(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 October 2013 (31.10.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/163628 A2

(51) International Patent Classification:
A61K 48/00 (2006.01)

(21) International Application Number:
PCT/US2013/038536

(22) International Filing Date:
26 April 2013 (26.04.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/639,227 27 April 2012 (27.04.2012) US

(71) Applicant: **DUKE UNIVERSITY** [US/US]; 2812 Erwin Road, Suite 306, Durham, NC 27705 (US).

(72) Inventors: **GERSBACH, Charles**; 3220 Wilderness Road, Durham, NC 27712 (US). **OUSTEROUT, David**; 1725 T.W. Alexander Drive, Unite 105, Durham, NC 27703 (US).

(74) Agent: **YEH, Sansun**; Michael Best & Friedrich LLP, 100 East Wisconsin Avenue, Suite 3300, Milwaukee, WI 53202-4108 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

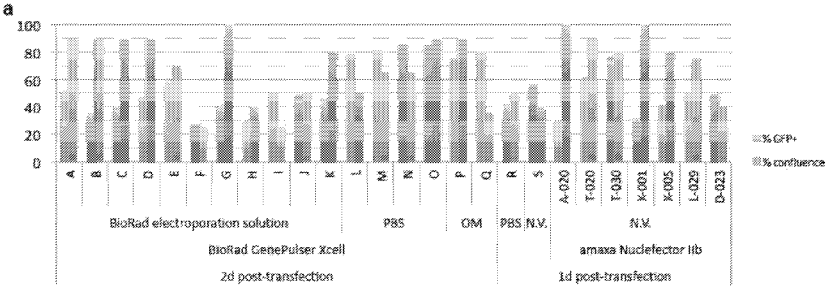
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: GENETIC CORRECTION OF MUTATED GENES

(57) Abstract: Disclosed herein are transcription activator-like effector nuclease (TALEN)-related compositions and methods of using said TALENs for correcting mutant genes.

a



b

Program	Conditions
A	1000uF, 120V, 1M cells/200uL, 10ug GFP + 10ug empty vector
B	500uF, 120V, 1M cells/200uL, 10ug GFP + 10ug empty vector
C	300uF, 120V, 1M cells/200uL, 10ug GFP + 10ug empty vector
D	500uF, 100V, 1M cells/200uL, 10ug GFP + 10ug empty vector
E	500uF, 150V, 1M cells/200uL, 10ug GFP + 10ug empty vector
F	500uF, 200V, 1M cells/200uL, 10ug GFP + 10ug empty vector
G	1000uF, 100V, 1M cells/200uL, 10ug GFP + 10ug empty vector
H	1000uF, 150V, 1M cells/200uL, 10ug GFP + 10ug empty vector
I	1000uF, 100V, 0.2M cells/200uL, 10ug GFP + 10ug empty vector
J	1000uF, 100V, 0.4M cells/200uL, 10ug GFP + 10ug empty vector
K	1000uF, 100V, 1M cells/200uL, 10ug GFP + 10ug empty vector
L	960uF, 190V, 0.5M cells/150uL, 5ug GFP
M	950uF, 190V, 1M cells/200uL, 10ug GFP
N	950uF, 160V, 0.5M cells/150uL, 5ug GFP
O	950uF, 160V, 1M cells/200uL, 10ug GFP
P	950uF, 220V, 2M cells/300uL, 10ug GFP
Q	950uF, 100V, 1M cells/200uL, 10ug GFP
R	950uF, 190V, 1M cells/200uL, 2ug pmaxGFP
S	950uF, 190V, 1M cells/100uL, 2ug pmaxGFP

Figure 1

WO 2013/163628 A2

GENETIC CORRECTION OF MUTATED GENES

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to U.S. Provisional Application No. 61/639,227, filed April 27, 2012, which is incorporated herein by reference in its entirety.

STATEMENT OF GOVERNMENT INTEREST

[0002] This invention was made with government support under federal grant number DP2-OD008586 awarded by NIH. The U.S. Government has certain rights to this invention.

SEQUENCE LISTING

[0003] The sequence listing is filed with the application in electronic format only and is incorporated by reference herein. The sequence listing text file "WO00_ASFILED_SequenceListing-Text" was created on April 26, 2013 and is 141,796 bytes in size.

TECHNICAL FIELD

[0004] The present disclosure relates to the field of genome engineering and genomic alteration of the mutated gene using transcription activator-like effector nucleases (TALENs).

BACKGROUND

[0005] Hereditary genetic diseases have devastating effects on children in the United States. These diseases currently have no cure and can only be managed by attempts to alleviate the symptoms. For decades, the field of gene therapy has promised a cure to these diseases, such as Duchenne muscular dystrophy (DMD), by introducing new genetic material into patient's cells. In contrast to gene addition, genome editing with engineered site-specific endonucleases selectively replace or correct disrupted genes. Technical hurdles regarding the safe and efficient delivery of therapeutic genes to cells and patients have limited these approaches. Scientists have

only been able to add new genetic material to cells without any control over where it is inserted into the genome. This strategy has led to a myriad of unforeseen negative consequences that can all be attributed to the inability to correct the existing mutated gene sequences. Current experimental gene therapy strategies for genetic diseases, such as DMD, use repeated administration of transient gene delivery vehicles or rely on permanent integration of foreign genetic material into the genomic DNA. Both of these methods have serious safety concerns. Furthermore, these strategies have been limited by an inability to deliver the large and complex gene sequences.

SUMMARY

[0006] The present disclosure is directed to a transcription activator-like effector nuclease (TALEN) protein that may bind to a dystrophin gene. The dystrophin gene may comprise a premature stop codon and the TALEN protein may bind upstream or downstream from the premature stop codon. The TALEN protein may bind to a nucleotide sequence comprising one of SEQ ID NOs: 16-46, 50, 52, 58, and 59, and the complement thereof. The TALEN protein may comprise a nuclease. The nuclease may comprise FokI. The TALEN protein may comprise 15-19 RVD modules. The TALEN protein may comprise an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, and SEQ ID NO: 15. The TALEN protein may bind to a region in exon 51 of the dystrophin gene. The TALEN protein may bind to a nucleotide sequence comprising one of SEQ ID NOs: 16-24, 26-41, 46, 50, 52, 58, and 59, and the complement thereof. The TALEN protein may comprise a nuclease. The nuclease may comprise FokI. The TALEN protein may comprise 15-19 RVD modules. The TALEN protein may comprise an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13. The TALEN protein may bind to a nucleotide sequence comprising SEQ ID NO: 19, or the complement thereof. The TALEN protein may comprise an amino acid sequence selected from the group consisting of SEQ ID NO: 2 and 4. The TALEN protein may bind to a region in the 5' UTR of the dystrophin gene. The TALEN protein may bind to a nucleotide sequence

comprising one of SEQ ID NOs: 25 and 42-45 and the complement thereof. The TALEN protein may comprise a nuclease. The nuclease may comprise FokI. The TALEN protein may comprise 15-19 RVD modules. The TALEN protein may comprise an amino acid sequence selected from the group consisting of SEQ ID NO: 14 and 15. The dystrophin gene may be a human dystrophin gene.

[0007] The present disclosure is directed to an isolated polynucleotide comprising a nucleotide sequence encoding said TALEN protein and a vector comprising said isolated polynucleotide. The present disclosure is directed to a cell comprising said isolated polynucleotide or said vector.

[0008] The present disclosure is directed to a composition comprising two or more of said TALENs, wherein a first TALEN binds to a first binding region and a second TALEN binds to a second binding region, wherein the first binding region and second binding region are located within a target region and the first binding region and second binding region are not the same. The first binding region and the second binding region may be separated by at least one nucleotide. The first binding region and the second binding region may be separated by 5-25 base pairs.

[0009] The present disclosure is directed to a method of treating a subject in need thereof having a mutant dystrophin gene, the method comprising administering to the subject said TALEN protein, said isolated polynucleotide, said vector, said cell or said composition. The subject may be suffering from Duchenne muscular dystrophy.

[0010] The present disclosure is directed to a method of correcting a mutant dystrophin gene in a cell, the method comprising administering to a cell containing a mutant dystrophin gene said TALEN protein, said isolated polynucleotide, said vector, or said composition. The correction of the mutant dystrophin gene may comprise homology-directed repair. The method may further comprise administering to the cell a donor DNA. The mutant dystrophin gene may comprise a frameshift mutation which causes a premature stop codon and a truncated gene product. The correction of the mutant dystrophin gene may comprise nuclease mediated non-homologous end joining.

[0011] The present disclosure is directed to a method of correcting a mutant gene in a cell, the method comprising administering to a cell containing a mutant gene a first TALEN and a second TALEN, wherein the first TALEN binds to a first binding region and a second TALEN binds to

a second binding region, wherein the first binding region and second binding region are located within a target region and the first binding region and second binding region are not the same, wherein the correction of the mutant dystrophin gene may comprise nuclease mediated non-homologous end joining, and wherein the correction restores the mutant gene.

[0012] The present disclosure is directed to a method of correcting a mutant dystrophin gene in a cell, the method comprising administering to a cell containing a mutant dystrophin gene a first TALEN and a second TALEN, wherein the first TALEN binds to a first binding region and a second TALEN binds to a second binding region, wherein the first binding region and second binding region are located within a target region and the first binding region and second binding region are not the same. The mutant dystrophin gene may comprise a premature stop codon and the target region is upstream or downstream of the premature stop codon. The correction of the mutant dystrophin gene may comprise nuclease mediated non-homologous end joining. The first binding region and the second binding region may be separated by at least one nucleotide. The first binding region and the second binding region may be separated by 5-25 base pairs. The method may not involve co-administration of exogenous DNA. The correction of the mutant dystrophin gene may comprise homology-directed repair. The method may further comprise administering to the cell a donor DNA.

[0013] The present disclosure is directed to a kit comprising said TALEN protein, said isolated polynucleotide, said vector, said cell or said composition.

[0014] The present disclosure is directed to a kit for correcting a mutant dystrophin gene, the kit comprising said TALEN protein, said isolated polynucleotide, said vector, said cell or said composition.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] Figure 1 shows the optimization of electroporation conditions for myoblasts. (a) DMD myoblast cells (cell line 1) were electroporated using BioRad Gene Pulser Xcell or Amaxa Nucleofector IIb devices using the indicated programs. Several different buffers were tested, including BioRad electroporation solution, Sigma phosphate-buffered saline product #D8537 (PBS), Invitrogen OptiMEM I (OM), or Amaxa Nucleofector solution V (N.V.). Conditions using the GenePulser device used infinite resistance. For nucleofection, 1 million cells/100 μL nucleofection solution and 2 μg of GFP vector were used according to the manufacturer's

specifications. Electroporation using the GenePulser device with program O in PBS solution was selected as the optimal conditions for electroporating myoblasts. (b) Conditions used to optimize BioRad Gene Pulser Xcell electroporation.

[0016] Figure 2 shows the full amino acid sequences of engineered TALENs targeted to exon 51 of the dystrophin gene. Fok ELD-S/KKR-S refers to the *Fok*I nuclease domain that was fused to the TALE DNA-binding domains. The RVD sequences are underlined.

[0017] Figure 3 shows the design of TALENs targeting exon 51 of the human dystrophin gene. (a) The possible reading frames of human dystrophin exon 51 and corresponding amino acid sequences after genome editing. (b) Combinations of TALEN pairs were designed to target immediately upstream of either out-of-frame stop codon (underline) in exon 51 (bold) of the human dystrophin gene. (c) Representative genetic mutations that cause DMD and the resulting dystrophin protein are shown (deletion of exons 45-50 or 46-50). Gene editing from (a) restored the dystrophin reading frame in both types of mutations to create a functional dystrophin protein.

[0018] Figure 4 shows the validation and characterization of TALENs. (a) Each TALEN construct was transfected independently into HEK293T cells to confirm full-length expression. All TALENs were approximately 95-110kDa. (b) Combinations of TALENs were co-transfected into HEK293T cells to screen for highly active TALEN pairs. Gene modification frequency was monitored at day 3 and day 10 to assess stable gene modification. Arrows denote predicted cleavage band sizes indicative of NHEJ activity. (c) Summary of TALEN spacer lengths. (d) Measured gene modification rates detected by the Surveyor assay from day 3 data in (b). (e) Measured indel signal changes between day 3 and day 10 from the data in (b). (f) Cytotoxicity assay in HEK293T cells for all TALEN combinations. I-SceI is a non-toxic meganuclease and GZF3 is a zinc-finger nuclease known to be cytotoxic to human cells. "n.d." means not detected.

[0019] Figure 5 shows the optimization of cytotoxicity assay using Lipofectamine 2000 in 293T cells. Varying amounts of plasmid encoding the non-toxic endonuclease I-SceI and toxic zinc-finger nuclease GZF3 were transfected into 293T cells and assessed for relative survival rates post-transfection. Based on these data, 100 ng of nuclease expression plasmid was used for the cytotoxicity studies.

[0020] Figure 6 shows (A) Surveyor assay to detect exon 51 gene modification in clonal cell populations of TALEN-treated DMD myoblasts. Asterisks indicate modified cells. (B) Sequences of modified clonal populations. The size of the insertion/deletion is indicated on the

right. The two asterisks represent a 183 nucleotide insertion. Four of eleven modified clones (of 27 total assayed clones) showed indels that lead to reading frame correction

[0021] Figure 7 shows the genetic correction of aberrant dystrophin reading frames by TALEN-mediated genome editing. (a) Isogenic clones were derived from human skeletal myoblasts treated with ten micrograms of each plasmid encoding TN3/8 and screened using the Surveyor assay to detect mutant alleles in reference to the parent (untreated) genomic DNA. Arrows denote predicted cleavage band sizes indicative of NHEJ activity. (b) Sanger sequencing of the TALEN target site in exon 51 in mutant clones identified in (a). The asterisk represents a 183 nucleotide insertion. (c) DMD human myoblast cell line 1 was treated with ten micrograms of each plasmid encoding the TN3/8 TALEN pair and isogenic clones were subsequently derived. Sanger sequencing was used to identify clones with small insertion or deletion mutations at the exon 51 genomic locus characteristic of NHEJ. Clone 106 had a 5 bp deletion that may restore the reading frame (boxed). All other clones had deletions that were not predicted to result in corrective frameshift events. (d) Clonal cell populations with NHEJ events detected at exon 51 were cultured in differentiation conditions for 7 days and analyzed by western blot for dystrophin expression at the predicted molecular weight (412 kDa).

[0022] Figure 8 shows chromatograms of clones from Figure 7.

[0023] Figure 9 shows the efficient genetic modification and protein restoration in a bulk population of cells treated with TN3/8. (a,b) Dose-dependent response of NHEJ activity with increasing amounts of TALEN pair TN3/8 measured by the Surveyor assay after transfection of the indicated amount of each TALEN plasmid into two different DMD myoblast lines, each carrying a novel deletion of exons 48-50 (Δ48-50). Arrows denote predicted cleavage band sizes indicative of NHEJ activity. (c) DMD myoblast line 1 was treated with five micrograms of each TALEN plasmid and dystrophin expression was assessed after 7 days of differentiation by western blot using the NCL-Dys2 antibody. (d) DMD myoblast 2 was treated with the indicated amount of each TALEN plasmid and dystrophin expression was assessed after 7 days of differentiation by western blot using the MANDYS8 antibody. Protein from wild-type human myoblasts differentiated in parallel was diluted 1:100 and loaded as a positive control for full-length dystrophin expression (427 kDa) relative to the truncated Δ48-50 product (412 kDa).

[0024] Figure 10 shows the Dystrophin reading frame restoration in primary dermal fibroblasts. (a) Primary DMD fibroblasts carrying a deletion of exons 46-50 (Δ46-50) were

electroporated with increasing doses of the indicated amount of each TALEN plasmid and gene modification rates were quantified with the Surveyor assay. Arrows denote predicted cleavage band sizes indicative of NHEJ activity. (b) Analysis of myogenin and dystrophin expression (MANDYS8) in wild-type and DMD fibroblasts after treatment with TN3/8 and 15 days of forced MyoD expression. Protein from wild-type dermal fibroblasts was included as a positive control for full-length dystrophin expression (427 kDa) relative to the truncated Δ46-50 product (400 kDa). (c) Immunofluorescence staining to detect myosin heavy-chain (MHC) after MyoD expression by lentiviral gene transfer.

[0025] Figure 11 shows TALENs targeting the 5' UTR of the mouse and human dystrophin gene for integrating a gene, such as the dystrophin or miniaturized dystrophin cDNA, under the control of the dystrophin promoter. (A) The target site of TN45/50 in the human and mouse dystrophin 5' UTR. (B) The RVD components of TN45/50. (C) Combinations of TALENs were co-transfected into HEK293T cells to screen for highly active TALEN pairs. Gene modification frequency was monitored at day 3 and day 10 to assess stable gene modification. (D) 10 micrograms each plasmid encoding each TN45 and TN50 were electroporated into human skeletal myoblasts from a DMD patients. TALEN-mediated modification of the target locus was determined by the Surveyor assay three days after transfection.

[0026] Figure 12 shows the full amino acid sequences of engineered TALENs targeted to the 5'UTR of the dystrophin gene. Fok ELDS/KKRS refers to the *Fok*I nuclease domain that was fused to the TALE DNA-binding domains.

[0027] Figure 13 shows (a) the schematic detailing PCR-based detection of targeted integration events, (b) identification of targeted clones by PCR in which a GFP or minidystrophin gene was inserted into the TALEN target site, asterisks indicate positive targeted events (c) minidystrophin expression in clones positive for targeted integration, (d) GFP expression in differentiation conditions in selected clones from treated DMD patient lines.

DETAILED DESCRIPTION

[0028] As described herein, certain methods and engineered transcription activator-like effector nuclease (TALEN)-related compositions have been discovered to be useful for correcting or reducing the effects of mutations in genes involved in genetic diseases, for example, the dystrophin gene for phenotypic correction of DMD. The present disclosure is

directed to genome editing with TALENs with a repair template or donor DNA, which can replace the entire gene or the region containing the mutation and restore the expression of a full-functional or partially-functional protein. The present disclosure is also directed to genome editing with TALENs without a repair template, which can efficiently correct the reading frame and restore the expression of a functional protein involved in a genetic disease. The disclosed TALENs and methods may involve using nuclease-mediated non-homologous end joining (NHEJ)-based correction approaches, which enable efficient correction in proliferation-limited primary cell lines that may not be amenable to homologous recombination or selection-based gene correction. This strategy integrates the rapid and robust assembly of active TALENs with an efficient gene editing method for the treatment of genetic diseases caused by mutations in nonessential coding regions that cause frameshifts or premature stop codons, with no TALEN-mediated off-target changes to the protein-coding regions of the genome.

[0029] Section headings as used in this section and the entire disclosure herein are merely for organizational purposes and are not intended to be limiting.

1. Definitions

[0030] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[0031] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms "a," "and" and "the" include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments "comprising," "consisting of" and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[0032] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[0033] "Binding region" as used herein refers to the region within the target region that is recognized and bound by a TALEN. The TALE DNA-binding domain of the TALEN recognizes and binds to the binding region. For example, the binding region in the dystrophin gene may include a nucleotide sequence found within a target region of SEQ ID NO: 16-25, 44, or 45, or a complement thereof. The binding region may include a nucleotide sequence of SEQ ID NO: 26-43, or a complement thereof.

[0034] "Coding sequence" or "encoding nucleic acid" as used herein means the nucleic acids (RNA or DNA molecule) that comprise a nucleotide sequence which encodes a protein. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The coding sequence may be codon optimize.

[0035] "Complement" or "complementary" as used herein means a nucleic acid can mean Watson-Crick (*e.g.*, A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[0036] "Correcting", "genome editing" and "restoring" as used herein refers to changing a mutant gene that encodes a truncated protein or no protein at all, such that a full-length functional or partially full-length functional protein expression is obtained. Correcting or restoring a mutant gene may include replacing the region of the gene that has the mutation or replacing the entire mutant gene with a copy of the gene that does not have the mutation with a repair mechanism such as HDR. Correcting or restoring a mutant gene may also include repairing a frameshift mutation that causes a premature stop codon by generating a double stranded break in the gene that is then repaired using NHEJ. NHEJ may add or delete at least one base pair during repair which may restore the proper reading frame and eliminate the premature stop codon.

[0037] "Donor DNA", "donor template" and "repair template" as used interchangeably herein refers to a double-stranded DNA fragment or molecule that includes at least a portion of the gene of interest. The donor DNA may encode a full-functional protein or a partially-functional protein.

[0038] "Duchenne Muscular Dystrophy" or "DMD" as used interchangeably herein refers to a recessive, fatal, X-linked disorder that results in muscle degeneration and eventual death. DMD is a common hereditary monogenic disease and occurs in 1 in 3500 males. DMD is the result of inherited or spontaneous mutations that cause nonsense or frame shift mutations in the dystrophin gene. The majority of dystrophin mutations that cause DMD are deletions of exons that disrupt the reading frame and cause premature translation termination in the dystrophin gene. DMD patients typically lose the ability to physically support themselves during childhood, become progressively weaker during the teenage years, and die in their twenties.

[0039] "Dystrophin" as used herein refers to a rod-shaped cytoplasmic protein which is a part of a protein complex that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. Dystrophin provides structural stability to the dystroglycan complex of the cell membrane that is responsible for regulating muscle cell integrity and function. The dystrophin gene or "DMD gene" as used interchangeably herein is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons code for the protein which is over 3500 amino acids.

[0040] "Exon 51" as used herein refers to the 51st exon of the dystrophin gene. Exon 51 is frequently adjacent to frame-disrupting deletions in DMD patients and has been targeted in clinical trials for oligonucleotide-based exon skipping. A clinical trial for the exon 51 skipping compound eteplirsen recently reported a significant functional benefit across 48 weeks, with an average of 47% dystrophin positive fibers compared to baseline. Mutations in exon 51 are ideally suited for permanent correction by NHEJ-based genome editing.

[0041] "Frameshift" or "frameshift mutation" as used interchangeably herein refers to a type of gene mutation wherein the addition or deletion of one or more nucleotides causes a shift in the reading frame of the codons in the mRNA. The shift in reading frame may lead to the alteration in the amino acid sequence at protein translation, such as a missense mutation or a premature stop codon.

[0042] "Functional" and "full-functional" as used herein describes protein that has biological activity. A "functional gene" refers to a gene transcribed to mRNA, which is translated to a functional protein.

[0043] "Genetic construct" as used herein refers to the DNA or RNA molecules that comprise a nucleotide sequence that encodes a protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered. As used herein, the term "expressible form" refers to gene constructs that contain the necessary regulatory elements operable linked to a coding sequence that encodes a protein such that when present in the cell of the individual, the coding sequence will be expressed.

[0044] "Genetic disease" as used herein refers to a disease, partially or completely, directly or indirectly, caused by one or more abnormalities in the genome, especially a condition that is present from birth. The abnormality may be a mutation, an insertion or a deletion. The abnormality may affect the coding sequence of the gene or its regulatory sequence. The genetic disease may be, but not limited to DMD, cystic fibrosis, Huntington's chorea, familial hypercholesterolemia (LDL receptor defect), hepatoblastoma, Wilson's disease, congenital hepatic porphyria, inherited disorders of hepatic metabolism, Lesch Nyhan syndrome, sickle cell anemia, thalassaemias, xeroderma pigmentosum, Fanconi's anemia, retinitis pigmentosa, ataxia telangiectasia, Bloom's syndrome, retinoblastoma, and Tay-Sachs disease.

[0045] "Homology-directed repair" or "HDR" as used interchangeably herein refers to a mechanism in cells to repair double strand DNA lesions when a homologous piece of DNA is present in the nucleus, mostly in G2 and S phase of the cell cycle. HDR uses a donor DNA template to guide repair and may be used to create specific sequence changes to the genome, including the targeted addition of whole genes. If a donor template is provided along with the TALENs, then the cellular machinery will repair the break by homologous recombination, which is enhanced several orders of magnitude in the presence of DNA cleavage. When the homologous DNA piece is absent, non-homologous end joining may take place instead.

[0046] "Identical" or "identity" as used herein in the context of two or more nucleic acids or polypeptide sequences means that the sequences have a specified percentage of residues that are the same over a specified region. The percentage may be calculated by optimally aligning the

two sequences, comparing the two sequences over the specified region, determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the specified region, and multiplying the result by 100 to yield the percentage of sequence identity. In cases where the two sequences are of different lengths or the alignment produces one or more staggered ends and the specified region of comparison includes only a single sequence, the residues of single sequence are included in the denominator but not the numerator of the calculation. When comparing DNA and RNA, thymine (T) and uracil (U) may be considered equivalent. Identity may be performed manually or by using a computer sequence algorithm such as BLAST or BLAST 2.0.

[0047] "Mutant gene" or "mutated gene" as used interchangeably herein refers to a gene that has undergone a detectable mutation. A mutant gene has undergone a change, such as the loss, gain, or exchange of genetic material, which affects the normal transmission and expression of the gene. A "disrupted gene" as used herein refers to a mutant gene that has a mutation that causes a premature stop codon. The disrupted gene product is truncated relative to a full-length undisrupted gene product.

[0048] "Non-homologous end joining (NHEJ) pathway" as used herein refers to a pathway that repairs double-strand breaks in DNA by directly ligating the break ends without the need for a homologous template. The template-independent re-ligation of DNA ends by NHEJ is a stochastic, error-prone repair process that introduces random micro-insertions and micro-deletions (indels) at the DNA breakpoint. This method may be used to intentionally disrupt, delete, or alter the reading frame of targeted gene sequences. NHEJ typically uses short homologous DNA sequences called microhomologies to guide repair. These microhomologies are often present in single-stranded overhangs on the end of double-strand breaks. When the overhangs are perfectly compatible, NHEJ usually repairs the break accurately, yet imprecise repair leading to loss of nucleotides may also occur, but is much more common when the overhangs are not compatible.

[0049] "Nuclease mediated NHEJ" as used herein refers to NHEJ that is initiated after a nuclease, such as a TALEN, cuts double stranded DNA.

[0050] "Nucleic acid" or "oligonucleotide" or "polynucleotide" as used herein means at least two nucleotides covalently linked together. The depiction of a single strand also defines the

sequence of the complementary strand. Thus, a nucleic acid also encompasses the complementary strand of a depicted single strand. Many variants of a nucleic acid may be used for the same purpose as a given nucleic acid. Thus, a nucleic acid also encompasses substantially identical nucleic acids and complements thereof. A single strand provides a probe that may hybridize to a target sequence under stringent hybridization conditions. Thus, a nucleic acid also encompasses a probe that hybridizes under stringent hybridization conditions.

[0051] Nucleic acids may be single stranded or double stranded, or may contain portions of both double stranded and single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine and isoguanine. Nucleic acids may be obtained by chemical synthesis methods or by recombinant methods.

[0052] "Operably linked" as used herein means that expression of a gene is under the control of a promoter with which it is spatially connected. A promoter may be positioned 5' (upstream) or 3' (downstream) of a gene under its control. The distance between the promoter and a gene may be approximately the same as the distance between that promoter and the gene it controls in the gene from which the promoter is derived. As is known in the art, variation in this distance may be accommodated without loss of promoter function.

[0053] "Partially-functional" as used herein describes a protein that is encoded by a mutant gene and has less biological activity than a functional protein but more than a non-functional protein.

[0054] "Premature stop codon" or "out-of-frame stop codon" as used interchangeably herein refers to nonsense mutation in a sequence of DNA, which results in a stop codon at location not normally found in the wild-type gene. A premature stop codon may cause a protein to be truncated or shorter compared to the full-length version of the protein.

[0055] "Promoter" as used herein means a synthetic or naturally-derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from

sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter and the CMV IE promoter.

[0056] "Repeat variable diresidue" or "RVD" as used interchangeably herein refers to a pair of adjacent amino acid residues within the DNA recognition motif (also known as "RVD module"), which includes 33-35 amino acids, of the TALE DNA-binding domain. The RVD determines the nucleotide specificity of the RVD module. RVD modules may be combined to produce an RVD array. The "RVD array length" as used herein refers to the number of RVD modules that corresponds to the length of the nucleotide sequence within the target region that is recognized by the TALEN, *i.e.*, the binding region.

[0057] "Spacers" and "spacer region" as used interchangeably herein refers to the region within the target region that is between, but not a part of, the binding regions for two TALENs.

[0058] "Subject" and "patient" as used herein interchangeably refers to any vertebrate, including, but not limited to, a mammal (*e.g.*, cow, pig, camel, llama, horse, goat, rabbit, sheep, hamsters, guinea pig, cat, dog, rat, and mouse, a non-human primate (for example, a monkey, such as a cynomolgous or rhesus monkey, chimpanzee, etc.) and a human). In some embodiments, the subject may be a human or a non-human. The subject or patient may be undergoing other forms of treatment.

[0059] "Target gene" as used herein refers to any nucleotide sequence encoding a known or putative gene product. The target gene may be a mutated gene involved in a genetic disease.

[0060] "Target region" as used herein refers to the region of the target gene to which two or more TALENs are designed to bind and cleave. The target region includes the binding regions for the TALENS and the spacer region, which occurs between the binding regions. The two TALENs bind to different binding regions within the target region, after which the target region is cleaved. For example, the target region in the dystrophin gene may include a nucleotide sequence of SEQ ID NO: 16-25, 44-46, 50, 52, 58, or 59, or a complement thereof. The target

region in the dystrophin gene may include one or more of SEQ ID NO: 26-43, or a complement thereof. Each of SEQ ID NO: 26-43, or a complement thereof, may represent a binding region.

[0061] "Transcription activator-like effector" or "TALE" as used herein refers to a protein structure that recognizes and binds to a particular DNA sequence. The "TALE DNA-binding domain" refers to a DNA-binding domain that includes an array of tandem 33-35 amino acid repeats, also known as RVD modules, each of which specifically recognizes a single base pair of DNA. RVD modules may be arranged in any order to assemble an array that recognizes a defined sequence.

[0062] A binding specificity of a TALE DNA-binding domain is determined by the RVD array followed by a single truncated repeat of 20 amino acids. A TALE DNA-binding domain may have 12 to 27 RVD modules, each of which contains an RVD and recognizes a single base pair of DNA. Specific RVDs have been identified that recognize each of the four possible DNA nucleotides (A, T, C, and G). Because the TALE DNA-binding domains are modular, repeats that recognize the four different DNA nucleotides may be linked together to recognize any particular DNA sequence. These targeted DNA-binding domains may then be combined with catalytic domains to create functional enzymes, including artificial transcription factors, methyltransferases, integrases, nucleases, and recombinases.

[0063] "Transcription activator-like effector nucleases" or "TALENs" as used interchangeably herein refers to engineered fusion proteins of the catalytic domain of a nuclease, such as endonuclease *Fok*I, and a designed TALE DNA-binding domain that may be targeted to a custom DNA sequence. A "TALEN monomer" refers to an engineered fusion protein with a catalytic nuclease domain and a designed TALE DNA-binding domain. Two TALEN monomers may be designed to target and cleave a target region.

[0064] TALENs may be used to introduce site-specific double strand breaks at targeted genomic loci. Site-specific double-strand breaks are created when two independent TALENs bind to nearby DNA sequences, thereby permitting dimerization of *Fok*I and cleavage of the target DNA. TALENs have advanced genome editing due to their high rate of successful and efficient genetic modification. This DNA cleavage may stimulate the natural DNA-repair machinery, leading to one of two possible repair pathways: homology-directed repair (HDR) or the non-homologous end joining (NHEJ) pathway. For example, a TALEN directed towards the dystrophin gene may include an amino acid sequence of any one of SEQ ID NO: 1-15.

[0065] "Variant" used herein with respect to a nucleic acid means (i) a portion or fragment of a referenced nucleotide sequence; (ii) the complement of a referenced nucleotide sequence or portion thereof; (iii) a nucleic acid that is substantially identical to a referenced nucleic acid or the complement thereof; or (iv) a nucleic acid that hybridizes under stringent conditions to the referenced nucleic acid, complement thereof, or a sequences substantially identical thereto.

[0066] "Variant" with respect to a peptide or polypeptide that differs in amino acid sequence by the insertion, deletion, or conservative substitution of amino acids, but retain at least one biological activity. Variant may also mean a protein with an amino acid sequence that is substantially identical to a referenced protein with an amino acid sequence that retains at least one biological activity. A conservative substitution of an amino acid, *i.e.*, replacing an amino acid with a different amino acid of similar properties (*e.g.*, hydrophilicity, degree and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes may be identified, in part, by considering the hydropathic index of amino acids, as understood in the art. Kyte *et al.*, *J. Mol. Biol.* 157:105-132 (1982). The hydropathic index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydropathic indexes may be substituted and still retain protein function. In one aspect, amino acids having hydropathic indexes of ±2 are substituted. The hydrophilicity of amino acids may also be used to reveal substitutions that would result in proteins retaining biological function. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide. Substitutions may be performed with amino acids having hydrophilicity values within ±2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties.

[0067] "Vector" as used herein means a nucleic acid sequence containing an origin of replication. A vector may be a viral vector, bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self-replicating extrachromosomal vector, and preferably, is a DNA plasmid. For example, the vector may encode a TALEN protein comprising the amino acid sequence of one of SEQ ID NO: 1-15.

[0068] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

2. TALENs

[0069] Provided herein are TALENs for use in genome editing and treating genetic diseases. The TALENs may be designed to target any gene involved in a genetic disease. The TALENs may include a nuclease and a TALE DNA-binding domain that binds to the target gene. The target gene may have a mutation such as a frameshift mutation or a nonsense mutation. If the target gene has a mutation that causes a premature stop codon, the TALEN may be designed to recognize and bind a nucleotide sequence upstream or downstream from the premature stop codon.

[0070] The TALE DNA-binding domain may have an RVD array length between 1-30 modules, between 1-25 modules, between 1-20 modules, between 1-15 modules, between 5-30 modules, between 5-25 modules, between 5-20 modules, between 5-15 modules, between 7-25 modules, between 7-23 modules, between 7-20 modules, between 10-30 modules, between 10-25 modules, between 10-20 modules, between 10-15 modules, between 15-30 modules, between 15-25 modules, between 15-20 modules, between 15-19 modules, between 16-26 modules, between 16-41 modules, between 20-30 modules, or between 20-25 modules in length. The RVD array length may be 5 modules, 8 modules, 10 modules, 11 modules, 12 modules, 13 modules, 14 modules, 15 modules, 16 modules, 17 modules, 18 modules, 19 modules, 20 modules, 22 modules, 25 modules or 30 modules.

a. Dystrophin

[0071] Dystrophin is a rod-shaped cytoplasmic protein which is a part of a protein complex that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. Dystrophin provides structural stability to the dystroglycan complex of the

cell membrane. The dystrophin gene is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons code for the protein which is over 3500 amino acids. Normal skeleton muscle tissue contains only small amounts of dystrophin but its absence of abnormal expression leads to the development of severe and incurable symptoms. Some mutations in the dystrophin gene lead to the production of defective dystrophin and severe dystrophic phenotype in affected patients. Some mutations in the dystrophin gene lead to partially-functional dystrophin protein and a much milder dystrophic phenotype in affected patients.

[0072] DMD is the result of inherited or spontaneous mutations that cause nonsense or frame shift mutations in the dystrophin gene. Naturally occurring mutations and their consequences are relatively well understood for DMD. It is known that in-frame deletions that occur in the exon 45-55 region contained within the rod domain can produce highly functional dystrophin proteins, and many carriers are asymptomatic or display mild symptoms. Furthermore, more than 60% of patients may theoretically be treated by targeting exons in this region of the dystrophin gene. Efforts have been made to restore the disrupted dystrophin reading frame in DMD patients by skipping non-essential exons during mRNA splicing to produce internally deleted but functional dystrophin proteins. The deletion of internal dystrophin exons retain the proper reading frame but cause the less severe Becker muscular dystrophy.

b. TALENs for dystrophin

[0073] TALENs specific for dystrophin gene are disclosed herein. The TALENs may include a nuclease and a TALE DNA-binding domain that binds to the dystrophin gene. The TALENs may bind and recognize a target region. TALEN target regions may be chosen immediately upstream of possible out-of-frame stop codons such that insertions or deletions during the repair process restore the dystrophin reading frame in either disrupted frame.

[0074] The TALE domain may bind to a nucleic acid sequence of SEQ ID NOs: 16-46, 50, 52, 58, or 59, or a complement thereof. The TALEN may include an amino acid sequence selected from the group consisting of SEQ ID NO: 1-15, or a complement thereof. The nuclease may be *Fok*I. For example, the disclosed TALENs were engineered to mediate highly efficient gene editing at exon 51 and the 5' UTR of the dystrophin gene. These TALENs restored dystrophin protein expression in cells from DMD patients, including skeletal myoblasts and dermal fibroblasts that were reprogrammed to the myogenic lineage by MyoD, with no TALEN-

mediated off-target changes to the protein-coding regions of the genome. The TALENS may have varying RVD array sequences and lengths. Examples of RVD arrays may be found in Table 2. The TALENS may have binding regions separated by varying spacer regions. Examples of spacer regions may be found in Table 3 as indicated as the nucleotide sequence that was not underlined.

(1) Exon 51

[0075] Exon 51 is frequently adjacent to frame-disrupting deletions in DMD and occurs in approximately 20% of all DMD patients. This class of DMD mutations is ideally suited for permanent correction by NHEJ-based genome editing and HDR. TALENs described herein have been developed for targeted modification of exon 51 in the human dystrophin gene. These TALENs were transfected into human DMD cells and mediated efficient gene modification and conversion to the correct reading frame. Protein restoration was concomitant with frame restoration and detected in a bulk population of TALEN-treated cells.

[0076] The TALE DNA-binding domain of these TALENs may bind to a nucleic acid sequence of SEQ ID NOs: 16-24, 26-41, 46, 50, 52, 58, or 59, or a complement thereof. The TALEN protein may include an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13, or a complement thereof. The nuclease may be *Fok*I.

(2) 5' UTR

[0077] TALENs described herein have been developed for targeted modification of the 5' untranslated region (UTR) of the human dystrophin gene. TALENs engineered to target the 5' UTR of the dystrophin gene allows the integration of a replacement dystrophin gene, such as a full-functional dystrophin gene or a minidystrophin expression cassette, to be under the control of the endogenous promoter and regulatory sequences in the normal chromosomal context. Targeting the gene region upstream of the dystrophin coding sequence thereby allows tissue-specific expression of the dystrophin protein at physiologic expression levels. This approach may be used to treat all of the possible mutations within the dystrophin-encoding sequence.

[0078] These TALENs may include a nuclease and a TALE DNA-binding domain that binds to 5' UTR of the human dystrophin gene. The TALE DNA-binding domain may bind to a nucleic acid sequence of SEQ ID NOs: 25 or 42-45, or a complement thereof. The TALEN

protein may include an amino acid sequence selected from the group consisting of SEQ ID NO: 14 and SEQ ID NO: 15, or a complement thereof. The nuclease may be *Fok*I.

3. Compositions

[0079] The present disclosure also provides compositions of two or more TALENs, as described above. These compositions may be used in genome editing. The composition includes a first TALEN, which binds to a first binding region, and a second TALEN, which binds to a second binding region. The first binding region and second binding region are located within a target region or the target gene. The first binding region and second binding region are not the same; the first TALEN binds to a binding region that is different binding region from where the second TALEN binds. The binding regions are separated by at least one base pair. The spacer region between the binding regions of the target region may be between 1-30 bp, between 1-25 bp, between 1-20 bp, between 1-15 bp, between 5-30 bp, between 5-25 bp, between 5-20 bp, between 5-15 bp, between 7-25 bp, between 7-23 bp, between 7-20 bp, between 10-30 bp, between 10-25 bp, between 10-20 bp, between 10-15 bp, between 14-25 bp, between 14-19 bp, between 15-30 bp, between 15-25 bp, between 15-20 bp, between 20-30 bp, or between 20-25 bp in length. The spacer region may be 5 bp, 10 bp, 11 bp, 12 bp, 13 bp, 14 bp, 15 bp, 16 bp, 17 bp, 18 bp, 19 bp, 20 bp, 21 bp, 22 bp, 23 bp, 24 bp, 25 bp, or 30 bp in length.

[0080] The two or more TALEN protein may be selected from TALEN proteins that include an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, or SEQ ID NO: 15, or a complement thereof. The composition may include two TALEN proteins having respectively the amino acid sequence of: SEQ ID NO: 1 and SEQ ID NO: 3; SEQ ID NO: 1 and SEQ ID NO: 4; SEQ ID NO: 1 and SEQ ID NO: 8; SEQ ID NO: 1 and SEQ ID NO: 9; SEQ ID NO: 1 and SEQ ID NO: 11; SEQ ID NO: 1 and SEQ ID NO: 13; SEQ ID NO: 2 and SEQ ID NO: 3; SEQ ID NO: 2 and SEQ ID NO: 4; SEQ ID NO: 2 and SEQ ID NO: 8; SEQ ID NO: 2 and SEQ ID NO: 9; SEQ ID NO: 2 and SEQ ID NO: 11; SEQ ID NO: 2 and SEQ ID NO: 13; SEQ ID NO: 5 and SEQ ID NO: 3; SEQ ID NO: 5 and SEQ ID NO: 4; SEQ ID NO: 5 and SEQ ID NO: 8; SEQ ID NO: 5 and SEQ ID NO: 9; SEQ ID NO: 5 and SEQ ID NO: 11; SEQ ID NO: 5 and SEQ ID NO: 13; SEQ ID NO: 6 and SEQ ID NO: 3; SEQ ID NO: 6 and SEQ ID NO: 4; SEQ ID NO: 6 and SEQ ID NO: 8; SEQ ID NO: 6 and SEQ ID NO:

9; SEQ ID NO: 6 and SEQ ID NO: 11; SEQ ID NO: 6 and SEQ ID NO: 13; SEQ ID NO: 7 and SEQ ID NO: 3; SEQ ID NO: 7 and SEQ ID NO: 4; SEQ ID NO: 7 and SEQ ID NO: 8; SEQ ID NO: 7 and SEQ ID NO: 9; SEQ ID NO: 7 and SEQ ID NO: 11; SEQ ID NO: 7 and SEQ ID NO: 13; SEQ ID NO: 10 and SEQ ID NO: 3; SEQ ID NO: 10 and SEQ ID NO: 4; SEQ ID NO: 10 and SEQ ID NO: 8; SEQ ID NO: 10 and SEQ ID NO: 9; SEQ ID NO: 10 and SEQ ID NO: 11; SEQ ID NO: 10 and SEQ ID NO: 13; SEQ ID NO: 12 and SEQ ID NO: 3; SEQ ID NO: 12 and SEQ ID NO: 4; SEQ ID NO: 12 and SEQ ID NO: 8; SEQ ID NO: 12 and SEQ ID NO: 9; SEQ ID NO: 12 and SEQ ID NO: 11; SEQ ID NO: 12 and SEQ ID NO: 13; or SEQ ID NO: 14 and SEQ ID NO:15.

4. Methods of Correcting a Mutant Gene and Treating a Subject

[0081] The present disclosure also provides methods of correcting a mutant gene in a cell and treating a subject suffering from a genetic disease, such as DMD. The method may include administering to a cell or subject a TALEN protein, a polynucleotide or vector encoding said TALEN protein, or composition of TALENs as described above. The method may include administering a first TALEN, which binds to a first binding region, and a second TALEN, which binds to a second binding region. The first binding region and second binding region are located within a target region or the target gene. The first binding region and second binding region are not the same; the first TALEN binds to a binding region that is different binding region from where the second TALEN binds. The binding regions are separated by at least one base pair. The spacer region between the binding regions of the target region may be between 1-30 bp, between 1-25 bp, between 1-20 bp, between 1-15 bp, between 5-30 bp, between 5-25 bp, between 5-20 bp, between 5-15 bp, between 7-25 bp, between 7-23 bp, between 7-20 bp, between 10-30 bp, between 10-25 bp, between 10-20 bp, between 10-15 bp, between 14-25 bp, between 14-19 bp, between 15-30 bp, between 15-25 bp, between 15-20 bp, between 20-30 bp, or between 20-25 bp in length. The spacer region may be 5 bp, 10 bp, 11 bp, 12 bp, 13 bp, 14 bp, 15 bp, 16 bp, 17 bp, 18 bp, 19 bp, 20 bp, 21 bp, 22 bp, 23 bp, 24 bp, 25 bp, or 30 bp in length. The two or more TALEN protein may be selected from the TALEN proteins described above. The method may involve homology-directed repair or non-homologous end joining.

a. Homology-Directed Repair

[0082] As described herein, the inventors show restoration of protein expression from an endogenous mutated gene using a method involving homology-directed repair. The method as

described above further includes administrating a donor template to the cell. The donor template may include a nucleotide sequence encoding a full-functional protein or a partially-functional protein. For example, the donor template may include a miniaturized dystrophin construct, termed minidystrophin ("minidys") or a full-functional dystrophin construct for restoring a mutant dystrophin gene.

b. Nuclease mediated non-homologous end joining

[0083] As described herein, the inventors show restoration of protein expression from an endogenous mutated gene through template-free NHEJ-mediated DNA repair. In contrast to a transient method targeting the target gene RNA, the correction of the target gene reading frame in the genome by transiently expressed TALENs may lead to permanently restored target gene expression by each modified cell and all of its progeny.

[0084] Nuclease mediated NHEJ gene correction may correct the mutated target gene and offers several potential advantages over the HDR pathway. For example, NHEJ does not require a donor template, which may cause nonspecific insertional mutagenesis. In contrast to HDR, NHEJ operates efficiently in all stages of the cell cycle and therefore may be effectively exploited in both cycling and post-mitotic cells, such as muscle fibers. This provides a robust, permanent gene restoration alternative to oligonucleotide-based exon skipping or pharmacologic forced read-through of stop codons and could theoretically require as few as one drug treatment. NHEJ-based gene correction using TALENs, as well as other engineered nucleases including meganucleases and zinc finger nucleases, may be combined with other existing *ex vivo* and *in vivo* platforms for cell- and gene-based therapies, in addition to the plasmid electroporation approach described here. For example, delivery of TALENs by mRNA-based gene transfer or as purified cell permeable proteins could enable a DNA-free genome editing approach that would circumvent any possibility of insertional mutagenesis.

c. Duchenne muscular dystrophy

[0085] The method, as described above, may be used for correcting the dystrophin gene and recovering full-functional or partially-functional protein expression of said mutated dystrophin gene. In some aspects and embodiments the disclosure provides a method for reducing the effects (*e.g.*, clinical symptoms/indications) of DMD in a patient. In some aspects and embodiments the disclosure provides a method for treating DMD in a patient. In some aspects and embodiments the disclosure provides a method for preventing DMD in a patient. In some

aspects and embodiments the disclosure provides a method for preventing further progression of DMD in a patient.

5. Constructs and Plasmids

[0086] The genetic constructs may comprise a nucleic acid sequence that encodes the TALEN disclosed herein. The genetic construct, such as a plasmid, may comprise a nucleic acid that encodes the TALEN. The genetic construct may be present in the cell as a functioning extrachromosomal molecule. The genetic construct may be a linear minichromosome including centromere, telomeres or plasmids or cosmids.

[0087] The genetic construct may also be part of a genome of a recombinant viral vector, including recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The genetic construct may be part of the genetic material in attenuated live microorganisms or recombinant microbial vectors which live in cells. The genetic constructs may comprise regulatory elements for gene expression of the coding sequences of the nucleic acid. The regulatory elements may be a promoter, an enhancer an initiation codon, a stop codon, or a polyadenylation signal.

[0088] The nucleic acid sequences may make up a genetic construct that may be a vector. The vector may be capable of expressing the TALEN in the cell of a mammal. The vector may be recombinant. The vector may comprise heterologous nucleic acid encoding the TALEN. The vector may be a plasmid. The vector may be useful for transfecting cells with nucleic acid encoding the TALEN, which the transformed host cell is cultured and maintained under conditions wherein expression of the TALEN takes place.

[0089] Coding sequences may be optimized for stability and high levels of expression. In some instances, codons are selected to reduce secondary structure formation of the RNA such as that formed due to intramolecular bonding.

[0090] The vector may comprise heterologous nucleic acid encoding the TALEN and may further comprise an initiation codon, which may be upstream of the TALEN coding sequence, and a stop codon, which may be downstream of the TALEN coding sequence. The initiation and termination codon may be in frame with the TALEN coding sequence. The vector may also comprise a promoter that is operably linked to the TALEN coding sequence. The promoter operably linked to the TALEN coding sequence may be a promoter from simian virus 40 (SV40), a mouse mammary tumor virus (MMTV) promoter, a human immunodeficiency virus (HIV)

promoter such as the bovine immunodeficiency virus (BIV) long terminal repeat (LTR) promoter, a Moloney virus promoter, an avian leukosis virus (ALV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter, Epstein Barr virus (EBV) promoter, or a Rous sarcoma virus (RSV) promoter. The promoter may also be a promoter from a human gene such as human actin, human myosin, human hemoglobin, human muscle creatine, or human metalothionein. The promoter may also be a tissue specific promoter, such as a muscle or skin specific promoter, natural or synthetic. Examples of such promoters are described in US Patent Application Publication No. US20040175727, the contents of which are incorporated herein in its entirety.

[0091] The vector may also comprise a polyadenylation signal, which may be downstream of the TALEN. The polyadenylation signal may be a SV40 polyadenylation signal, LTR polyadenylation signal, bovine growth hormone (bGH) polyadenylation signal, human growth hormone (hGH) polyadenylation signal, or human β-globin polyadenylation signal. The SV40 polyadenylation signal may be a polyadenylation signal from a pCEP4 vector (Invitrogen, San Diego, CA).

[0092] The vector may also comprise an enhancer upstream of the TALEN. The enhancer may be necessary for DNA expression. The enhancer may be human actin, human myosin, human hemoglobin, human muscle creatine or a viral enhancer such as one from CMV, HA, RSV or EBV. Polynucleotide function enhancers are described in U.S. Patent Nos. 5,593,972, 5,962,428, and WO94/016737, the contents of each are fully incorporated by reference. The vector may also comprise a mammalian origin of replication in order to maintain the vector extrachromosomally and produce multiple copies of the vector in a cell. The vector may also comprise a regulatory sequence, which may be well suited for gene expression in a mammalian or human cell into which the vector is administered. The vector may also comprise a reporter gene, such as green fluorescent protein ("GFP") and/or a selectable marker, such as hygromycin ("Hygro").

[0093] The vector may be expression vectors or systems to produce protein by routine techniques and readily available starting materials including Sambrook *et al.*, Molecular Cloning and Laboratory Manual, Second Ed., Cold Spring Harbor (1989), which is incorporated fully by reference. In some embodiments the vector may comprise the nucleic acid sequence encoding the TALEN comprising the amino acid sequence of at least one of SEQ ID NOs:1-15

6. Pharmaceutical Compositions

[0094] The TALEN may be in a pharmaceutical composition. The pharmaceutical composition may comprise about 1 ng to about 10 mg of DNA of the TALEN or TALEN protein. The pharmaceutical compositions according to the present invention are formulated according to the mode of administration to be used. In cases where pharmaceutical compositions are injectable pharmaceutical compositions, they are sterile, pyrogen free and particulate free. An isotonic formulation is preferably used. Generally, additives for isotonicity may include sodium chloride, dextrose, mannitol, sorbitol and lactose. In some cases, isotonic solutions such as phosphate buffered saline are preferred. Stabilizers include gelatin and albumin. In some embodiments, a vasoconstriction agent is added to the formulation.

[0095] The TALEN may further comprise a pharmaceutically acceptable excipient. The pharmaceutically acceptable excipient may be functional molecules as vehicles, adjuvants, carriers, or diluents. The pharmaceutically acceptable excipient may be a transfection facilitating agent, which may include surface active agents, such as immune-stimulating complexes (ISCOMS), Freunds incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs, vesicles such as squalene and squalene, hyaluronic acid, lipids, liposomes, calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents.

[0096] The transfection facilitating agent is a polyanion, polycation, including poly-L-glutamate (LGS), or lipid. The transfection facilitating agent is poly-L-glutamate, and more preferably, the poly-L-glutamate is present in the TALEN at a concentration less than 6 mg/ml. The transfection facilitating agent may also include surface active agents such as immune-stimulating complexes (ISCOMS), Freunds incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs and vesicles such as squalene and squalene, and hyaluronic acid may also be used administered in conjunction with the genetic construct. In some embodiments, the DNA vector TALEN may also include a transfection facilitating agent such as lipids, liposomes, including lecithin liposomes or other liposomes known in the art, as a DNA-liposome mixture (see for example W09324640), calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents. Preferably, the transfection facilitating agent is a polyanion, polycation, including poly-L-glutamate (LGS), or lipid.

7. Methods of Delivery

[0097] Provided herein is a method for delivering the pharmaceutical formulations, preferably TALENs, for providing genetic constructs and proteins of the TALENs. The delivery of the TALEN may be the transfection or electroporation of the TALEN as a nucleic acid molecule that is expressed in the cell and delivered to the surface of the cell. The nucleic acid molecules may be electroporated using BioRad Gene Pulser Xcell or Amaxa Nucleofector IIb devices. Several different buffers may be used, including BioRad electroporation solution, Sigma phosphate-buffered saline product #D8537 (PBS), Invitrogen OptiMEM I (OM), or Amaxa Nucleofector solution V (N.V.). Transfections may include a transfection reagent, such as Lipofectamine 2000.

[0098] Upon delivery of the TALEN to the cell, and thereupon the vector into the cells of the mammal, the transfected cells will express the TALEN. The TALEN may be administered to a mammal to correct the dystrophin gene in a mammal. The mammal may be human, non-human primate, cow, pig, sheep, goat, antelope, bison, water buffalo, bovids, deer, hedgehogs, elephants, llama, alpaca, mice, rats, or chicken, and preferably human, cow, pig, or chicken.

8. Routes of Administration

[0099] The TALEN proteins and compositions may be administered to a subject by different routes including orally, parenterally, sublingually, transdermally, rectally, transmucosally, topically, via inhalation, via buccal administration, intrapleurally, intravenous, intraarterial, intraperitoneal, subcutaneous, intramuscular, intranasal intrathecal, and intraarticular or combinations thereof. For veterinary use, the composition may be administered as a suitably acceptable formulation in accordance with normal veterinary practice. The veterinarian may readily determine the dosing regimen and route of administration that is most appropriate for a particular animal. The TALEN proteins and compositions may be administered by traditional syringes, needleless injection devices, "microprojectile bombardment gone guns", or other physical methods such as electroporation ("EP"), "hydrodynamic method", or ultrasound.

[00100] The vector encoding a TALEN protein may be delivered to the mammal by several technologies including DNA injection (also referred to as DNA vaccination) with and without in vivo electroporation, liposome mediated, nanoparticle facilitated, recombinant vectors such as recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus.

[00101] The nucleotide encoding a TALEN protein may be introduced into a cell to genetically correct the target gene. For example, a nucleotide encoding a TALEN protein directed towards a mutant dystrophin gene may be introduced into a myoblast cell from a DMD patient. The genetically corrected myoblast cell may be treated with MyoD to induce differentiation into fibroblasts, which may be implanted into subjects, such as the damaged muscles of a subject to verify that the corrected dystrophin protein was functional and/or to treat the subject.

9. Cell types

[00102] Any of these delivery methods and/or routes of administration could be utilized with a myriad of cell types currently under investigation for cell-based therapies, including immortalized myoblast cells, such as wild-type and DMD patient derived lines, for example Δ48-50 DMD, C25C14 and DMD-7796 cell lines, primal DMD dermal fibroblasts, induced pluripotent stem cells, bone marrow-derived progenitors, skeletal muscle progenitors, human skeletal myoblasts from DMD patients, CD133+ cells, mesoangioblasts, and MyoD- or Pax7-transduced cells, or other myogenic progenitor cells. Immortalization of human myogenic cells may be used for clonal derivation of genetically corrected myogenic cells.

[00103] Cells may be modified *ex vivo* to isolate and expand clonal populations of immortalized DMD myoblasts that contain a genetically corrected dystrophin gene and are free of nuclease-introduced mutations in protein coding regions of the genome. Alternatively, transient *in vivo* delivery of nucleases by non-viral or non-integrating viral gene transfer, or by direct delivery of purified proteins containing cell-penetrating motifs may enable highly specific correction *in situ* with minimal or no risk of exogenous DNA integration.

10. Kits

[00104] Provided herein is a kit, which may be used to correct a mutated gene. The kit comprises at least one component for correcting a mutated gene and instructions for using the TALEN. Instructions included in kits may be affixed to packaging material or may be included as a package insert. While the instructions are typically written or printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (*e.g.*, magnetic discs, tapes, cartridges, chips), optical media (*e.g.*, CD ROM), and the like. As used herein, the term "instructions" may include the address of an internet site that provides the instructions.

[00105] At least one component may include at least one TALEN, as described above, that specifically binds and cleaves the mutated gene when a second TALEN is present nearby. Two or more TALENs, as described above, may be included in the kit to specifically bind and target a particular target region in the mutated gene. The TALEN may be specific for a mutated dystrophin gene, as described above. The kit may include donor DNA, as described above.

11. Examples

[00106] It will be readily apparent to those skilled in the art that other suitable modifications and adaptations of the methods of the present disclosure described herein are readily applicable and appreciable, and may be made using suitable equivalents without departing from the scope of the present disclosure or the aspects and embodiments disclosed herein. Having now described the present disclosure in detail, the same will be more clearly understood by reference to the following examples, which are merely intended only to illustrate some aspects and embodiments of the disclosure, and should not be viewed as limiting to the scope of the disclosure. The disclosures of all journal references, U.S. patents, and publications referred to herein are hereby incorporated by reference in their entireties.

[00107] The present invention has multiple aspects, illustrated by the following non-limiting examples.

Example 1

Materials and Methods

[00108] ***Cell culture and transfection.*** HEK293T cells were obtained from the American Tissue Collection Center (ATCC) through the Duke Cell Culture Facility and were maintained in DMEM supplemented with 10% bovine calf serum and 1% penicillin/streptomycin. Immortalized myoblasts (Mamchaoui, K. *et al. Skelet Muscle* 1, 1-11 (2011)) (one from a wild-type donor, and two Δ48-50 DMD patient derived lines) were maintained in skeletal muscle media (PromoCell) supplemented with 20% bovine calf serum (Sigma), 50 μg/ml fetuin, 10 ng/ml human epidermal growth factor (Sigma), 1 ng/ml human basic fibroblast growth factor (Sigma), 10 μg/ml human insulin (Sigma), 1% GlutaMAX (Invitrogen), and 1% penicillin/streptomycin (Invitrogen). Primary DMD dermal fibroblasts were obtained from the Coriell Cell repository (GM05162A, Δ46-50) and maintained in DMEM supplemented with 10%

fetal bovine serum, 1 ng/mL human basic fibroblast growth factor, and 1% penicillin/streptomycin. All cell lines were maintained at 37°C and 5% CO_2.

[00109] HEK293T cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol in 24 well plates. Immortalized myoblasts and primary fibroblasts were transfected by electroporation using the Gene Pulser XCell (BioRad) with PBS as an electroporation buffer using optimized conditions for each line (Figure 1). Transfection efficiencies were measured by delivering an EGFP expression plasmid and using flow cytometry. These efficiencies were routinely ≥95% for HEK293T and ≥70% for the primary fibroblasts and immortalized myoblasts. For all experiments, the indicated mass of electroporated plasmid corresponds to the amount used for each TALEN monomer.

[00110] ***TALE nuclease assembly and off-target site prediction.*** TALENs targeted to exon 51 of the human dystrophin gene were designed *in silico* using the TALE-NT webserver (Cermak, T. *et al. Nucleic Acids Res* 39, e82 (2011)). TALEN target sites were chosen to include half-site targets approximately 15-19 bp in length, preceded by a 5'-T (Miller, J.C. *et al. Nat Biotechnol* 29, 143-148 (2011)). Plasmids encoding these TALENs were assembled using the Golden Gate assembly method (Cermak, T. *et al. Nucleic Acids Res* 39, e82 (2011)) and standard cloning techniques into a modified pcDNA3.1 (Invitrogen) destination vector containing the Δ152/+63 TALEN architecture (Miller, J.C. *et al. Nat Biotechnol* 29, 143-148 (2011)) derived from the pTAL3 expression vector provided in the Golden Gate kit from Addgene. The *Fok*I endonuclease domains were codon optimized and contained the ELD/KKR obligate heterodimer (Doyon, Y. *et al. Nat Methods* 8, 74-79 (2010)) and *Sharkey* mutations (Guo, J. *et al. J Mol Biol* (2010)) as described previously (Perez-Pinera, P. *et al. Nucleic Acids Res* 40, 3741-3752 (2012)). Complete sequences, including TN3/8 sequences, are provided in Figure 2. Potential off-target sites for TALEN pairs, including TN3/8, in the human genome were predicted *in silico* using the Paired Target Finder tool on the TALE-NT 2.0 webserver (Doyle, E.L. *et al. Nucleic Acids Res* 40, W117-122 (2012)). All predicted off-target sites were scanned using the following parameters: recommended score cutoff (3.0), spacers of range 12-23 bp, and upstream base set to "T only". Valid likely potential off-target sites were only considered as those with up to 4 mismatches per TALEN half-site binding sequence (maximum of 8 mismatches per TALEN pair target site). Plasmid DNA encoding the modules that were used to construct the TALEs was obtained from the Addgene non-profit plasmid repository.

[00111] ***Cel-I quantification of endogenous gene modification.*** TALEN-induced lesions at the endogenous target site were quantified using the Surveyor nuclease assay (Guschin, D.Y. *et al. Meth Mol Biol* 649, 247-256 (2010)), which can detect mutations characteristic of nuclease-mediated NHEJ. After electroporation, cells were incubated for 3 or 10 days at 37°C and genomic DNA was extracted using the DNeasy Blood and Tissue kit (QIAGEN). The target locus was amplified by 30 cycles of PCR with the AccuPrime High Fidelity PCR kit (Invitrogen) using primers 5'- GAGTTTGGCTCAAATTGTTACTCTT-3' (SEQ ID NO: 60) and 5'-GGGAAATGGTCTAGGAGAGTAAAGT-3' (SEQ ID NO: 61). The resulting PCR products were randomly melted and reannealed in a PCR machine with the program: 95°C for 240 s, followed by 85°C for 60 s, 75°C for 60 s, 65°C for 60 s, 55°C for 60 s, 45°C for 60 s, 35°C for 60 s, and 25°C for 60 s with a -0.3°C/s rate between steps. Following reannealing, 8 μL of PCR product was mixed with 1 μL of Surveyor Nuclease S and 1 μL of Enhancer S (Transgenomic) and incubated at 42°C for 1 hour. After incubation, 6 μL of digestion product was loaded onto a 10% TBE polyacrylamide gel and run at 200V for 30 min. The gels were stained with ethidium bromide and quantified using ImageLab (Bio-Rad) by densitometry as previously described (Guschin, D.Y. *et al. Meth Mol Biol* 649, 247-256 (2010)).

[00112] ***Cytotoxicity assay.*** To quantitatively assess potential TALEN cytotoxicity, HEK293T cells were transfected with 10 ng of a GFP reporter and 100 ng of each nuclease using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). The percentage of GFP positive cells was assessed at 2 and 5 days by flow cytometry. The survival rate was calculated as the decrease in GFP positive cells from days 2 to 5 and normalized to cells transfected with an empty nuclease expression vector as described (Cornu, T.I. *et al. Meth Mol Biol* 649, 237-245 (2010)).

[00113] ***Clone isolation procedure.*** Immortalized DMD myoblasts were electroporated with 10 μg of each TALEN plasmid (20 μg total). After 7 days, isogenic clones were isolated by clonal dilution in hypoxic conditions (5% O_2) to accelerate myoblast growth. Genomic DNA was extracted from clones using the QuickExtract Kit (Epicentre) and the target locus amplified by PCR using the Cel-I primers and conditions above. The resulting PCR products were either mixed with equal amounts of PCR product from untreated cells and analyzed by the Surveyor assay, as described above, or directly submitted for conventional Sanger sequencing to identify modified clones.

[00114] ***Viral transduction and forced MyoD overexpression in primary fibroblasts.*** 300,000 fibroblasts were plated transduced in 10 cm plates with a lentiviral vector encoding a full-length human MyoD cDNA under the control of a dox-inducible promoter and a constitutive puromycin resistance cassette. Two days post-transduction, fibroblasts were selected for 6 days in 1 μg/mL puromycin (Sigma) to enrich for transduced cells. Fibroblasts were then plated at a density of 200,000 cells in 10 cm dishes and MyoD expression was induced by adding 3 μg/mL doxycycline (Fisher Scientific) to the media, which was exchanged every two days.

[00115] ***Western blot analysis.*** To assess dystrophin expression, immortalized myoblasts were differentiated into myofibers by replacing the growth medium with DMEM supplemented with 1% insulin-transferrin-selenium (Invitrogen) and 1% antibiotic/antimycotic (Invitrogen) for 4-7 days. Fibroblasts were transdifferentiated into myoblasts by inducing MyoD overexpression and incubating the cells in DMEM supplemented with 1% insulin-transferrin-selenium (Invitrogen), 1% antibiotic/antimycotic (Invitrogen) and 3 μg/mL doxycycline for 15 days. TALEN expression was assessed at 3 days after transfecting HEK293T cells. Cells were collected and lysed in RIPA buffer (Sigma) supplemented with a protease inhibitor cocktail (Sigma) and the total protein amount was quantified using the bicinchoninic acid assay according to the manufacturer's instructions (Pierce). Samples were then mixed with NuPAGE loading buffer (Invitrogen) and 5% β-mercaptoethanol and heated to 85°C for 10 minutes. Twenty-five micrograms of protein were separated on 4-12% NuPAGE Bis-Tris gels (Invitrogen) with MES buffer (Invitrogen). Proteins were transferred to nitrocellulose membranes for 1-2 hours in transfer buffer containing 10-20% methanol and 0.01% SDS. The blot was then blocked for 1 hour with 5% milk-TBST at room temperature. Blots were probed with the following primary antibodies: NCL-Dys2 (1:25, Leica), MANDYS8 (1:100, Sigma), GAPDH (1:5000, Cell Signaling), anti-FLAG-HRP (1:2000, Cell Signaling), or anti-myogenin F5D (1:200, Santa Cruz). Dystrophin expression was detected using MANDYS8 in DMD myoblast line 2 and the DMD fibroblast line or NCL-Dys2 in DMD myoblast line 1. TALEN expression was detected using anti-FLAG. Blots were then incubated with mouse or rabbit horseradish peroxidase-conjugated secondary antibodies (Santa Cruz) and visualized using the ChemiDoc chemilumescent system (BioRad) and Western-C ECL substrate (BioRad).

[00116] ***Immunofluorescence.*** Fibroblasts were plated on cover slips in 24 well plates at a density of 30,000 cells/well and MyoD expression was induced for 15 days as described above.

Cells were then fixed in 4% paraformaldehyde and blocked for 1 hour at room temperature with PBS containing 5% BSA, 2% goat serum and 0.2% Triton X-100. Cells were then stained overnight at 4°C with MF20 (1:200, Developmental Studies Hybridoma Bank) primary antibody and then for 1 hour at room temperature with anti-mouse AlexaFluor 488 (Molecular Probes) secondary antibody. Cover slips were mounted with ProLong Gold antifade (Molecular Probes).

[00117] ***Exome sequencing and analysis.*** The exomes of four clonally derived DMD myoblast lines carrying known TALEN-mediated deletions in exon 51 of the dystrophin gene, as well as the parent line for these cells, were analyzed. Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (QIAGEN) and 3 μg of DNA were submitted to the Duke Institute for Genome Sciences and Policy's Genome Sequencing & Analysis Core. Illumina-compatible libraries were made and enriched for exonic regions using the SureSelect Human All Exon V4 Kit (Agilent). Five total libraries were prepared from the four treatment samples and one parental line reference sample. The libraries were indexed and sequenced on one lane of Illumina HiSeq2000 (100-bp paired-end sequencing). Bioinformatics analyses were performed by Duke Genome Sequencing & Analysis Core. The analysis pipeline includes the initial QC to remove sequencing adaptors and low quality bases to facilitate mapping. Sequence depth of targeted regions was calculated as >97% at 10x coverage, >91% at 20x coverage, and >82% for 30x coverage (Table 1). Table 1 shows the exome capture statistics. DOWT was the parent DMD myoblast cell line used as the reference sample for analysis. DO32, DO106, DO127, and DO141 were the four clonally derived DMD myoblast lines carrying predetermined on-target NHEJ events at the exon 51 dystrophin locus.

Table 1 Exome capture statistics

Sample Name	DO106	DO127	DO141	DO32	DOWT	Agilent-Human All Exon V4
capture efficiency						
Reads on Target	79.33	79.28	79.27	75.95	79.35	75
Reads On-Target +/-100bp	86.84	88.96	89.15	86.11	89.15	85
Coverage						
1x	99.87	99.88	99.88	99.87	99.88	99
10x	97.4	97.71	97.48	97.53	97.46	90
20x	91.26	92.28	91.41	91.62	91.41	80
30x	82.66	84.45	82.8	83.27	82.94	
50x	63.51	66.35	63.54	64.35	63.9	
100x	28.13	31.12	27.85	28.45	28.62	

[00118] Each sequencing reaction generated >64 million reads with >93% of reads above a quality score of 30 and an overall mean quality score of >36.4. High quality reads were mapped to the human reference genome (hg19) using bwa 0.5.9. An exome capture pipeline developed at the Duke Sequencing Core was used to assess the exome capture efficiency. Picard v1.74 was used for removing PCR duplicates. The GATK (v1.6-13) toolkit was used for variant calling, read realignment around INDELs, quality score recalibration and QC filtering. The filtering step discards the variants with 1) low coverage (coverage (<30x), 2) strand-bias, 3) low SNP quality score (< 50) and 4) low allelic frequency (<0.5). Each candidate point mutation or INDEL were reviewed manually by IGV to identify false negative artifacts due to insufficient coverage of the parental line. Identical point mutations and INDELs that occurred in more than two of the four clones were verified as artifacts due to coverage of the reference parent cell line and were discarded. Common point mutations and INDELs were removed by comparing to human dbSNP135. The remaining point mutations and INDELs were annotated using Annovar and classified using a perl script written by the Duke Sequencing Core. The non-exonic point mutations were not considered. All point mutations and INDELs were individually visualized and validated on IGV. The flanking 100bp of each validated mutation was screened for any potential sequence similarity to the TN3/8 target site using the Paired Target Finder tool on the TALE-NT 2.0 webserver using the parameters: recommended score cutoff (3.0), spacers of range 1-30bp, and upstream base set to "T only."

Example 2

Results

[00119] **Design and validation of TALENs targeted to the dystrophin gene.** To evaluate TALEN-mediated genetic correction by NHEJ, several TALENs were designed to target exon 51 in the dystrophin gene. Plasmids encoding the exemplary TALENs described herein were electroporated into conditionally immortalized myoblasts (muscle cells) isolated from DMD patients. TALEN target sites were chosen immediately upstream of the two possible out-of-frame stop codons (Figure 3a), such that insertions or deletions could restore the dystrophin reading frame in either disrupted frame. Variable lengths of spacers between TALEN monomers and TALEN RVD array lengths were tested to optimize nuclease activity (Figure 3b, Table 2), as done previously (Miller, J.C. *et al. Nat Biotechnol* 29, 143-148 (2011)).

Table 2 Target sequences of engineered TALENs (TN) and repeat variable diresidues (RVDs) that determine DNA-binding specificity

	Target sequences	RVD sequence
TN1	attttagctcctact (SEQ ID NO:26)	NI NG NG NG NG NI NN HD NG HD HD NG NI HD NG
TN2	tttagctcctactcaga (SEQ ID NO:27)	NG NG NG NI NN HD NG HD HD NG NI HD NG HD NI NN NI
TN3	agctcctactcagact (SEQ ID NO:28)	NI NN HD NG HD HD NG NI HD NG HD NI NN NI HD NG
TN4	cctactcagactgtt (SEQ ID NO:29)	HD HD NG NI HD NG HD NI NN NI HD NG NN NG NG
TN5	aaccacaggttgtgtca (SEQ ID NO:30)	NI NI HD HD NI HD NI NN NN NG NG NN NG NN NG HD NI
TN6	agtaaccacaggttgt (SEQ ID NO:31)	NI NN NG NI NI HD HD NI HD NI NN NN NG NG NN NG
TN8	ccttagtaaccacaggt (SEQ ID NO:32)	HD HD NG NG NI NN NG NI NI HD HD NI HD NI NN NN NG
TN9	gcaaaaacccaaaatat (SEQ ID NO:33)	NN HD NI NI NI NI NI HD HD HD NI NI NI NI NG NI NG
TN10	ttgcaaaaacccaaaat (SEQ ID NO:34)	NG NG NN HD NI NI NI NI NI HD HD HD NI NI NI NI NG
TN12	cctttttgcaaaaaccc (SEQ ID NO:35)	HD HD NG NG NG NG NG NN HD NI NI NI NI NI HD HD HD
TN13	caccagagtaacagtct (SEQ ID NO:36)	HD NI HD HD NI NN NI NN NG NI NI HD NI NN NG HD NG
TN17	gtcaccagagtaacagtct (SEQ ID NO:37)	NN NG HD NI HD HD NI NN NI NN NG NI NI HD NI NN NG HD NG
TN26	ccaaactagaaatgccat (SEQ ID NO:38)	HD HD NI NI NI HD NG NI NN NI NI NI NG NN HD HD NI NG
TN27	gaaatctgccagagcaggt (SEQ ID NO:39)	NN NI NI NI NG HD NG NN HD HD NI NN NI NN HD NI NN NN NG
TN28	ctatagatgagattatct (SEQ ID NO:40)	HD NG NI NG NI NN NI NG NN NI NN NI NG NG NI NG HD NG
TN29	agcaatgcggccatccct (SEQ ID NO:41)	NI NN HD NI NI NG NN HD NN NN HD HD NI NG HD HD HD NG

[00120] Western blots confirmed full-length and robust expression of the TALENs following transfection of TALEN-encoding plasmids into HEK293T cells (Figure 4a). All combinations of left and right TALENs were then transfected into HEK293T cells and the genomic DNA was assessed for modification by the Surveyor assay, which can detect the frequency of allelic modifications with a dynamic range of ~1-50%. Several TALENs with spacers of 14-19 bp were highly active with gene editing efficiencies exceeding modification of 10% of total alleles

(Figures 4b-4d). The gene editing frequencies were stable from day 3 to day 10 (Figures 4b and 4e), confirming that these TALENs were well tolerated in human cells. Furthermore, the engineered TALENs showed minimal cytotoxicity in human cells similar to the well-characterized non-cytotoxic homing endonuclease I-SceI (Figures 4f and 5). TN3/8 was highly active and a well-tolerated TALEN pair and was used for subsequent experiments.

[00121] TALEN activity in these cells was assessed by the Surveyor assay, which detects DNA cleavage and religation. This analysis showed that several TALEN pairs were active (Figures 6 and 9a,b). To verify that the targeted changes to exon 51 mediated by the nucleases were leading to restoration of gene expression, dystrophin expression was determined by Western blot. Myoblasts from DMD patients harboring deletions of exons 45 through 50 were electroporated with TALENs (TN3/TN8) that target exon 51 prior to the premature stop codon. Because no homologous donor was provided, DNA repair by NHEJ to lead to small insertions and deletions (indels) was predicted. Because the size of these indels was random, one-third of the modified cells were predicted to contain indels that restored dystrophin expression. After these cells were moved to differentiation conditions, new dystrophin expression that was not present in untreated cells was detected with both an antibody recognizing the rod domain that was N-terminal of the premature truncation (MANDYS-8, Sigma) and the C-terminus (NCL-Dys2, Leica) (Figures 9c and 9d). Dilution of the protein lysates from the differentiated wild-type human myoblasts suggested that the level of correction achieved in the current unoptimized protocol was close to 10% (Figures 9c and 9d). The data demonstrated that TALEN delivery to muscle cells from DMD patients resulted in nuclease activity at the target gene (Figures 6, 9a, and 9b) and dystrophin protein expression in these cells (Figures 9c and 9d) that otherwise was not possible due to DMD mutations.

[00122] In order to further characterize the genotype and phenotype of the corrected cells, clonal cell populations were derived from the nuclease-treated human DMD myoblasts. TALEN-treated cells that showed restored dystrophin protein expression (Figures 9c and 9d) were clonally diluted. The exon 51 PCR product from genomic DNA of 27 clonal populations was diluted 1:1 with the same PCR product from untreated DMD cells and analyzed by the Surveyor assay, which detects mismatches between DNA sequences (Figure 6A). Eleven of the 27 clones showed DNA sequences distinct from the untreated cells at the location of TALEN cleavage, suggesting a gene modification rate of 41% (Figure 6A). This was significantly higher

than what was detected in bulk Surveyor assays (~10-15%, Figure 4b), but was similar to what has been detected with other highly optimized nucleases. The PCR amplicons from these eleven clones were sequenced and 4 of them (36%) showed indels that would be predicted to restore the dystrophin reading frame (Figure 6B).

[00123] These results show correction of dystrophin expression in cells from DMD patients harboring mutations that otherwise would not allow for any dystrophin expression. Although these results show correction by the random generation of indels by NHEJ gene repair in exon 51, these TALENs could be useful for gene correction via other mechanisms, such as exon deletion and gene repair by homologous recombination. Full sequences of all TALENs targeting in and around exon 51 are included in Table 2. Table 2 shows the target sequences and RVDs for TALENs. All target sequences were preceded by a prerequisite 5'T. Figure 2 shows the amino acid sequences of engineered TALENs targeted to exon 51 of the dystrophin gene. Fok ELD-S/KKR-S refers to the *Fok*I nuclease domain that was fused to the TALE DNA-binding domains and the RVD diresidues are underlined. Table 3 shows the target sequences of the TALEN pairs.

Table 3 Target sequences of the TALEN pairs.

TN2/TN6	tttagctcctactcagactgttactctggtgacacaacctgtggttact (SEQ ID NO:16)
TN2/TN8	tttagctcctactcagactgttactctggtgacacaacctgtggttactaagg (SEQ ID NO:17)
TN3/6	agctcctactcagactgttactctggtgacacaacctgtggttact (SEQ ID NO:18)
TN3/TN8	agctcctactcagactgttactctggtgacacaacctgtggttactaagg (SEQ ID NO:19)
TN9/TN17	gcaaaaacccaaaatattttagctcctactcagactgttactctggtgac (SEQ ID NO:20)
TN10/TN17	ttgcaaaaacccaaaatattttagctcctactcagactgttactctggtgac (SEQ ID NO:21)
TN12/TN13	ccttttgcaaaaacccaaaatattttagctcctactcagactgttactctggtg (SEQ ID NO:22)
TN26/TN27	ccaaactagaaatgccatcttccttgatgttggaggtacctgctctggcagatttc (SEQ ID NO:23)
TN28/TN29	ctatagatgagattatctgcccatgactggcgcagggatggccgcattgct (SEQ ID NO:24)

[00124] TN3/8 mediates high efficiency conversion to all three reading frames. NHEJ-based gene modification created indels of random length and may cause conversion to any of the three reading frames in an exonic sequence. To validate the overall gene modification rate and possible reading frames generated following TALEN-induced NHEJ, clonal cell populations were derived from human skeletal myoblasts that had been electroporated with TN3/8-encoding

plasmids. These clones were assayed for NHEJ events occurring at the dystrophin exon 51 locus using the Surveyor assay to detect sequence differences relative to untreated cells (Figure 7a). Eleven of 28 (39%) clonal cell populations were modified and subsequent sequencing of the alleles from these clones confirmed indels characteristic of NHEJ (Figure 7b). Deletions were heavily favored. The random length of these indels verified that conversion to any of the three reading frames was possible. The conversion rate to any one of the three reading frames was observed to be roughly proportional to 1/3 of the total NHEJ events (Figure 7b). Interestingly, several small deletions were observed that did not alter the original reading frame, demonstrating that this approach could be used to delete aberrant stop-codons (Figure 7b).

[00125] Reading frame correction leads to restored protein expression. Whether correction of the dystrophin reading frame by TALEN-mediated NHEJ results in restored dystrophin protein expression was assessed. Immortalized human myoblasts derived from DMD patients with a frame-disrupted dystrophin gene caused by deletion of exons 48-50 (Δ48-50) were electroporated with plasmids encoding TN3/8. Clonal cell populations were isolated and screened by PCR amplification of genomic DNA and Sanger sequencing to identify indels characteristic of NHEJ. Approximately 5% of clones contained modifications in exon 51, including one clone with an NHEJ event that may correct the dystrophin reading frame (Figures 7c and 8). Following myogenic differentiation, restored dystrophin protein expression was detected by western blot at its predicted size (~412 kDa) only in the corrected clone, and not in clones with non-corrective NHEJ events (Figure 7d). These data demonstrate that NHEJ events that restore the dystrophin reading frame also rescue dystrophin protein expression.

[00126] TALEN-mediated genetic correction in bulk-treated DMD myoblasts. Efficient *in situ* frame correction in the absence of selection is a powerful use of NHEJ-based gene correction. Accordingly, the restoration of dystrophin expression in TALEN-treated bulk populations of DMD myoblast lines derived from two different patients containing different deletions of exons 48-50 in the dystrophin gene was investigated. The frequency of gene modification increased with the dose of electroporated TN3/8-encoding plasmids with indels detected in up to 12.7% and 6.8% of alleles, in the two patient lines as measured by the Surveyor assay (Figures 9a and 9b). Following 7 days of myogenic differentiation induced by serum removal, restored dystrophin expression was detected in the bulk cell populations at the predicted size (~412 kDa) relative to expression from wild-type cells (427 kDa) (Figures 9c and 9d). The

increase in dystrophin protein expression with TALEN dose was concomitant with the level NHEJ events detected by the Surveyor assay.

[00127] Gene restoration in primary DMD dermal fibroblasts. The simplicity of this NHEJ-based approach may enable efficient correction in proliferation-limited primary cell lines that may not be amenable to homologous recombination or selection-based gene correction. For example, DMD patient-derived primary dermal fibroblasts carrying a frame-disrupting deletion of exons 46-50 (Δ46-50) were electroporated with plasmids encoding TN3/8, resulting in high frequency gene modification in a dose-dependent manner (Figure 10a). These treated fibroblasts were then transduced with a lentivirus expressing MyoD under an inducible promoter to stimulate transdifferentiation into the myogenic lineage and dystrophin expression. Expression of myogenin (Figure 10b) and myosin heavy chain (Figure 10c) confirmed efficient transdifferentiation of wild type and DMD patient fibroblasts. Rescued dystrophin expression was detected in TALEN-treated MyoD-induced fibroblasts in a dose-dependent manner at the predicted size of approximately 400 kDa (Figure 10b), similar to the results obtained in skeletal myoblasts (Figures 9c and 9d).

[00128] Analysis of off-target effects induced by TN3/8. A concern for all genome editing strategies was the potential for off-target gene modification events. TN3/8 did not show significant cytotoxicity and was well tolerated by human cells (Figures 4b, 4e, and 4f), suggesting specific gene targeting. Potential off-target sites were assessed *in silico* using the TALE-NT 2.0 Paired Target Finder Prediction webserver to scan the human genome for sequences containing up to 4 mismatches per TALEN half-site (up to 8 total mismatches per target site) separated by spacers of any length between 12 and 23 bases. This analysis did not produce any potential off-target sites that met these criteria. To further examine unpredicted off-target DNA modifications, the whole exomes of clonally derived DMD myoblasts that were previously confirmed to contain NHEJ events at the on-target exon 51 locus were sequenced (Figures 6d and 7c). Notably, the only insertion or deletion events characteristic of NHEJ were detected at the on-target exon 51 locus of the dystrophin gene in all four clonal lines analyzed, confirming the specificity of these TALENs (Tables 1 and 4). In Table 4, "syn. SNV" refers to synonymous single nucleotide variants and "nonsyn. SNV" refers to nonsynonymous single nucleotide variants.

Table 4 Summary of clonal sequence variants detected by exome sequencing.

Clone	Mutation Type	Category	AA From	AA To	Gene	Chr	Location	Ref Base	Mutant Base
32	Transition	syn. SNV	S	S	ANKS1B	12	99201691	C	T
	Transversion	nonsyn. SNV	F	L	ZNF836	19	52659835	G	C
	Transition	syn. SNV	P	P	SASH1	6	148664242	T	C
	Transversion	syn. SNV	L	L	DAXX	6	33287597	T	G
	Transversion	syn. SNV	L	L	CDH7	18	63525175	T	A
	Deletion	frameshift	-	-	DMD	X	31792285	ACCAG	-
106	Transition	nonsyn. SNV	E	G	ENG	9	130582267	T	C
	Transition	nonsyn. SNV	N	D	CCDC36	3	49294344	A	G
	Transition	syn. SNV	V	V	TARBP1	1	234556520	C	T
	Transition	nonsyn. SNV	Q	R	UGT3A1	5	35988575	T	C
	Transversion	nonsyn. SNV	L	I	SOWAHB	4	77817679	G	T
	Transversion	nonsyn. SNV	Q	P	MEF2A	15	100252738	A	C
	Transversion	nonsyn. SNV	R	L	RFC1	4	39306505	C	A
	Transition	stopgain SNV	Q	X	ELN	7	73474508	C	T
	Deletion	frameshift	-	-	DMD	X	31792285	ACCAG	-
127	Transition	nonsyn. SNV	A	V	PLEKHH1	14	68041071	C	T
	Transition	syn. SNV	P	P	RASAL2	1	178269222	C	T
	Transversion	nonsyn. SNV	S	C	IGDCC4	15	65676357	G	C
	Transition	syn. SNV	A	A	LMTK3	19	49001482	G	A
	Transition	stopgain SNV	W	X	PLEKHS1	10	115526378	G	A
	Transition	nonsyn. SNV	V	I	FAM110C	2	45848	C	T
	Transition	nonsyn. SNV	G	E	TRAK1	3	42251610	G	A
	Transversion	syn. SNV	S	S	C15orf39	15	75499997	A	T
	Transition	syn. SNV	L	L	GPBAR1	2	219127549	C	T
	Deletion	nonframeshift	-	-	DMD	X	31792284	CAC	-
141	Transition	syn. SNV	L	L	MUC16	19	8999474	T	C
	Transversion	nonsyn. SNV	F	C	RP1	8	55538286	T	G
	Transversion	nonsyn. SNV	E	A	PPP1R10	6	30569808	C	G
	Transition	syn. SNV	T	T	CAMKV	3	49896829	T	C
	Transition	syn. SNV	F	F	AK2	1	33478842	G	A
	Deletion	nonframeshift	-	-	DMD	X	31792287	CAG	-

[00129] Consistent with known genomic mutation rates that normally occur during clonal expansion, the exome sequencing revealed several single nucleotide variants (SNVs) in each clone relative to the parental cell line. Using the TALE-NT 2.0 Paired Target Site Prediction webserver, the immediate region around each mutation was scanned for any sequence similarity to the TN3/8 target site to determine if the TALENs could be responsible for the observed SNVs. No target sites with similarity to our TALEN target site with spacers of 1-30 bases were found in the flanking 100 bp of any SNV. Because NHEJ-mediated mutagenesis rarely results in substitutions relative to indels, the detected SNVs were likely to have arisen during clonal expansion as observed in other studies. In summary, there was no apparent off-target activity related to TALEN-mediated, NHEJ-based genetic correction in these clonally derived cells.

[00130] As shown above, the TALENs had high specificity as demonstrated by *in silico* analysis, cytotoxicity assays, and exome sequencing of clonally-derived modified cells.

Example 3

TALENs Targeting the 5' UTR of the Mouse and Human Dystrophin Gene

[00131] TALENs were designed to target the promoter region of the Dp427m dystrophin isoform, which has been shown to drive high expression of dystrophin in skeletal and cardiac muscle tissues. A panel of TALEN pairs was designed to optimize spacer length and gene editing activity. These TALENs were designed to target a homologous region in both mouse and human dystrophin promoter sequences to facilitate future animal studies, such as preclinical studies in a small animal model without creating a transgenic animal that carries human sequences. Combinations of TALENs were transfected into human 293T cells to assess activity by the Surveyor assay, as described above (see Figure 11C).

[00132] One TALEN pair, TN45/50, was identified to have high activity that was stable at 3 and 10 days post-transfection (see Figure 11A). The RVD components of TN45/50 are shown in Figure 11B. Full amino acid sequences of engineered TALENs targeted to the 5'UTR of the dystrophin gene are provided, as TN45 (Fok-KKRS) (SEQ ID NO: 14) and TN50 (Fok-ELDS) (SEQ ID NO: 15) (see Figure 12). Fok ELDS/KKRS refers to the *Fok*I nuclease domain that was fused to the TALE DNA-binding domains.

[00133] 10 micrograms of each plasmid encoding each of TN45 and TN50 were electroporated into human skeletal myoblasts from DMD patients. TALEN-mediated modification of the target locus was determined by the Surveyor assay three days after. As shown in Figure 11D, the TN45/50 TALENs were active at the chromosomal locus in skeletal myoblasts from human DMD patients.

Example 4

Integration of minidystrophin into 5'UTR of dystrophin

[00134] Functional replacement of mutated dystrophin genes may be accomplished by expression of a miniaturized dystrophin construct, termed minidystrophin ("minidys") (Wang *et al*, *Proc Natl Acad Sci U S A*. (2000) 97(25):13714-9). The dystrophin 5'UTR was targeted with TALENs (Figure 13a) to mediate site-specific integration of minidys that would be under control of the endogenous dystrophin promoter. GFP was targeted to the 5' UTR of the dystrophin gene

by co-transfecting DMD patient myoblasts with the TN45/50 TALEN pair and a donor construct containing the GFP transgene with a selection marker. Results indicated that integration of a gene encoding GFP into the 5'UTR resulted in robust expression from the endogenous promoter in many targeted clones following myogenic differentiation (Figure 13b, d). Similarly, DMD patient myoblast clones, C25C14 and DMD-7796 (provided by Vincent Mouly, Université Pierre et Marie Curie, Paris, France), were generated with TALEN-mediated integration of a construct containing the minidystrophin gene and a selection marker at the 5'UTR. As with the GFP construct, minidystrophin expression was detected only in targeted clones by western blot (Figure 13b, c).

[00135] It is understood that the foregoing detailed description and accompanying examples are merely illustrative and are not to be taken as limitations upon the scope of the invention, which is defined solely by the appended claims and their equivalents.

[00136] Various changes and modifications to the disclosed embodiments will be apparent to those skilled in the art. Such changes and modifications, including without limitation those relating to the chemical structures, substituents, derivatives, intermediates, syntheses, compositions, formulations, or methods of use of the invention, may be made without departing from the spirit and scope thereof.

CLAIMS

What is claimed is:

1. A transcription activator-like effector nuclease (TALEN) protein that binds to a dystrophin gene.

2. The TALEN protein of claim 1, wherein the dystrophin gene comprises a premature stop codon and the TALEN protein binds upstream or downstream from the premature stop codon.

3. The TALEN protein of claim 1 or 2, wherein the TALEN protein binds to a nucleotide sequence comprising one of SEQ ID NOs: 16-46, 50, 52, 58, and 59, and the complement thereof.

4. The TALEN protein of claim 3, wherein the TALEN protein comprises a nuclease.

5. The TALEN protein of claim 4, wherein the nuclease comprises *Fok*I.

6. The TALEN protein of claim 1 or 2, wherein the TALEN protein comprises 15-19 RVD modules.

7. The TALEN protein of claim 2 or 3, wherein the TALEN protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, and SEQ ID NO: 15.

8. The TALEN protein of claim 1 or 2, wherein the TALEN protein binds to a region in exon 51 of the dystrophin gene.

9. The TALEN protein of claim 8, wherein the TALEN protein binds to a nucleotide sequence comprising one of SEQ ID NOs: 16-24, 26-41, 46, 50, 52, 58, and 59, and the complement thereof.

10. The TALEN protein of claim 9, wherein the TALEN protein comprises a nuclease.

11. The TALEN protein of claim 10, wherein the nuclease comprises *Fok*I.

12. The TALEN protein of claim 8, wherein the TALEN protein comprises 15-19 RVD modules.

13. The TALEN protein of claim 8, wherein the TALEN protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13.

14. The TALEN protein of claim 8, wherein the TALEN protein binds to a nucleotide sequence comprising SEQ ID NO: 19, or the complement thereof.

15. The TALEN protein of claim 14, wherein the TALEN protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 2 and 4.

16. The TALEN protein of claim 1 or 2, wherein the TALEN protein binds to a region in the 5' UTR of the dystrophin gene.

17. The TALEN protein of claim 16, wherein the TALEN protein binds to a nucleotide sequence comprising one of SEQ ID NOs: 25 and 42-45 and the complement thereof.

18. The TALEN protein of claim 17, wherein the TALEN protein comprises a nuclease.

19. The TALEN protein of claim 18, wherein the nuclease comprises *Fok*I.

20. The TALEN protein of claim 16, wherein the TALEN protein comprises 15-19 RVD modules.

21. The TALEN protein of claim 16, wherein the TALEN protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 14 and 15.

22. The TALEN protein of any one of the preceding claims, wherein the dystrophin gene is a human dystrophin gene.

23. An isolated polynucleotide comprising a nucleotide sequence encoding the TALEN protein of any one of claims 1-22.

24. A vector comprising the isolated polynucleotide of claim 23.

25. A cell comprising the isolated polynucleotide of claim 23 or the vector of claim 24.

26. A composition comprising two or more TALENs according to any one of claims 1-22, wherein a first TALEN binds to a first binding region and a second TALEN binds to a second binding region, wherein the first binding region and second binding region are located within a target region and the first binding region and second binding region are not the same.

27. The composition of claim 26, wherein the first binding region and the second binding region are separated by at least one nucleotide.

28. The composition of claim 26, wherein the first binding region and the second binding region are separated by 5-25 base pairs.

29. A method of treating a subject in need thereof having a mutant dystrophin gene, the method comprising administering to the subject the TALEN protein of any one of claims 1-22, the isolated polynucleotide of claim 23, the vector of claim 24, the cell of claim 25 or the composition of any one of claims 26-28.

30. The method of claim 29, wherein the subject is suffering from Duchenne muscular dystrophy.

31. A method of correcting a mutant dystrophin gene in a cell, the method comprising administering to a cell containing a mutant dystrophin gene the TALEN protein of any one of claims 1-22, the isolated polynucleotide of claim 23, the vector of claim 24, or the composition of any one of claims 26-28.

32. The method of claim 31, wherein the correction of the mutant dystrophin gene comprises homology-directed repair.

33. The method of claim 32, further comprising administering to the cell a donor DNA.

34. The method of claim 31, wherein the mutant dystrophin gene comprises a frameshift mutation which causes a premature stop codon and a truncated gene product.

35. The method of claim 34, wherein the correction of the mutant dystrophin gene comprises nuclease mediated non-homologous end joining.

36. A method of correcting a mutant gene in a cell, the method comprising administering to a cell containing a mutant gene a first TALEN and a second TALEN, wherein the first TALEN binds to a first binding region and a second TALEN binds to a second binding region, wherein the first binding region and second binding region are located within a target region and the first binding region and second binding region are not the same, wherein the correction of the mutant gene comprises nuclease mediated non-homologous end joining, and wherein the correction restores the mutant gene.

37. A method of correcting a mutant dystrophin gene in a cell, the method comprising administering to a cell containing a mutant dystrophin gene a first TALEN and a second TALEN, wherein the first TALEN binds to a first binding region and a second TALEN binds to a second binding region, wherein the first binding region and second binding region are located within a target region and the first binding region and second binding region are not the same.

38. The method of claim 36 or 37, wherein the mutant dystrophin gene comprises a premature stop codon and the target region is upstream or downstream of the premature stop codon.

39. The method of claim 37, wherein the correction of the mutant dystrophin gene comprises nuclease mediated non-homologous end joining.

40. The method of claim 36 or 37, wherein the first binding region and the second binding region are separated by at least one nucleotide.

41. The method of claim 36 or 37, wherein the first binding region and the second binding region are separated by 5-25 base pairs.

42. The method of any one of claims 36-41, wherein the method does not involve co-administration of exogenous DNA.

43. The method of claim 37, wherein the correction of the mutant dystrophin gene comprises homology-directed repair.

44. The method of claim 44, further comprising administering to the cell a donor DNA.

45. A kit comprising the TALEN protein of any one of claims 1-22, the isolated polynucleotide of claim 23, the vector of claim 24, the cell of claim 25 or the composition of any one of claims 26-28.

46. A kit for correcting a mutant dystrophin gene, the kit comprising the TALEN protein of any one of claims 1-22, the isolated polynucleotide of claim 23, the vector of claim 24, the cell of claim 25 or the composition of any one of claims 26-28.

a

b

Program	Conditions
A	1000uF, 120V, 1M cells/200uL, 10ug GFP + 10ug empty vector
B	500uF, 120V, 1M cells/200uL, 10ug GFP + 10ug empty vector
C	300uF, 120V, 1M cells/200uL, 10ug GFP + 10ug empty vector
D	500uF, 100V, 1M cells/200uL, 10ug GFP + 10ug empty vector
E	500uF, 150V, 1M cells/200uL, 10ug GFP + 10ug empty vector
F	500uF, 200V, 1M cells/200uL, 10ug GFP + 10ug empty vector
G	1000uF, 100V, 1M cells/200uL, 10ug GFP + 10ug empty vector
H	1000uF, 150V, 1M cells/200uL, 10ug GFP + 10ug empty vector
I	1000uF, 180V, 0.2M cells/200uL, 10ug GFP + 10ug empty vector
J	1000uF, 180V, 0.4M cells/200uL, 10ug GFP + 10ug empty vector
K	1000uF, 180V, 1M cells/200uL, 10ug GFP + 10ug empty vector
L	950uF, 190V, 0.5M cells/150uL, 5ug GFP
M	950uF, 190V, 1M cells/200uL, 10ug GFP
N	950uF, 160V, 0.5M cells/150uL, 5ug GFP
O	950uF, 160V, 1M cells/200uL, 10ug GFP
P	950uF, 220V, 2M cells/300uL, 10ug GFP
Q	950uF, 180V, 1M cells/200uL, 10ug GFP
R	950uF, 190V, 1M cells/200uL, 2ug pmaxGFP
S	950uF, 190V, 1M cells/100uL, 2ug pmaxGFP

Figure 1

TN2 (Fok-ELDS) (SEQ ID NO:1):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QDHGLTPDQVVATASNIGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSE
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQT
RDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN3 (Fok-ELDS) (SEQ ID NO:2):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALESIVAQLSRPD
PALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVM
EFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNY
KAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN6 (Fok-KKRS) (SEQ ID NO:3):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALESIVAQLSRPD
PALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVM
EFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNY
KAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Figure 2

TN8 (Fok-KKRS) (SEQ ID NO:4):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKS
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQT
RNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN9 (Fok-ELDS) (SEQ ID NO:5):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKS
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQT
RDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN10 (Fok-ELDS) (SEQ ID NO:6):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKS
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQT
RDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Figure 2 (continued)

TN12 (Fok-ELDS) (SEQ ID NO:7):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKS
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQT
RDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN13 (Fok-KKRS) (SEQ ID NO:8):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKS
ELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQT
RNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN17 (Fok-KKRS) (SEQ ID NO:9):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALESIVAQLS
RPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELLEIARNPTQDRILEM
KVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFK
GNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Figure 2 (continued)

TN26 (Fok-ELDS) (SEQ ID NO:10):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQLS
RPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEM
KVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFK
GNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN27 (Fok-KKRS) (SEQ ID NO:11):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALESIVAQLS
RPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEM
KVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFK
GNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN28 (Fok-ELDS) (SEQ ID NO:12):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQLS
RPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPI
DYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIK
AGTLTLEEVRRKFNNGEINF

Figure 2 (continued)

TN29 (Fok-KKRS)(SEQ ID NO:13):

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLT
VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH
PDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
GLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLP
HAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPI
DYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNRKTNCNGAVLSVEELLIGGEMIK
AGTLTLEEVRRKFNNGEINF

Figure 2 (continued)

a

		Reading frame	SEQ ID NO:
Intron 50	Exon 51		
TATTTTAGCTCCTACTCAGACTGTTACTCTGGTGACACAACCTGTGGTTACTAAG			46
L L L R L L L L W STOP			
S Y S D C Y S G D T T C G Y STOP		+2 (out-of-frame)	47
P T Q T V T L L V T T Q P P V V T T K		+1 (out-of-frame)	48
		+3 (in-frame)	49

Nuclease cleaves DNA and induces DNA damage response, gene editing during repair restores reading frame

	SEQ ID NO:
TATTTTAGCTCCTACTCAGACTGTTA (N) CTCTGGTGACACAACCTGTGGTTACTAAG	50
L L L R L L L X L V T Q P P V V T T K Corrected +2	51
	SEQ ID NO:
TATTTTAGCTCCTACTCAGACTGTTA (NN) CTCTGGTGACACAACCTGTGGTTACTAAG	52
S Y S D C X X L V T Q P V V T T K Corrected +1	53

b

TALEN	Left target	Right target	SEQ ID NO:
TN1	5′-ATTTTAGCTCCTACT		54
TN2	5′-TTTAGCTCCTACTCAGA		55
TN3	5′-AGCTCCTACTCAGACT		56
TN4	5′-CCTACTCAGACTGTT		57
	5′-AAAATATTTTAGCTCCTACTCAGACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGA		58
	TTTTATAAAATCGAGGATGAGTCTGACAATGAGACCACTGTGTTGGACACCAATGATTCCT-5′		59
TN5		ACTGTTGGACACCAA-5′	30
TN6		TGTTGGACACCAATGA-5′	31
TN8		TGGACACCAATGATTCC-5′	32

c

Figure 3

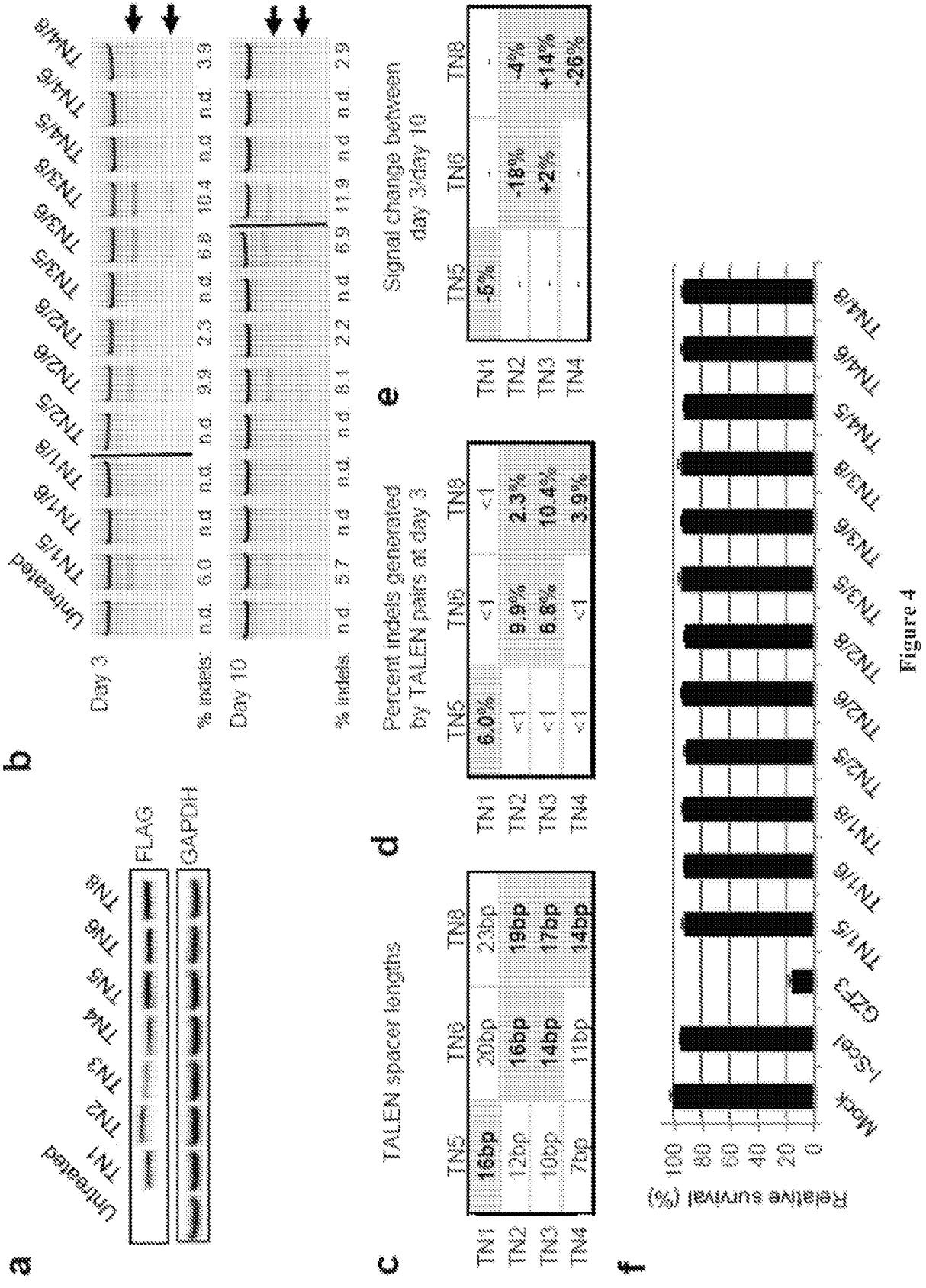

a
Untreated
TN1
TN2
TN3
TN4
TN5
TN6
TN8
FLAG
GAPDH
b
Day 3
Day 10
Untreated
TN1/5
TN1/6
TN1/8
TN2/5
TN2/6
TN2/8
TN3/5
TN3/6
TN3/8
TN4/5
TN4/6
TN4/8
% indels: n.d. 6.0 n.d. n.d. n.d. 9.9 2.3 n.d. 6.8 10.4 n.d. n.d. 3.9
% indels: n.d. 5.7 n.d. n.d. n.d. 8.1 2.2 n.d. 6.9 11.9 n.d. n.d. 2.9
c
TALEN spacer lengths
TN5 TN6 TN8
TN1 16bp 20bp 23bp
TN2 12bp 16bp 19bp
TN3 10bp 14bp 17bp
TN4 7bp 11bp 14bp
d
Percent indels generated by TALEN pairs at day 3
TN5 TN6 TN8
TN1 6.0% <1 <1
TN2 <1 9.9% 2.3%
TN3 <1 6.8% 10.4%
TN4 <1 <1 3.9%
e
Signal change between day 3/day 10
TN5 TN6 TN8
TN1 -5% - -
TN2 - -18% -4%
TN3 - +2% +14%
TN4 - - -26%
f
Relative survival (%)
0 20 40 60 80 100
Mock
I-SceI
GZF3
TN1/5
TN1/6
TN1/8
TN2/5
TN2/6
TN2/8
TN3/5
TN3/6
TN3/8
TN4/5
TN4/6
TN4/8

Figure 4

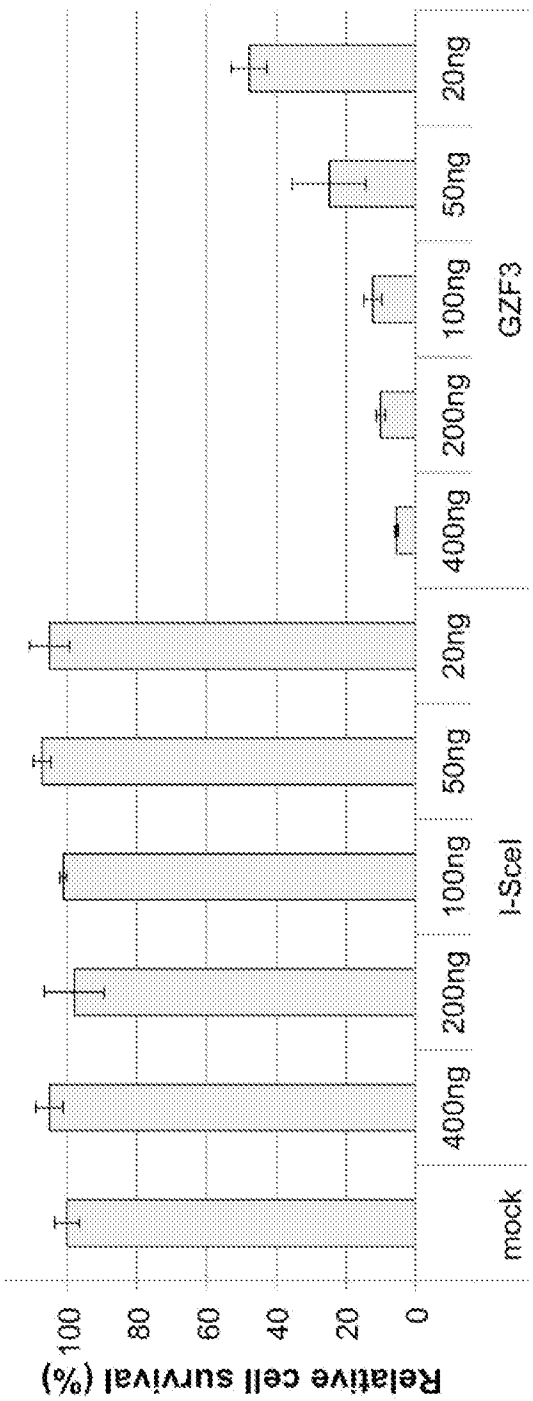

Relative cell survival (%)
0
20
40
60
80
100
mock
400ng
200ng
100ng
50ng
20ng
I-SceI
400ng
200ng
100ng
50ng
20ng
GZF3

Figure 5

A
Clones
Neg control
B
SEQ ID NO:
WT AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACTCT-GGTGACACAACCTGTGG 63
3 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACT----GTGACACAACCTGTGG -3 64
4 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGT----------GG -21 65
12 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACTCT------TGACACAACCTGTGG -2 66
13 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTAC----------ACAACCTGTGG -9 67
16 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACT---GGTGACACAACCTGTGG -2 68
17 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACTC*%GGTGACACAACCTGTGG +183 69
18 AAAACCCCAAAATTTTTTAGCTCCTTCTCAGACTTTTT--TCT-GGTGACACAACCTGTGG -2 70
20 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACTCTTTGGTGACACAACCTGTGG +1 71
23 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACTCT----------GCAACCTGTGG -6 72
46 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTAC-------GGTGACACAACCTGTGG -3 73
47 AAAAACCCAAAATATTTTAGCTCCTACTCAGACTGTTACTCT-GGTGACACAACCTGTGG 0 74
Frame restored: 4/11 (~36.3%)

Figure 6

a
NHEJ events: 1128 = 39.3%
Clone:
Clone:
b
Clone
Exon 51 sequence
Frameshift
SEQ ID NO:
Deletions:
WT
AGACTGTTACTCTGGTGACACAACCTGTGGTTAC
83
1
AGACTGTTACT---GTGACACAACCTGTGGTTAC
-3
84
4
AGACTGT----------GGTTAC
-21
85
11
AGACTGTTACTCT--TGACACAACCTGTGGTTAC
-2
86
13
AGACTGTTAC--------ACAACCTGTGGTTAC
-8
87
14
AGACTGTTACT--GGTGACACAACCTGTGGTTAC
-2
88
15
AGACTGTTACTC*GTGACACAACCTGTGGTTAC
+163
89
16
AGACTGTTACT--GGTGACACAACCTGTGGTTAC
-2
90
18
AGACTGTTACTCTG------CAACCTGTGGTTAC
-6
91
26
AGACTGTTAC---GGTGACACAACCTGTGGTTAC
-3
92
Insertions:
WT
AGACTGTTACTCT-GGTGACACAACCTGTGGTTAC
93
17
AGACTGTTACTCTTGGTGACACAACCTGTGGTTAC
+1
62
Frame +1: 2/11
Frame +2: 4/11
Frame +3: 5/11
c
NHEJ events (0/160 = -5%)
Clone
Exon 51 sequence
Frameshift
SEQ ID NO:
WT
AGACTGTTACTCTGGTGACACAACCTGTGGTTAC
75
32
AGACTGTT----TGGTGACACAACCTGTGGTTAC
-4
76
40
AGACTGTTACTCTG-TGACACAACCTGTGGTTAC
-1
77
67
AGACTG----------TGACACAACCTGTGGTTAC
-9
78
106
AGACTGTTAC----TGACACAACCTGTGGTTAC
-5
79
137
AGACTGTTACTC---TGACACAACCTGTGGTTAC
-3
80
141
AGACTGTTACT---GTGACACAACCTGTGGTTAC
-3
81
145
AGACTGTTACTC---TGACACAACCTGTGGTTAC
-3
82
d
127
141
32
40
106
Clone
N
N
N
N
Y
Frame Restored
Dystrophin
GAPDH

Figure 7

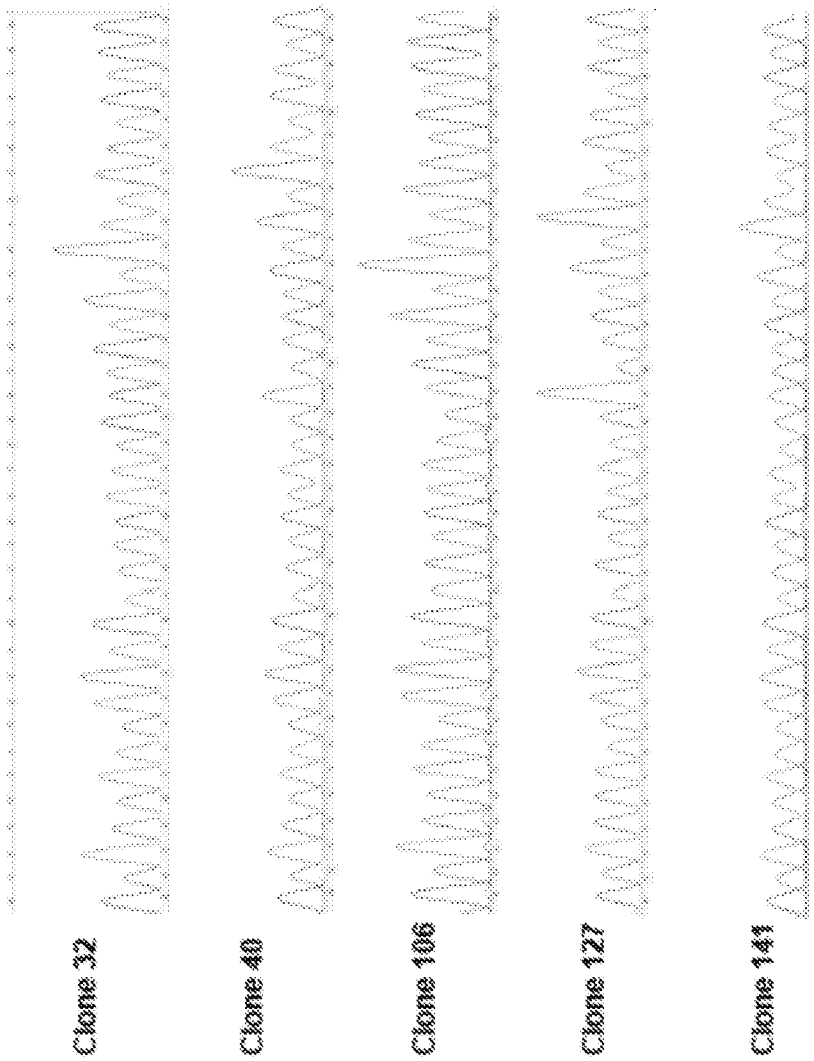


Figure 8

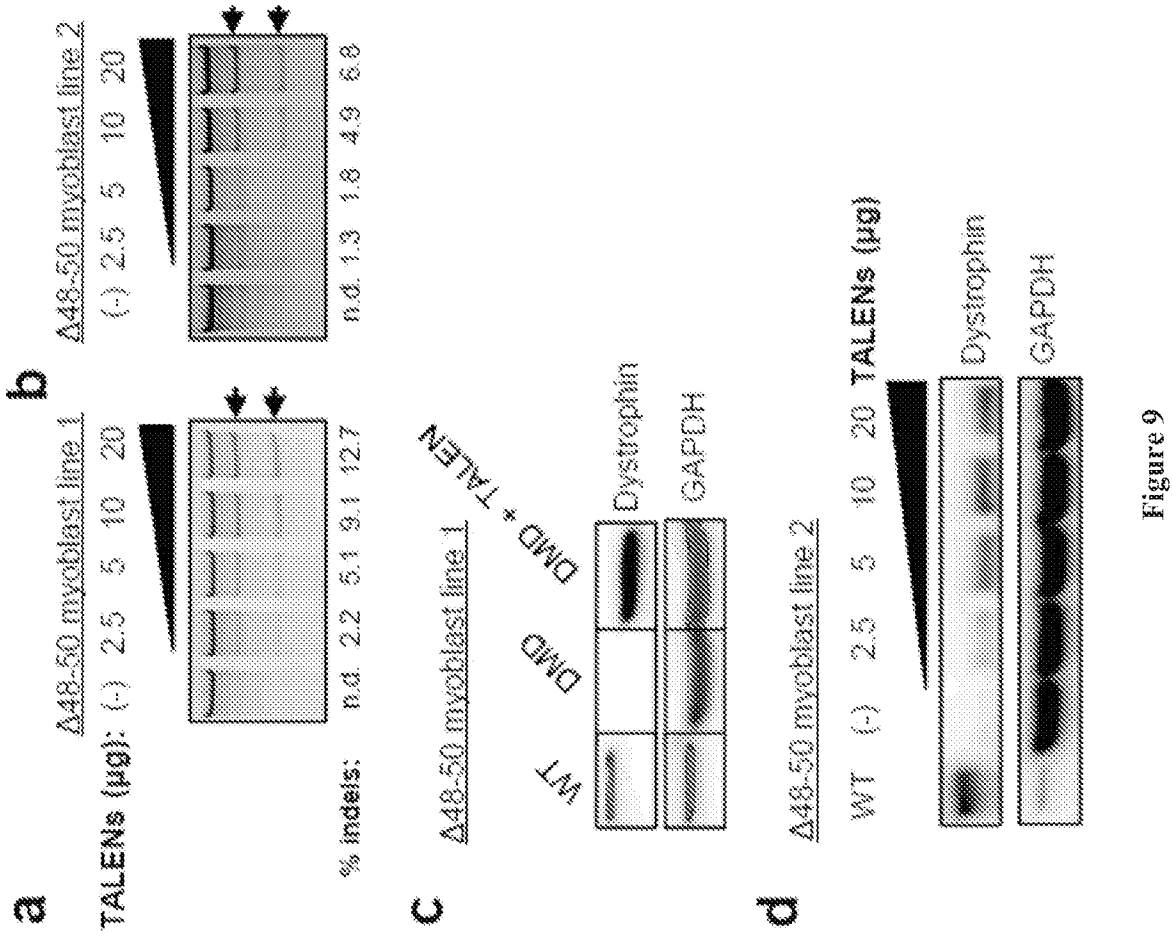

a
Δ48-50 myoblast line 1
TALENs (μg): (-) 2.5 5 10 20
% indels: n.d. 2.2 6.1 9.1 12.7
b
Δ48-50 myoblast line 2
(-) 2.5 5 10 20
n.d. 1.3 1.8 4.9 6.8
c
Δ48-50 myoblast line 1
WT
DMD
DMD + TALEN
Dystrophin
GAPDH
d
Δ48-50 myoblast line 2
WT (-) 2.5 5 10 20 TALENs (μg)
Dystrophin
GAPDH

Figure 9

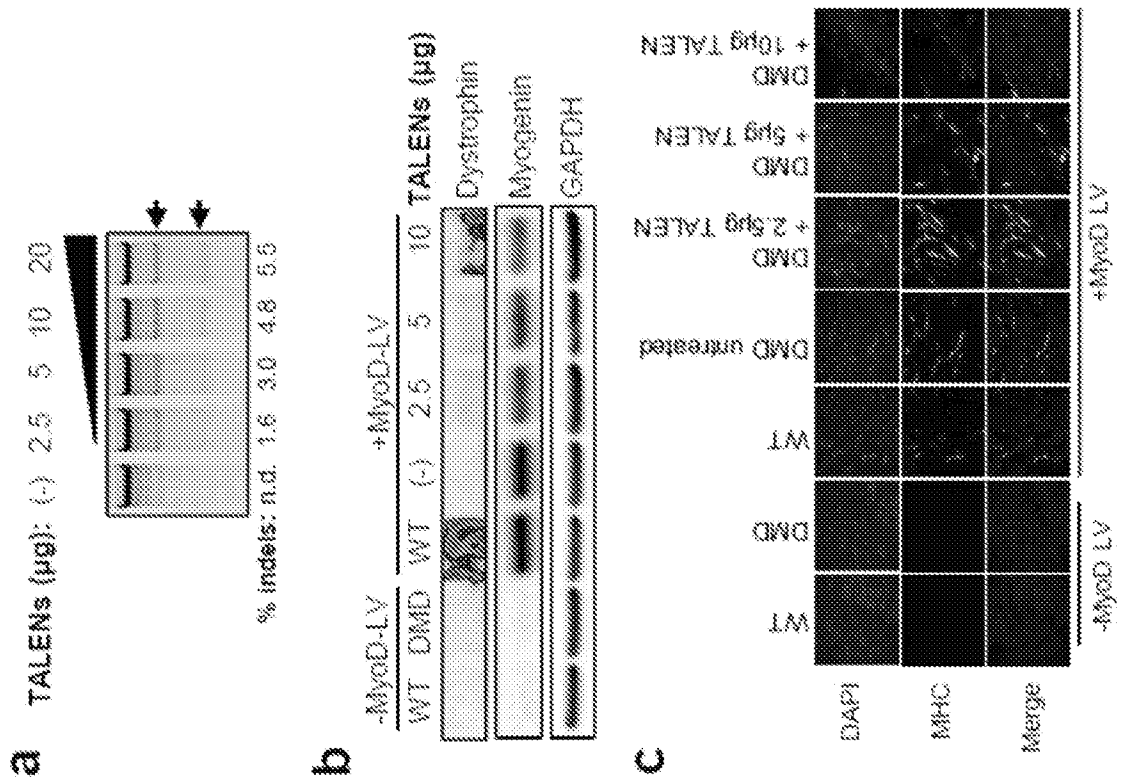

a
TALENs (μg): (-) 2.5 5 10 20
% indels: n.d. 1.6 3.0 4.6 6.6
b
-MyoD-LV
+MyoD-LV
WT DMD WT (-) 2.5 5 10 TALENs (μg)
Dystrophin
Myogenin
GAPDH
c
WT
DMD
WT
DMD untreated
DMD + 2.5μg TALEN
DMD + 5μg TALEN
DMD + 10μg TALEN
-MyoD LV
+MyoD LV
DAPI
MHC
Merge

Figure 10

A

Target site (5'->3')

TN45 (Right TALEN) tAACTGATGCCAGGATT (- strand) SEQ ID NO: 42
TN50 (Left TALEN) tGGAATTTGAAATATCC SEQ ID NO: 43

Full target site tGGAATTTGAAATATCCNNNNNNNNNNNNNNNNAATCCTGGCATCAGTTa SEQ ID NO: 44
Human target site tGGAATTTGAAATATCCGGGGCCTCTACAGAATCCTGGCATCAGTTa SEQ ID NO: 25
Mouse target site tGGAATTTGAAATATCCAGAGGCCTCTACAGAATCCTGGCATCAGTTa SEQ ID NO: 45

B

RVDs

NI NI HD NG NN NI NG NN HD HD NI NN NN NI NG MG HD HD
NN NN NI NI NG NG NG NN NI NI NI NG NI NG HD HD

C

TALEN - +

Figure 11

TN45 (Fok-KKRS) (SEQ ID NO:14):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ
VVATASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDH
PALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELITRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVM
EFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNY
KAQLTRLNRKTNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

TN50 (Fok-ELDS) (SEQ ID NO:15):
MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGRGSVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSG
ARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALESIVAQLSRPD
PALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVAQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNPTQDRILEMKVM
EFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTRDKHLNPNEWWKVYPSSVTEFKFLFVSGHFKGNY
KAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Figure 12

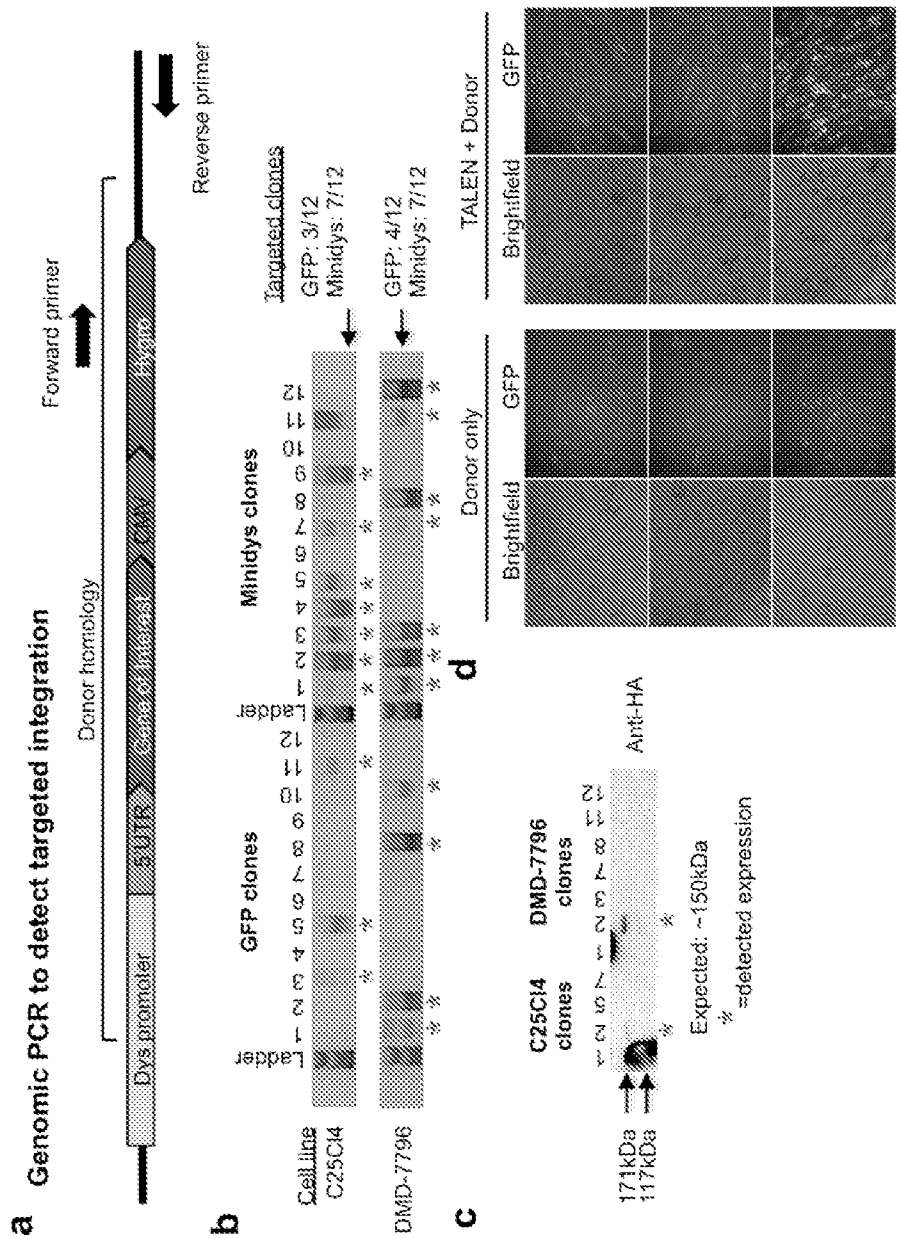

a
Genomic PCR to detect targeted integration
Forward primer
Donor homology
Dys promoter
5'UTR
Reverse primer
b
GFP clones
Minidys clones
Targeted clones
Ladder
Cell line
C25CI4
DMD-7796
GFP: 3/12
Minidys: 7/12
GFP: 4/12
Minidys: 7/12
c
C25CI4 clones
DMD-7796 clones
Anti-HA
171kDa
117kDa
Expected: ~150kDa
*=detected expression
d
Donor only
TALEN + Donor
Brightfield
GFP

Figure 13