, about 32%, about 33%, about 34%, about 35%, about 36%, about 37%, about 38%, about 39%, about 40%, about 41%, about 42%, about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%,

about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, or more.

[0099] In another embodiment at least one of the guide RNA, modified nucleic acid editing system, and donor DNA are administered to a tissue or cell at the same time, such as on the same delivery vehicle, and one or more component (i.e., the modified nucleic acid editing system, and guide RNA) is under the control of an inducible promoter. As an example, in one embodiment the inducible promoter may be for example a small molecule-induced promoter such as tetracycline-inducible promoter.

[0100] The delivery vehicles (whether viral vector or non-viral vector or RNA conjugate material) may be administered by any method known in the art, including injection, optionally by direct injection to target tissues or cells. Nucleic acid modification can be monitored over time by, for example, periodic biopsy with PCR amplification and/or sequencing of the target region from genomic DNA, or by RT-PCR and/or sequencing of the expressed transcripts. Alternatively, nucleic acid modification can be monitored by detection of a reporter gene or reporter sequence. Alternatively, nucleic acid modification can be monitored by expression or activity of a modified gene product or a therapeutic effect in the cell or tissue or in a subject.

[0101] In some embodiments, the cell or tissue is in a subject. For example the subject may be a human, in particular a human in need of therapeutic or prophylactic intervention. Alternatively, the subject is an animal, including livestock, poultry, domesticated animal, or laboratory animal. In various embodiments, the subject is a mammal, such as a human, horse, cow, dog, cat, rodent, or pig.

[0102] In some embodiments, the methods provided herein include obtaining a cell or population of cells from a subject and modifying a target polynucleotide in the cell or cells *ex vivo*, using the delivery systems, compositions, methods, and/or kits disclosed herein. In further embodiments, the *ex vivo* modified cell or cells may be re-introduced into the subject following *ex vivo* modification. Thus, the present disclosure provides methods for treating a disease or disorder in a subject, comprising obtaining one or more cells from the subject, modifying one or more target nucleotide sequences in the cell *ex vivo* using both conventional and the modified nucleic acid editing systems described herein and re-introducing of the cell with the modified target nucleotide sequence back into the subject having the disease or disorder. In some embodiments, cells in which nucleotide sequence

modification has occurred are expanded *in vitro* prior to reintroduction into the subject having the disease or disorder.

[0103] In other embodiments, at least one of the modified nucleic acid editing system, guide RNA and donor DNA are administered to a cell *in vitro*.

[0104] In some embodiments, at least one component (e.g., the guide RNA, donor, modified nucleic acid editing system or nucleic acid vector) accumulates in a cell or tissue which may be, for example, liver, heart, lung (including airway epithelial cells), skeletal muscle, CNS (e.g., nerve cells), endothelial cells, blood cells, bone marrow cells, blood cell precursor cells, stem cells, fat cells, or immune cells. Tissue targeting or distribution can be controlled by selection and design a viral delivery vehicle, or in some embodiments is achieved by selection and design of lipid or polymeric delivery vehicles.

[0105] In some embodiments, the percentage efficiency of target sequence modification (editing) using the methods disclosed herein is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% at least 95%, or at least 99%.

[0106] In some embodiments, the efficiency of target sequence modification (editing) using the methods and compositions disclosed herein provides a 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, a 7-fold increase or more compared to methods of making the same modification (edit) without tethering the donor DNA.

[0107] In some embodiments, the efficiency of target sequence modification is less than 100%, or wherein an effect on fewer than 100% of the cells has a therapeutic effect. For example, a therapeutic effect may be achieved when the efficiency of nucleic acid modification of about 0.01% to about 100%, about 0.01 % to about 50%, about 0.05% to about 40%, about 0.1% to about 30%, about 0.5% to about 25%, about 1 % to about 20%, about 1 % to about 15%, about 1% to about 10%, or about 1% to about 5%. Thus, even if the efficiency of nucleotide sequence modification is relatively low (e.g., less than 50%, or less than 40%, or less than 30%, or less than 20%, or less than 10%, or less than 5%, or less than 1%, or less than 0.5%, or less than 0.1%), modest expression of the introduced or corrected or modified gene product may result in a therapeutic effect in the disease or disorder.

[0108] In some embodiments, the delivery systems and compositions disclosed herein are formulated such that the ratio of the components is optimized for consistent delivery to the target sequence. In one embodiment, the ratio of the guide RNA and modified nucleic acid editing system is optimized for consistent delivery to the target sequence. In another embodiment, the ratio of the donor DNA to the guide RNA and/or to the modified nucleic

acid editing system is optimized for consistent delivery to the target sequence. For example, in one embodiment, the ratio of modified Cas9:guideRNA:donor is from about 1:1:1 to about 1:1:100. In a further embodiment, the ratio is from about 1:1:2 to about 1:1:90, from about 1:1:5 to about 1:1:75, or from about 1:1:10 to about 1:1:50. In other embodiments, the ratio is about 1:1:1 or below, such as from about 1:1:0.01 to about 1:1:1, from about 1:1:0.02 to about 1:1:0.75, or about 1:1:0.05 to about 1:1:0.5, or about 1:1:0.1 to about 1:1:0.5. In other embodiments, wherein the composition does not include a guide RNA, the ratio of modified nucleic acid editing system:donor DNA is from about 1:100 to about 100:1, or about 1:50 to about 50:1, or about 1:25 to about 25:1, or about 1:10 to about 10:1, or about 1:5 to about 5:1, or about 1:2 to about 2:1, or about 1:1.

Kits

[0109] In one aspect, there is provided kits containing any one or more of the components disclosed in the above methods, compositions, and delivery systems. Kit components may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. In some embodiments, the kits disclosed herein comprise one or more reagents for use in the embodiments disclosed herein. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular method, or in a form that requires addition of one or more other components before use (e.g. in concentrate or lyophilized form). Suitable buffers include, but are not limited to, phosphate buffered saline, sodium carbonate buffer, sodium bicarbonate buffer, borate buffer, Tris buffer, MOPS buffer, HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10.

[0110] For example, a kit may comprise: a donor DNA and a modified nucleic acid editing system. The kit may further comprise a guide RNA. The kit may provide an expression system providing for expression of either or both of the modified nucleic acid editing system and guide RNA in a target cell. The kit may provide one or more doses of an RNA delivery system, each dose providing for expression of the modified nucleic acid editing system in the target cell or tissue.

[0111] The kit may be custom made for use with user defined target sequences.

[0112] It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

EXAMPLES

Example 1: Plasmid for tethering Crispr/Cas9 complex to a genomic target DNA sequence.

[0113] Figure 2 is a map of a plasmid vector containing an active Cas 9 gene sequence and its guide RNA. The coding sequence of Cas9 is contiguous with a lac repressor DNA binding domain. This fusion is operably linked to a CMV promoter. When expressed lac repressor DNA binding domain binds the lac operator sequence in the plasmid backbone sequence.

[0114] Donor DNA complementary to a genomic target DNA sequence is also cloned into the vector and provides a template for homologous recombination between the Crispr/Cas9 generated double-strand break in the target DNA sequence. In one embodiment the donor DNA sequence is modified to prevent binding of the Crispr/Cas9 nuclease to plasmid sequence and contains selectable markers to aid in identification of recombinant cell lines. It is contemplated that the donor DNA sequence may contain mutant DNA sequences to change the function of the target chromosomal gene or it may contain a 'wild type' sequence to correct a mutant target DNA sequence.

Example 2: Plasmids for a binary tethering system.

[0115] Figures 3 and 4 are maps of plasmids used in a binary system where the Crispr/Cas9 gene sequence containing the lac repressor fusion and guide RNA are present on a tether expression vector (Figure 3) and the donor DNA sequence is present on a second tethered gene targeting vector (Figure 4). Co-transfection of both vectors is necessary for expression of the target-specific Crispr/Cas9/lac repressor nuclease binds the lac operator sequence on the tethered targeting plasmid (Figure 4) thus localizing the gene targeting plasmid to the target DNA.

[0116] The sequences of complementary oligonucleotides used to clone the lactose operator sequence or tetracycline resistance operator sequence into HindIII/Sall restriction endonuclease digested pUC19. The correct orientation allows subsequent cloning of LacO and TetO duplexes into pUC19 to generate plasmids with one or more sequential operators that can be used to clone donor DNA molecules for gene editing by homology directed recombination in mammalian cells.

Table 1. LacO and TetO Oligonucleotides.

<u>Name</u>	<u>Sequence</u>	<u>SEQ ID NO</u>
tetO w	TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG	59

tetO c	TCGACTTTCACCTTTTCTCTATCACTGATAGGGAGTGGTAAAC	60
lacO w	TCGAGTTTAGTGGAATTGTGAGCGGATAACAATTTCACTGAAAG	61
lacO c	TCGACTTTCAGTGAAATTGTTATCCGCTCACAATTCCACTAAAC	62

Example 3: plasmids for an *in vitro* tethering system.

[0117] Figures 5 and 6 are plasmid maps of vectors for an *in vitro* tethering system whereby purified Caspr/Cas9 is used to bind a gene targeting vector prior to transfection into cells. The His-tagged Caspr/Cas9-lac repressor fusion is expressed from the plasmid shown in Figure 5, the fusion is then purified by immobilized metal affinity chromatography. Figure 5 indicates that the Lac repressor protein operator-binding domain is fused to the c-terminus of the CAS9 9. However, as set out in SEQ ID NOS: 8 and 9 a preferable arrangement is that the lac repressor is fused to the N –terminus of Cas9 preferably using a linker such as the XTEN linker. In any case the purified fusion protein is the mixed with a gene targeting vector (i.e. Figure 4). The complex of purified fusion protein and gene targeting vector is transfected into cells with the guide RNA expression plasmid (Figure 6).

Example 4. Gene Editing of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene by GeneTethered Donor DNA Fragments.

[0118] Cystic fibrosis is one of the most prevalent genetic diseases found in the Caucasian population with as many as 1/27 individuals carrying a mutation in the CFTR gene. The most common mutation found in CF patients is the delF508, or F508del, mutation that is the result of an in frame 3 base pair deletion in exon 11 resulting in a loss of phenylalanine at residue 508 in the protein. Several immortalized human cell lines have been generated from CF patients, including the CFBE41o- cell line homozygous for the delF508 mutation. The CFTR gene sequence in HEK293 cells, however, is a normal, or wild type, CFTR gene sequence.

[0119] In this example, the px458 Cas9 vector was modified by engineering a lactose repressor protein, or tetracycline repressor protein, fused in frame with the Cas9 protein sequence to create a “GeneTether”. This GeneTether modification enables binding of donor DNA molecules used for gene editing to the functional Cas9 protein and localizes the donor DNA molecule at the Cas9-generated double strand break to enhance homology directed repair, reduce on-target mutations, and reduce induction of the P53 DNA repair system. These modified Cas9 proteins were transfected into cells with donor DNA molecules containing lactose operator or tetracycline operator DNA sequences to measure the gene editing efficiency introducing the delF508 mutation into the normal HEK293 CFTR gene

Construction of Lac and Tet Tether modified Cas9.

[0120] The lactose and tetracycline repressor proteins bind well-defined operator sequences with high specificity and affinity. GeneTethers are lactose or tetracycline repressor proteins fused with Cas9 (and other Cas proteins), TALENS, or ZNF with that will bind to DNA molecules containing the respective operator sequences. Binding of GeneTether-modified Cas proteins, TALENS, or ZNF to their genomic target will thereby physically localize any DNA molecules, bound to the repressor protein fusion, to the same genomic site (Figure 1). Localization of DNA molecules, homologous to the genomic DNA target sequences, enhance the efficiency of gene editing by the homologous directed recombination DNA repair pathways and minimize DNA mutations.

Construction of Lactose Repressor-modified Cas9

[0121] The lactose repressor gene was amplified by polymerase chain reaction from *Escherichia coli* DH5 α using lacI Primer 2f/Primer 2r (Table 2) and the Q5 high fidelity thermostable polymerase (New England Biolabs) according to vendor instructions. The 1083 base pair product was gel-purified (Monarch DNA Gel Extraction Kit, New England Biolabs) and used for further modifications. Two BbsI restriction endonuclease sites in the Lactose repressor gene sequence were inactivated using polymerase chain reaction amplification with the mutagenic primers lacIBbs2f, lacIBbs2r, lacIBbs3f, and lacIBbs3r to generate A to G transitions in codons 164 and 277, retaining glutamic acid codons (Figure LacIBbsAssembled). Three gel-purified, Q5 high fidelity polymerase, polymerase chain reaction products using primer pairs lacI primer 1f/lacI Bbs Primer 2r, lacI Bbs Primer 2f/lacI Bbs Primer 3r, and lacI Bbs Primer 3f/lacI Primer 1r (Table 2) were used to reconstruct the full length 1083 base pair BbsI-inactivated Lactose repressor protein gene (NEBuilder HiFi Assembly, New England Biolabs).

Table 2. PCR Primer DNA Sequences.

Primer	Sequence	SEQ ID NO
CF1Bf	CCTTCTCTGTGAACCTCTATCA	10
CF1f	GCAGAGTACCTGAAACAGGA	11
CF5r	CATTACAGTAGCTTACCCA	12
CF7Cr	ATAGGAAACACCAAAGATGA	13
CF8Cr	ATAGGAAACACCAATGATAT	14
CF8Crfull	ATAGGAAACACCAATGATATTTTCTTTAATGGTGCCAGGC	15
CF9Cf	GAAAATATCATTGGTGTTCCTATGATGAATATAGATACAG	16

CF96250f	TGAGTTAGATGTTTGACGC	17
CF96310f	GCTGTGCATTTTCCTCTGGGTAATACTTTAG	18
CF98236f	GTCTCTATTACTTAATCTGTACCT	19
CF98328f	CTGTGAAGATTAAATAAATTAATATAGTTAAAGCAC	20
CF99287r	ATGCTCATTCCATTAGGCTATAGTATTA	21
CF99310r	CTAATTCTCTGCTGGCAGATCAATGC	22
CF101310r	CAAGACGTTGTGTTAGGTACATTACATGTACATC	23
lacI Bbs Primer 2f	CTCCCATGAGGACGGTACGCGACTGGGC	24
lacI Bbs Primer 2r	GCCCAGTCGCGTACCGTCCTCATGGGAG	25
lacI Bbs Primer 3f	GTGGGATACGACGATACCGAGGACAGCTCATGTTATATC	26
lacI Bbs Primer 3r	GATATAACATGAGCTGTCCTCGGTATCGTCGTATCCCAC	27
lacI Primer 1f	ATGAAACCAGTAACGTTATACGATGTCGC	28
lacI Primer 1r	TCACTGCCCGCTTTCCAG	29
lacI Primer 2f	GGTATCCACGGAGTCCCAGCAGCCATGAAACCAGTAACG TTAT	30
Primer1fAg e	CTGGAGCACCTGCCTGAAATCAC	31
Primer1fXb a	CGCGTGCGCCAATTCTGCAGACAAATG	32
Primer1r	TGCTGGGACTCCGTGGATACCGACCTTCCGCTTC	33
Primer2f	GGTATCCACGGAGTCCCAGCAGCCGTGAAACCAGTAACG TTAT	34
Primer2r	GGCGGACTCTGAGGTCCCGGGAGTCTCGCTGCCGCTCT GCCCGCTTTCCAG	35
Primer3f	CGGGACCTCAGAGTCCGCCACACCCGAAAGTGACAAGAA GTACAGCATC	36
Primer3r	CGTCCACCTTGGCCATCTCGTTGCTG	37
lacOsymCF	GTTCCGAATATAAATTGTGAGCGCTCACAATTAAGCTTGC	38

1f	AGAGTACCTGAAACAGGA	
lacOsymCF 5r	GTTTCGGAATATAAATTGTGAGCGCTCACAATTAAGCTTCA TTCACAGTAGCTTACCCA	39
lacOsymCF 5kf	GTTTCGGAATATAAATTGTGAGCGCTCACAATTAAGCTTGC TGTGCATTTTCCTCTGGGT	40
lacOsymCF 5kr	GTTTCGGAATATAAATTGTGAGCGCTCACAATTAAGCTTCA AGACGTTGTGTAGGTACATTACATGTAC	41
pUC19poly CF5kr	GAATTCGAGCTCGGTACCCGGGGATCCCAAGACGTTGTG TTAGGTACATTACATGTAC	42
pUC19poly CF5r	GAATTCGAGCTCGGTACCCGGGGATCCCATTACAGTAG CTTACCCA	43
tetoCF5kf	CACTCCCTATCAGTGATAGAGAAAAGAAAGCTGTGCATTT TCCTCTGGGT	44
tetoCF5kr	CACTCCCTATCAGTGATAGAGAAACAAGACGTTGTGTTAG GTACATTACATGTACATC	45
tetoCF1f	CACTCCCTATCAGTGATAGAGAAAAGGCAGAGTACCTGAA ACAGGA	46
tetoCF5r	CACTCCCTATCAGTGATAGAGAAAAGCATTACAGTAGCT TACCCA	47
10635f	CGGAGCCTATGGAAAAACGCCAGC	48

[0122] Construction of the lactose repressor-Cas9 fusion was performed as outlined in Figure 8. Three overlapping DNA fragments amplified by polymerase chain reaction with the Q5 high fidelity polymerase were cloned into AgeI/BglII digested pSpCas9 BB-2A-GFP(px458) (Genescript, Inc) to generate the GeneTether lactose repressor-Cas9 fusion plasmid pGT1 (Figure 8D, SEQ ID NO 63) using the NEBuilder system (New England Biolabs). The three, polymerase chain reaction products used for this assembly were generated using the polymerase chain reaction primers to the px458 vector Primer1fAge/Primer1r, and Primer3f/Primer3r, and the BbsI-inactivated Lactose repressor gene was amplified using Primer2f/Primer2r (Table 2). The resulting plasmid, pGT1, was screened for by colony polymerase chain reaction and the Lactose repressor-Cas9 fusion gene sequence was confirmed by DNA sequencing (Quintara Biosciences), restriction endonuclease mapping, and diagnostic polymerase chain reaction amplification.

Construction of Tetracycline Repressor-modified Cas9.

[0123] The gene encoding the class B TN10 tetracycline repressor, with human codon preference, was ordered as a synthetic DNA molecule (Genescript, Inc) cloned into the

pUC57 vector (Figure 9 and SEQ ID NO: 65). Construction of the tetracycline-repressor-Cas9 fusion was performed similar to the pGT1 lactose repressor-Cas9 fusion (Figure 10). The three, polymerase chain reaction products used for this assembly were generated using the polymerase chain reaction primers to the px458 vector Primer1fAge/Primer1r, and Primer3f/Primer3r, and the Tetracycline repressor gene was amplified using the primers NLS Linker primer1f/ tet linker Cas9 Primer 1r. The resulting plasmid, pGT9 (SEQ ID NO: 63), was screened for by colony polymerase chain reaction and the Tetracycline repressor-Cas9 fusion gene sequence was confirmed by DNA sequencing (Quintara Biosciences), restriction endonuclease mapping, and diagnostic polymerase chain reaction amplification. [0124] Guide RNA Cloning Guide RNA sequences were designed to bind the exon 11 gene sequence of the wild type CFTR gene and the CFTR del-F508 mutant gene sequence (Table 3). The guide DNA oligonucleotides were annealed to create duplex molecules with Bbs1 compatible overhang and were subsequently ligated into gel-purified BbsI restriction endonuclease digested px458, pGT1, and pGT9.

Table 3. Guide Sequences. Shown are the DNA oligonucleotides for cloning into the BbsI sites in pGT1, pGT9, and pX458 Cas9 vectors

Name	Sequence	SEQ ID NO
CFWT 130117 fw	CACCGATTAAAGAAAATATCATCTT	49
CFWT 130117 fc	AAACAAGATGATATTTTCTTTAATC	50
CFWT 130121 rw	CACCGAAAGATGATATTTTCTTTAA	51
CFWT 130121 rc	AAACTTAAAGAAAATATCATCTTTC	52
delF 130127 fw	CACCGACCATTAAGAAAATATCAT	53
delF 130127 fc	AAACATGATATTTTCTTTAATGGTC	54
delF 130138 rw	CACCGACCAATGATATTTTCTTTAA	55
delF 130138 rc	AAACTTAAAGAAAATATCATTGGTC	56

[0125] aCFTR Donor DNA Generation 5 kilobasepair CFTR DNA fragments approximately centered on the CFTR gene exon 11 were generated by polymerase chain reaction using the Q5 high fidelity DNA polymerase and primers CF96310f and CF101310r. Donor DNA fragments containing the CFTR delF508 mutation were generated using genomic DNA from the CFBE41o- cell line that is homozygous for the delF508 mutation. Donor DNA fragments containing the wild type (nonmutant) CFTR gene sequence were generated using genomic DNA from the HEK293 cell line.

[0126] Donor DNA fragments used for gene editing were modified at the 5' or 3' end of the fragment to contain either the lactose operator sequence (AATTGTGAGCGCTCACAATT, SEQ ID NO: 57) or tetracycline operator sequence (CACTCCCTATCAGTGATAGAGAAA, SEQ ID NO: 58). To generate 500 base pair donor DNA molecules, polymerase chain reaction amplification using the Q5 high fidelity thermostable DNA polymerase with the primer pairs lacOsymCF1f/CF5r to add the lactose operator sequence to the 5' end of the donor DNA or primer pairs Cf1f/lacOsymCF5r to add the lactose operator to the 3' end of the donor DNA fragment (Figure Donor DNA Molecules). Donor DNA molecules with the tetracycline operator sequence were generated using the primer pairs tetoCF1f/CF5r and CF1f/ tetoCF5r. Donor DNA fragments were gel-purified prior to use in cell transfections for gene editing.

Cell Transfection with px458, pGT1, pGT9 Cas9 vectors and 500 base pair donor DNA fragments

[0127] HEK293 cells were transfected in 12-well plates using Lipofectamine 3000 (Invitrogen, Inc) with 500 ng of plasmid DNA and 500 ng of gel purified donor DNA fragments. The px458, pGT1, and pGT9 plasmids contain a green fluorescent protein reporter gene that is expressed in transfected cells and allows monitoring of transfection efficiencies. Green fluorescent protein expression was visualized by fluorescence microscopy at approximately 48 hours post transfection. At 48-72 hours post transfection, 12-well plates of transfected cells were washed once with phosphate buffered saline and stored at -80°C until DNA was harvested for analysis of gene editing efficiencies.

Analysis for the delF508 mutation by Allele-Specific Polymerase Chain Reaction Analysis

[0128] The presence of the delF508 mutation in a population of CFTR wild type cells can be detected using polymerase chain reaction with Taq polymerase and the primer pair CF1Bf/CF8Cr. Since the CF1Bf primer is located 5' or outside of the donor DNA fragment, the 388 basepair product is generated from an "inside-out" approach and is specifically diagnostic for gene edited events and will not amplify randomly integrated or unintegrated donor DNA molecules.

[0129] Genomic DNA was prepared directly from each well of a 12-well plate (Genejet Genomic Purification Kit, ThermoFisher, Inc) and DNA concentration determined by ultraviolet spectroscopy. Allele specific polymerase chain reactions were performed on 100ng of genomic DNA and 500μM primer, 1.5 mM MgCl₂, HotStart Taq polymerase (New England Biolabs). The thermocycler program for semiquantitative amplification of the delF508 mutant DNA sequence was 95 °C, 2 minutes; 35 cycles of 95 °C, 30 seconds, 50 °C, 30 seconds, 72 °C, 1 minute; followed by 72 °C, 8 minutes. The 388 base pair

polymerase chain reaction product was visualized with ultraviolet light on a 1.5% agarose gel stained with Gelred (Biotium, Inc). For semi-quantification of the PCR product, standard genomic DNA samples with varying ratios of wild type and delF508 mutant DNA were amplified in parallel to experimental DNA samples.

Cell Transfection

[0130] Two different tethers (lactose-repressor-Cas9 fusion or tetracycline repressor-Cas9 fusion) were used in combination with donor DNA molecules containing a 5' lactose or tetracycline operator sequence or a 3' lactose or tetracycline operator sequence (Figures 11 and 12) and containing the delF508 deletion sequence. Homology directed repair by homologous recombination between the genomic CFTR target Cas9-induced double strand break and donor fragment would catalyze transfer of the delF508 DNA sequence to the genomic target (Figure 12).

[0131] Cas9 guides complementary to the nonmutated wild type CFTR exon 11 DNA sequence, but not to the delF508 DNA sequence were selected to allow recognition of the genomic DNA target and prevent the Cas9 enzyme from recognizing, and cleaving, the delF508 donor DNA fragment.

[0132] Approximately 5×10^5 HEK293 cells per well were seeded, 24 hours before transfection, into Corning 12-well tissue culture plates such that the wells were 50-80% confluent at the time of transfection. The media on the cells was changed prior to transfection.

[0133] For transfections, 500 ng of px458, pGT1 or pGT9 was mixed with 500 ng of donor DNA fragment in 50 μ L of DMEM medium. 2 μ L of the P3000 reagent was then added to the DNA/DMEM solution, mixed well, and allowed to incubate at room temperature for 5 minutes. A solution of Lipofectamine 3000 was made by adding 1.5 μ L of undiluted Lipofectamine 3000 to 50 μ L of DMEM and equal volumes of DNA/DMEM solutions and Lipofectamine/DMEM solutions were mixed and incubated at room temperature for 12-15 minutes. 100 μ L of the DNA/P3000/Lipofectamine solution was then added per well of Corning 12-well plates. Two days after transfection the cells are examined with fluorescent microscopy to assess the extent of successful transfection.

GeneTethers Increase Gene Editing Efficiency

[0134] Several variables for the effect of GeneTether modified Cas9 vectors on gene editing efficiency were tested: a lactose repressor-Cas9 fusion protein, tetracycline repressor-Cas9 fusion protein, Cas9 guides complementary to the top or bottom strand of genomic target, and 500 bp donor DNA molecules containing the delF508 deletion with the lactose operator sequence at the 5' or 3' end of the donor DNA fragment.

[0135] The combination of pGT1 vector (lactose repressor-Cas9 fusion) with the 130117 guide (genomic target forward strand) and lactose operator on the 3' end of donor DNA (Figure 13, lane 7) demonstrates higher gene editing efficiency as compared to the px458 vector with unmodified Cas9 and same donor DNA fragment (compare 388 basepair products lane 3 and lane 7, Figure 13. Measurement of band intensities using ImageJ software (<https://imagej.nih.gov>) indicates that the tethered donor DNA complex has an approximate 7-fold higher gene editing efficiency compared to the px458 unmodified Cas9. Indeed, comparison of lane 7 band intensity to the reconstructed mixture of WT:delF508 DNA (lanes 14-19) suggests the delF508 mutation is present in approximately 10% of the genomic DNA. Since transfection efficiencies for these experiments were 10%, or less, up to 100% of cells transfected with the combination of pGT1/130117guide/3' lactose operator sequence underwent successful gene editing.

[0136] Other combinations of pGT1/guide/donor DNA performed equal to or, slightly better than, unmodified Cas9 (Figure 13; lane 9 vs lane 5) for gene editing, demonstrating that the Cas9 protein activity was not affected by the lactose repressor fusion. The placement of the lactose operator sequence may slightly favor the 3' placement (Figure 13; lanes 7 and 9 vs lanes 3 and 5). Donor DNA transfected with px458 not containing guide sequence showed low levels of gene editing (Figure 13; lanes 10 and 11). Some faint 388 bp product is evident in lanes 10, 12, and 13 that is likely due to artifactual low-level mispriming of PCR primers used for the allele-specific PCR.

[0137] Gene editing with the pGT9 tetracycline repressor-Cas9 fusion vector, guide sequences, and donor fragments did not yield appreciably better gene editing frequencies than the control px458 vector (Figure 14, lanes 2-5 vs lanes 6-9). The placement of the tetracycline at the 3' end of the donor DNA fragment appears to result in slightly better gene editing efficiencies (Figure 14 lanes 2, 4, 6, 8 vs lanes 3, 5, 7) similar to placement of the lactose operator sequence.

[0138] These results demonstrate that the GeneTether lactose repressor-Cas9 fusion protein encoded by pGT1 can significantly increase gene editing efficiency as compared to the unmodified Cas9 found in the px458 vector. Indeed, it is possible that the combination of pGT1/130117 guide/delF508 lacO 500 successfully caused gene editing of WT to delF508 in almost 100% of transfected cells.