Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2902954 A1 2014/09/04

(21) 2 902 954

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2014/02/28
(87) **Date publication PCT/PCT Publication Date:** 2014/09/04
(85) **Entrée phase nationale/National Entry:** 2015/08/27
(86) **N° demande PCT/PCT Application No.:** US 2014/019322
(87) **N° publication PCT/PCT Publication No.:** 2014/134412
(30) **Priorité/Priority:** 2013/03/01 (US61/771,735)

(51) **Cl.Int./Int.Cl.** ***A01H 1/06*** (2006.01)

(71) **Demandeur/Applicant:**
REGENTS OF THE UNIVERSITY OF MINNESOTA, US

(72) **Inventeurs/Inventors:**
OSBORN, MARK J., US;
TOLAR, JAKUB, US;
BLAZAR, BRUCE, US;
VOYTAS, DANIEL F., US

(74) **Agent:** MARKS & CLERK

(54) **Titre : CORRECTION DE GENE A BASE DE TALEN**
(54) **Title: TALEN-BASED GENE CORRECTION**

(57) **Abrégé/Abstract:**
The invention is directed to transcription activator-like effector nuclease (TALEN)-mediated DNA editing of disease-causing mutations in the context of the human genome and human cells to treat patients with compromised genetic disorders. Epidermolysis bullosa (EB) is a group of genetic conditions that cause the skin to be very fragile and to blister easily. Blisters and skin erosions form in response to minor injury or friction, such as rubbing or scratching. Recessive dystrophic epidermolysis bullosa (RDEB), the most severe and classical form of the disease, is characterized by extensive blistering and scarring of the skin and mucosal membranes.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 September 2014 (04.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/134412 A1

(51) International Patent Classification:
*A01H 1/06* (2006.01)

(21) International Application Number:
PCT/US2014/019322

(22) International Filing Date:
28 February 2014 (28.02.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/771,735 1 March 2013 (01.03.2013) US

(71) Applicant: **REGENTS OF THE UNIVERSITY OF MINNESOTA** [US/US]; 600 Mcnamara Alumni Center, 200 Oak Street Se, Minneapolis, Minnesota 55455-2020 (US).

(72) Inventors: **OSBORN, Mark, J.**; St. Paul, Minnesota (US). **TOLAR, Jakub**; Minneapolis, Minnesota (US). **BLAZAR, Bruce**; Golden Valley, Minnesota (US). **VOYTAS, Daniel, F.**; St. Paul, Minnesota (US).

(74) Agent: **MURPHY, Joseph, F.**; CAESAR, RIVISE, BERNSTEIN, COHEN & POKOTILOW, LTD., Seven Penn Center, 1635 Market Street, 12th Floor, Philadelphia, Pennsylvania 19103 (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, IR, IS, JP, KE, KG, 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, SA, 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, KM, ML, MR, NE, SN, TD, TG).

**Published**:

— *with international search report (Art. 21(3))*

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

WO 2014/134412 A1

(54) Title: TALEN-BASED GENE CORRECTION

(57) Abstract: The invention is directed to transcription activator-like effector nuclease (TALEN)-mediated DNA editing of disease-causing mutations in the context of the human genome and human cells to treat patients with compromised genetic disorders. Epidermolysis bullosa (EB) is a group of genetic conditions that cause the skin to be very fragile and to blister easily. Blisters and skin erosions form in response to minor injury or friction, such as rubbing or scratching. Recessive dystrophic epidermolysis bullosa (RDEB), the most severe and classical form of the disease, is characterized by extensive blistering and scarring of the skin and mucosal membranes.

# TALEN-BASED GENE CORRECTION

## SPECIFICATION

## CROSS-REFERENCE TO RELATED APPLICATIONS

This international application claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Patent Application No. 61/771,735, filed March 1, 2013, the entirety of which is incorporated herein.

### Sequence Listing

The instant application contains a sequence listing which has been submitted in ascii format and is hereby incorporated by reference in its entirety. Said ascii copy, created on February 27, 2014, is named J110020003_st25.txt and is 74,494 byte in size.

### Background of the Invention

Epidermolysis bullosa (EB) is a group of genetic conditions that cause the skin to be very fragile and to blister easily. Blisters and skin erosions form in response to minor injury or friction, such as rubbing or scratching. Recessive dystrophic epidermolysis bullosa (RDEB), the most severe and classical form of the disease, is characterized by extensive blistering and scarring of the skin and mucosal membranes. The COL7A1 mutations associated with RDEB impair the ability of collagen 7 to connect the epidermis and dermis; and subsequent separation of the epidermis and dermis as a result of friction or minor injury causes the severe blistering and extensive scarring of the skin associated with RDEB. People with RDEB exhibit incurable, often fatal skin blistering and are at increased risk for aggressive squamous cell carcinoma1. Gene augmentation therapies are promising, but run the risk of insertional mutagenesis. Current gene therapy tools (e.g., viral-mediated gene-addition) rely on the provision of functional copies of a therapeutic gene that integrate at random or semi-random into the genome. The consequences of the random integration are perturbation of the locus where the cargo lands and potential gene inactivation or dysregulation (off target effects). These can result in life threatening side effects to the patient. It is therefore described herein engineered transcription activator like effector nucleases (TALENs) for precision genome-editing in cells of patients with, for example, RDEB, and other genetic disorders.

All references cited herein are incorporated by reference in their entireties.

## Summary of the Invention

The present invention overcomes the off target effects by providing site specific correction of the mutation. The correction of the mutation may be accomplished by transformation or transfection of a cell. The cell may be selected from the group consisting of a fibroblast, keratinocyte, inducible pluripotent stem cell, hematopoietic stem cell, mesenchymal stem cell, embryonic stem cell, hematopoietic progeny cell, T-cell, B-cell, glial cell, neural cell, neuroglial progenitor cell, neuroglial stem cell, muscle cell, lung cell, pancreatic cell, liver cell and a cell of the reticular endothelial system

One embodiment provides a method to treat a genetic disease or disorder caused by a genetic mutation comprising contacting a cell with one or more nucleic acids encoding a TALEN and a nucleic acid donor sequence, wherein TALEN protein is expressed in the cell and induces a site-specific double stranded DNA break in a target gene, wherein the donor sequence is a template for DNA repair resulting in a correction of the genetic mutation and provides correct gene expression, so as to treat the genetic disease or disorder. In one embodiment, the cell is a fibroblast, keratinocyte, inducible pluripotent-, hematopoietic-, mesenchymal-, or embryonic stem cell, hematopoietic progeny cell (such as a T-cell or B-cell), glia and neural cell, neuroglial progenitor and stem cell, muscle cell, lung cell, pancreatic and/or liver cell and/or a cell of the reticular endothelial system. The invention further provides for the use of one or more nucleic acids to treat a genetic disease or disorder caused by a genetic mutation, where said one or more nucleic acids encode a transcription activator like effector nuclease (TALEN) and a nucleic acid donor sequence, wherein when TALEN protein is expressed in a cell and induces a site-specific double stranded DNA break in a target gene, and wherein the donor sequence is a template for DNA repair, results in a correction of the genetic mutation and provides correct gene expression, so as to treat the genetic disease or disorder.

In the one embodiment, the TALEN is a left TALEN and further comprising a right TALEN that cooperates with the left TALEN to make the double strand break in the target gene. In another embodiment, the nucleic acid encoding the TALEN and/or the nucleic acid donor sequence is part of a vector or plasmid. In one embodiment, the TALEN includes a spacer (e.g., the spacer sequence is 12 to 30 nucleotides in length).

In one embodiment, the target gene is a gene with a genetic alteration/mutation. For example, in one embodiment, the target gene is COL7A1 (one with a mutation causing, for example, aberrant expression of the protein).

In one embodiment, the genetic disease is epidermolysis bullosa, osteogenesis imperfecta, dyskeratosis congenital, the mucopolysaccharidoses, muscular dystrophy, cystic fibrosis (CFTR), fanconi anemia, the sphingolipidoses, the lipofuscinoses, adrenoleukodystrophy, severe combined immunodeficiency, sickle-cell anemia or thalassemia.

One embodiment provides a method to treat a genetic disease or disorder caused by a genetic mutation comprising a) introducing into a cell (i) a first nucleic acid encoding a first transcription activator-like (TAL) effector endonuclease monomer, (ii) a second nucleic acid encoding a second TAL effector endonuclease monomer, and (iii) and a donor sequence, wherein each of said first and second TAL effector endonuclease monomers comprises a plurality of TAL effector repeat sequences and a FokI endonuclease domain, wherein each of said plurality of TAL effector repeat sequences comprises a repeat-variable diresidue, wherein said first TAL effector endonuclease monomer comprises the ability to bind to a first half-site sequence of a target DNA within said cell and comprises the ability to cleave said target DNA when said second TAL effector endonuclease monomer is bound to a second half-site sequence of said target DNA, wherein said target DNA comprises said first half-site sequence and said second half-site sequence separated by a spacer sequence, and wherein said first and second half-sites have the same nucleotide sequence or different nucleotide sequences, wherein said donor sequence comprises homology to the target at least at the 5' and 3's ends of the target sequence and the preselected genetic alteration and is a template for DNA repair resulting in a correction of the genetic mutation; and (b) culturing the cell under conditions in which the first and second TAL effector endonuclease monomers are expressed, so as to correct the mutation and restores correct gene expression. Each of the first and second nucleic acids may comprise a spacer (distinct from the spacer sequence). The spacer sequence may be located between the plurality of TAL effector repeat sequences and the FokI endonuclease domain. The spacer sequence may be 12 to 30 nucleotides. In a further embodiment, the invention provides for the use of one or more nucleic acids to treat a genetic disease or disorder caused by a genetic

mutation, wherein (i) a first nucleic acid encodes a first transcription activator-like (TAL) effector endonuclease monomer, (ii) a second nucleic acid encodes a second TAL effector endonuclease monomer, and (iii) and a donor sequence, wherein each of said first and second TAL effector endonuclease monomers comprises a plurality of TAL effector repeat sequences and a FokI endonuclease domain, wherein each of said plurality of TAL effector repeat sequences comprises a repeat-variable diresidue, wherein said first TAL effector endonuclease monomer comprises the ability to bind to a first half-site sequence of a target DNA within said cell and comprises the ability to cleave said target DNA when said second TAL effector endonuclease monomer is bound to a second half-site sequence of said target DNA, wherein said target DNA comprises said first half-site sequence and said second half-site sequence separated by a spacer sequence, and wherein said first and second half-sites have the same nucleotide sequence or different nucleotide sequences, wherein said donor sequence comprises homology to the target at least at the 5' and 3's ends of the target sequence and the preselected genetic alteration and is a template for DNA repair resulting in a correction of the genetic mutation; and wherein (b) culturing the cell under conditions in which the first and second TAL effector endonuclease monomers are expressed, so as to correct the mutation and restore correct gene expression.

Another embodiment provides a nucleic acid comprising a donor sequence, wherein the donor sequence is a template for site specific DNA repair resulting in a correction of a genetic mutation, wherein the donor sequence comprises homology to at least the 5' and 3' ends of the target sequence, wherein a portion of the donor sequence comprises a repair sequence to correct the target sequence for use in conjunction with a TALEN protein. In one embodiment, the donor comprises SEQ ID NO:22. In another embodiment, the target is COL7A1 (a gene with a mutation). In one embodiment, the 5' and 3' ends of the donor each have at least 100 bases of sequence identity to the target.

In another embodiment, the nucleic acid comprises SEQ ID NO:29 or 30. One embodiment provides the proteins coded for or expressed by the TALEN nucleic acids.

One embodiment provides a vector or plasmid comprising a donor sequence, wherein the donor sequence is a template for site specific DNA repair resulting in a

correction of a genetic mutation, wherein the donor sequence comprises homology to at least the 5' and 3' ends of the target sequence, wherein a portion of the donor sequence comprises a repair sequence to correct the target sequence for use in conjunction with a TALEN protein. In one embodiment, the donor comprises SEQ ID NO:22. In one embodiment, the target is COL7A1 (with a mutation). In one embodiment, the 5' and 3' ends of the donor each have at least 100 bases of sequence identity to the target. One embodiment provides a vector or plasmid comprising one or more of SEQ ID NOs: 22, 31, 28, 29 or 30. Another embodiment provides an isolated host cell comprising one or more of exogenous SEQ ID NOs: 22, 31, 28, 29 or 30 or the proteins expressed from such sequences. Another embodiment provides a transfected cell line comprising SEQ ID NOs: 22, 31, 28, 29 or 30 or the proteins expressed from such sequences.

One embodiment provides a method to treat a genetic disease or disorder caused by a genetic mutation comprising contacting a cell with a nucleic acid encoding a TALEN, wherein the TALEN corrects the mutation and for example, restores correct gene expression, or enhances gene expression. In one embodiment, the cell is a fibroblast. In another embodiment, the TALEN is a left TALEN and further comprising a right TALEN that cooperates with the left TALEN to make a double strand cut in a DNA. In one embodiment, the nucleic acid molecule is a vector. In another embodiment, the nucleic acid molecule is a plasmid. In one embodiment, the TALEN includes a spacer, such as 12 to 30 nucleotides in length. In one embodiment, the genetic disease is epidermolysis bullosa.

Another embodiment provides a method to treat a genetic disease or disorder caused by a genetic mutation comprising a) introducing into a cell (i) a first nucleic acid encoding a first transcription activator-like (TAL) effector endonuclease monomer, and (ii) a second nucleic acid encoding a second TAL effector endonuclease monomer, wherein each of said first and second TAL effector endonuclease monomers comprises a plurality of TAL effector repeat sequences and a FokI endonuclease domain, wherein each of said plurality of TAL effector repeat sequences comprises a repeat-variable di-residue, wherein said first TAL effector endonuclease monomer comprises the ability to bind to a first half-site sequence of a target DNA within said cell and comprises the ability to cleave said target DNA when said second TAL effector endonuclease monomer is bound to a second half-

site sequence of said target DNA, wherein said target DNA comprises said first half-site sequence and said second half-site sequence separated by a spacer sequence, and wherein said first and second half-sites have the same nucleotide sequence or different nucleotide sequences; and (b) culturing the cell under conditions in which the first and second TAL effector endonuclease monomers are expressed, so as to correct the mutation and restores correct gene expression.

The invention provides a nucleic acid encoding a TALEN and a nucleic acid donor sequence, wherein when the TALEN protein is expressed in a cell it induces a site-specific double stranded DNA break in a target gene, and further wherein the donor sequence is a template for DNA repair, which results in a correction of the genetic mutation and provides correct gene expression, so as to treat the genetic disease or disorder. The invention provides the nucleic acid, wherein the cell is a fibroblast, keratinocyte, inducible pluripotent-, hematopoietic-, mesenchymal-, or embryonic stem cell, hematopoietic progeny cell (such as a T-cell or B-cell), glia and neural cell, neuroglial progenitor and stem cell, muscle cell, lung cell, pancreatic and/or liver cell and/or a cell of the reticular endothelial system. The invention provides the nucleic acid, wherein the TALEN is a left TALEN and further comprising a right TALEN that cooperates with the left TALEN to make the double strand break in the target gene. The right TALEN may be encoded by the nucleic acid or a second nucleic acid. The left TALEN and the right TALEN may comprise a plurality of TAL effector repeat sequences and an endonuclease domain. Each of the left and right TALENS may comprise a spacer (distinct from the spacer sequence). The spacer sequence may be located between the plurality of TAL effector repeat sequences and the endonuclease domain. The spacer sequence may be encoded by a sequence of 12 to 30 nucleotides. The invention provides the nucleic acid, wherein said nucleic acid encoding the TALEN and/or the nucleic acid donor sequence is part of a vector or plasmid. The invention provides the nucleic acid, wherein the target gene is a gene with a genetic alteration/mutation. The invention provides the nucleic acid, wherein the target gene is COL7A1. The invention provides the nucleic acid, wherein the TALEN includes a spacer. The invention provides the nucleic acid wherein the spacer sequence is 12 to 30 nucleotides in length. The invention provides the nucleic acid, wherein the genetic disease is epidermolysis bullosa, osteogenesis imperfecta, dyskeratosis congenital,

the mucopolysaccharidoses, muscular dystrophy, cystic fibrosis (CFTR), fanconi anemia, the sphingolipidoses, the lipofuscinoses, adrenoleukodystrophy, severe combined immunodeficiency, sickle-cell anemia or thalassemia. The invention provides the nucleic acid, where in the genetic disease is epidermolysis bullosa. The invention provides at least one nucleic acid comprising (i) a first nucleic acid encoding a first transcription activator-like (TAL) effector endonuclease monomer, (ii) a second nucleic acid encoding a second TAL effector endonuclease monomer, and (iii) and a donor sequence, wherein each of said first and second TAL effector endonuclease monomers comprises a plurality of TAL effector repeat sequences and a FokI endonuclease domain, wherein each of said plurality of TAL effector repeat sequences comprises a repeat-variable diresidue, wherein said first TAL effector endonuclease monomer comprises the ability to bind to a first half-site sequence of a target DNA within said cell and comprises the ability to cleave said target DNA when said second TAL effector endonuclease monomer is bound to a second half-site sequence of said target DNA, wherein said target DNA comprises said first half-site sequence and said second half-site sequence separated by a spacer sequence, and wherein said first and second half-sites have the same nucleotide sequence or different nucleotide sequences, wherein said donor sequence comprises homology to the target at least at the 5' and 3's ends of the target sequence and the preselected genetic alteration and is a template for DNA repair resulting in a correction of the genetic mutation; and (b) culturing the cell under conditions in which the first and second TAL effector endonuclease monomers are expressed, so as to correct the mutation and restores correct gene expression. The invention provides a protein coded for or expressed by the nucleic acid. The invention provides a vector or plasmid comprising the nucleic acid. The invention provides an isolated host cell comprising the nucleic acid.

The invention provides for the use of the nucleic acids, vectors, host cells, and proteins of the invention to treat a genetic disease or disorder caused by a genetic mutation.

## Brief Description of the Drawings

Figures 1A-F. TALEN targeting, nuclease architecture and modification of COL7A1 gene. (a) COL7A1 target site on chromosome 3 and TALEN array binding. A schematic of human chromosome three and the region in exon 13 that was

targeted is shown. Arrows refer to primer sets used for subsequent analyses, and the line with mottled grey box is the donor used in (f). (b) COL7A1 target site and the core constituents of the nuclease complex. The TALEN is comprised of an N-terminal deletion of 152 residues of *Xanthomonas* TALEs, followed by the repeat domain, and a +63 C-terminal subregion fused to the catalytic domain of the FokI nuclease. (SEQ ID NO: 33; SEQ ID NO: 34) (c) Repeat Variable Diresidue (RVD) base recognition. The RVDs NN, NI, HD, and NG (that bind guanine, adenine, cytosine, and thymine, respectively) are coded to the corresponding full array in 1b. (d) Sketch of TALEN-generated (lightning bolt) double-stranded DNA break (DSB) and possible cellular repair mechanisms used for break repair. (SEQ ID NO: 35; SEQ ID NO: 36). (e) Error-prone non-homologous end-joining assessment by Sanger sequencing of TALEN-treated cells. Limiting cycle PCR was performed, followed by shotgun cloning; 75 clones were sequenced, with 64 showing 100% alignment to the genome database and 11 exhibiting non-homologous end joining (NHEJ)-induced deletions that are represented as dashes. The TALEN left and right target sites are in bold capital letters, and the spacer sequence is in lower-case letters. Total bases deleted are represented at right and signified as "del" followed by numbers of bases lost. (f) Homology-directed repair (HDR). The single-stranded oligonucleotide donor (ssODN) contained 65 bp of COL7A1 gene homology on the left arm and 101 bp on the right with a short, foreign sequence that serves as a unique primer site (mottled, grey box). Three primer PCR results in amplification with endogenous primer pairs (indicated with arrows labeled i. and iii.). TALEN insertion of the ODN results in a second, smaller PCR product size generated by primer pairs ii. and iii. The number at the bottom of the TALEN-treated cells indicates the rate of HDR determined by densitometry. (SEQ ID NOS: 37 to (SEQ ID NO: 48).

Figure 2. TALEN modification of *COL7A1* gene assessed by Surveyor nuclease assay. NHEJ assessment by Surveyor nuclease in RDEB fibroblasts. Limiting cycle PCR of a ~350 bp fragment was performed followed by Surveyor mismatch assay. TALEN induced NHEJ is evidenced by the predictable banding pattern of ~200 and 300 bp (arrows). At right is the unmodified COL7A1 locus in control cells.

Figures 3A-C. TALEN COL7A1 donor design and homology-directed repair.

(a) COL7A1 locus with mutation indicated by asterisk. Below is the donor, in alignment to its relation with the endogenous locus that is comprised of COL7A1 genomic sequences of a left arm 706 bp long and 100% homologous to the genomic locus. In between the left and right arms, designed so that it would be knocked into the intron between exons 12 and 13, is a floxed PGK puromycin cassette (box, loxp sites indicated by flanking arrows). The right arm was 806 bp long and contained 5 base changes. Four of these were silent point mutation polymorphisms (SPMPs) (referred to as upstream and downstream) that served as markers for identification of HDR-based events; the last was the normalized base that corrects the premature termination codon. The box represents three of the SPMPs that were located within 10 bp of one another. The normal (i.e., mutation reversion) base is denoted by the box and the terminal (downstream) SPMP that removes an ApaI restriction enzyme site is represented by a black box. Lightning bolt indicates the TALEN target site and the PCR primers (black arrows), designed so one was in the donor arm and the other outside it; utilized for analyses as shown. (SEQ ID NO: 49). SPMP detection in RDEB fibroblasts. TALEN treatment and PCR amplification followed by digestion with ApaI and Sanger sequencing shows the (b) presence of the ApaI-resistant SPMP that is derived from the donor and can only be present following TALEN cutting and homology-directed repair using the exogenous donor as the template, (SEQ ID NO: 50) (c) the unmodified base (ApaI sensitive) showing that a heterozygous HDR event occurred (SEQ ID NO: 51).

Figure 4A-B. *Cre* recombinase excision of PGK-puromycin. (a) Sketch of donor with floxed PGK puromycin. Introduction of a *Cre*-recombinase plasmid into puromycin resistant fibroblasts resulted in removal of the puromycin transgene. (b) Genomic loxp/*COL7A1* junction. PCR was used to demonstrate the presence of a loxP footprint (triangle/sequence below) in the intron between exons 12 and 13 in the RDEB TALEN/donor treated cells. (SEQ ID NO: 52).

Figure 5A-D. Early crossover event sequence analysis. (a) key for marker sequences introduced into the donor. Arrow=upstream SPMPs, line=the 1837 base causative for RDEB, arrow=downstream SPMPs. (SEQ ID NO: 53). Upstream crossover event. Sanger sequencing showing the incorporation of the upstream SPMPs (b) the maintenance of the mutation at base 1837 (SEQ ID NO: 54; (SEQ ID NO: 55) (c) and the absence of the downstream SPMP (SEQ ID NO: 56; (SEQ ID

NO: 57) (d) indicating that HDR occurred from the donor but failed to correct the mutation. Legend has been fixed to include D (SEQ ID NO: 58; (SEQ ID NO: 59).

Figure 6A-D. Sketch of putative early cross over event. (a) TALEN arrays are shown binding to the target sequence and the donor is shown below. (b) binding to target site and TALEN dimerization mediate a double stranded DNA break (lightning) and stimulation of HDR using the donor as the repair template. (c) Theoretical cross-over events. Alignment of the endogenous DNA and the donor results in a cross over event (Cross Over #1) where genetic material is exchanged in a manner where the upstream SPMPs (box) are incorporated while the second crossover (arrow/Cross Over #2) event happens upstream of the corrective base and downstream SPMP. (d) Resolved genomic sequence containing partial donor sequences (lines and box) with maintenance of the mutated base (box).

Figure 7A-C. Schematic of HDR and normal mRNA production. (a) Mutated endogenous *COL7A1* locus with TALEN target site indicated by lightning. Mutated base is shown and underneath is the donor that results in the (b) repair of the locus with permanent presence of donor-derived sequences from exon 12 through the intron between exons 15 and 16. (c) mRNA analysis. The indicated primers amplified a product that contains the corrective base (box and the *Apa*I SPMP black box) in the same amplicon.

Figure 8A. Sequence analysis of TALEN cutting of donor. (SEQ ID NO: 60). (a) cDNA from TALEN treated RDEB fibroblasts was analyzed by direct Sanger sequencing. The TALEN site is outlined in a red box (note that it is a partial TALEN sequence as the remainder of the site is within the adjacent intron. Arrow shows an exon/exon boundary). The RDEB mutation is underlined and showed a reversion to the wild type status (mutant=T, normal=C). The downstream *Apa*I SPMP is present and shown. Sequence alignment is of the cDNA sequence expected to be encoded by the donor on top and the recovered sequence on the bottom. The dashes/gaps show the deletions likely due to post-HDR TALEN cutting that induced subsequent NHEJ (non-homologous end joining). (SEQ ID NO:61; SEQ ID NO: 62).

Figures 9A-F. TALEN-mediated gene editing of COL7A1 with HDR and resultant normalized gene and protein expression. (a) TALEN-corrected cells with conversion of the mutation to wild-type status, (SEQ ID NO: 64) and (b) restoration of collagen type VII production assessed by immunofluorescence. (c) Homozygous

RDEB premature termination codon cDNA sequencing, (SEQ ID NO: 65) and (d) absence of type VII collagen protein production. (e) Sanger sequencing of wild-type COL7A1 locus, (SEQ ID NO: 66) and (f) type VII collagen expression. Cells were stained simultaneously and confocal microscopy exposure times and instrument setting were identical. Nuclei are stained with DAPI and show as blue.

Figure 10A-B. Sanger sequencing of mRNA from TALEN corrected fibroblasts. (a)Fibroblast clone 1-19 (SEQ ID NO: 67; SEQ ID NO: 68) and (b) 1-21 showed the presence of the corrected base (line) and the downstream SPMP (arrow). (SEQ ID NO:69; SEQ ID NO: 70).

Figures 11A-D. TALEN integration mapping profile. (a) Schematic of TALEN-induced DNA break that accepts the GFP cargo, permanently marking the genomic locus. (b) TALEN and IDLV co-expression in 293 cells resulted in stable GFP cells (flow cytometry analysis performed 6 weeks post TALEN and IDLV delivery). (c) Schema for linear amplification-mediated PCR. Blue arrow denotes the LAM PCR primer, and the dashed lines represent the products of linear amplification that were subsequently cloned and mapped to determine the TALEN-induced IDLV genomic fusion fragment. (d) (nr)LAM PCR/PCR identified integrants. LAM PCR sequence recovery and genome database search revealed five sites into which the IDLV integrated. Sequences mapped to the spacer region of the COL7A1 target site and four off-target sites at chromosomes 7, 16, 1, and 5 (none of the latter sequences were derived from a coding exon). (SEQ ID NOs: 71-75).

Figure 12A-B. Integrase deficient lentivirus. (a) sketch of GFP viral cassette that was produced with a defective integrase. (b) 293 IDLV GFP expression time course in the absence of TALENs over sequential analyses over 9 days showing rapid loss of GFP.

Figures 13 and 14 depict constructs.

## Detailed Description of the Invention

The invention is directed to transcription activator-like effector nuclease (TALEN)-mediated DNA editing of disease-causing mutations in the context of the human genome and human cells to treat patients with compromised genetic disorders. This is an advance over previous gene therapy trials/tools that rely on the provision of functional copies of a therapeutic gene that integrate at random or semi-random into the genome. The consequences of the previous gene therapy methods

are perturbation of the locus where the cargo lands and potential gene inactivation or dysregulation. These can result in life threatening side effects. The approach described herein maximizes safety and efficacy by employing a tailor made TALEN for, for example, the human genes that corrects the mutation spot alone while preserving the remainder of the genome in pristine condition – in other words, there is no disruption of the remaining genome, thus eliminating the off targets effects associated with the existing technology (e.g., viral-mediated gene-addition). This is a novel approach and is the first personalized gene therapy with TALEN-mediated transgene-free correction of disease causing mutation in cells, for example, human cells. Thus, the technology can be used in cells, such as human cells, such that a loss-of-function mutation can be seamlessly corrected with restoration of normal cellular function. In other embodiments, gene expression can be enhanced.

Definitions

In describing and claiming the invention, the following terminology will be used in accordance with the definitions set forth below. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention. Specific and preferred values listed below for radicals, substituents, and ranges are for illustration only; they do not exclude other defined values or other values within defined ranges for the radicals and substituents.

As used herein, the articles "a" and "an" refer to one or to more than one, i.e., to at least one, of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The term "about," as used herein, means approximately, in the region of, roughly, or around. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20%.

The term "isolated" refers to a factor(s), cell or cells which are not associated with one or more factors, cells or one or more cellular components that are associated with the factor(s), cell or cells *in vivo*.

"Cells" include cells from, or the "subject" is, a vertebrate, such as a mammal, including a human. Mammals include, but are not limited to, humans, farm animals, sport animals and companion animals. Included in the term "animal" is dog, cat, fish, gerbil, guinea pig, hamster, horse, rabbit, swine, mouse, monkey (e.g., ape, gorilla, chimpanzee, or orangutan), rat, sheep, goat, cow and bird.

A "control" subject is a subject having the same characteristics as a test subject, such as a similar type of disease, etc. The control subject may, for example, be examined at precisely or nearly the same time the test subject is being treated or examined. The control subject may also, for example, be examined at a time distant from the time at which the test subject is examined, and the results of the examination of the control subject may be recorded so that the recorded results may be compared with results obtained by examination of a test subject.

A "test" subject is a subject being treated.

A "disease" is a state of health of a subject wherein the subject cannot maintain homeostasis, and wherein if the disease is not ameliorated then the subject's health continues to deteriorate. In contrast, a "disorder" in a subject is a state of health in which the subject is able to maintain homeostasis, but in which the subject's state of health is less favorable than it would be in the absence of the disorder. However, the definitions of "disease" and "disorder" as described above are not meant to supersede the definitions or common usage related to specific addictive diseases or disorders.

A disease, condition, or disorder is "alleviated" if, for example, the severity of a symptom of the disease or disorder, the frequency with which such a symptom is experienced by a patient, or both, are reduced.

As used herein, an "effective amount" means, for example, an amount sufficient to produce a selected effect, such as alleviating symptoms of a disease or disorder.

The term "measuring the level of expression" or "determining the level of expression" as used herein refers to, for example, any measure or assay which can be used to correlate the results of the assay with the level of expression of a gene or protein of interest. Such assays include measuring the level of mRNA, protein levels, etc. and can be performed by assays such as northern and western blot analyses, binding assays, immunoblots, etc. The level of expression can include

rates of expression and can be measured in terms of the actual amount of an mRNA or protein present.

As used herein, the term "pharmaceutically acceptable carrier" includes, for example, any of the standard pharmaceutical carriers, such as a phosphate buffered saline solution, water, emulsions such as an oil/water or water/oil emulsion, and various types of wetting agents. The term also encompasses any of the agents approved by a regulatory agency of the US Federal government or listed in the US Pharmacopeia for use in animals, including humans.

The term "pharmaceutically-acceptable salt" refers to, for example, salts which retain the biological effectiveness and properties of the compounds of the present invention and which are not biologically or otherwise undesirable. In many cases, the compounds of the present invention are capable of forming acid and/or base salts by virtue of the presence of amino and/or carboxyl groups or groups similar thereto.

By the term "specifically binds," as used herein, is meant, for example, a molecule which recognizes and binds a specific molecule, but does not substantially recognize or bind other molecules in a sample.

The term "symptom," as used herein, refers to, for example, any morbid phenomenon or departure from the normal in structure, function, or sensation, experienced by the patient and indicative of disease.

As used herein, the term "treating" may include prophylaxis of the specific disease, disorder, or condition, or alleviation of the symptoms associated with a specific disease, disorder or condition and/or preventing or eliminating the symptoms. A "prophylactic" treatment is, for example, a treatment administered to a subject who does not exhibit signs of a disease or exhibits only early signs of the disease for the purpose of decreasing the risk of developing pathology associated with the disease. "Treating" is used interchangeably with "treatment" herein.

A "therapeutic" treatment is, for example, a treatment administered to a subject who exhibits symptoms of pathology for the purpose of diminishing or eliminating those symptoms.

A "therapeutically effective amount" of a compound is, for example, that amount of compound which is sufficient to provide a beneficial effect to the subject to which the compound is administered.

As used herein, "amino acids" are represented by the full name thereof, by the three letter code corresponding thereto, or by the one-letter code corresponding thereto, as indicated in the following table:

| Full Name | Three-Letter Code | One-Letter Code |
|---|---|---|
| Aspartic Acid | Asp | D |
| Glutamic Acid | Glu | E |
| Lysine | Lys | K |
| Arginine | Arg | R |
| Histidine | His | H |
| Tyrosine | Tyr | Y |
| Cysteine | Cys | C |
| Asparagine | Asn | N |
| Glutamine | Gln | Q |
| Serine | Ser | S |
| Threonine | Thr | T |
| Glycine | Gly | G |
| Alanine | Ala | A |
| Valine | Val | V |
| Leucine | Leu | L |
| Isoleucine | Ile | I |
| Methionine | Met | M |
| Proline | Pro | P |
| Phenylalanine | Phe | F |
| Tryptophan | Trp | W |

The expression "amino acid" as used herein is meant to include both natural and synthetic amino acids, and both D and L amino acids. "Standard amino acid" means any of the twenty standard L-amino acids commonly found in naturally occurring peptides. "Nonstandard amino acid residue" means any amino acid, other than the standard amino acids, regardless of whether it is prepared synthetically or derived from a natural source. As used herein, "synthetic amino acid" also encompasses chemically modified amino acids, including but not limited to salts, amino acid derivatives (such as amides), and substitutions. Amino acids contained within the peptides of the present invention, and particularly at the carboxy- or amino-terminus, can be modified by methylation, amidation, acetylation or substitution with other chemical groups which can change the peptide's circulating half-life without adversely affecting their activity. Additionally, a disulfide linkage may be present or absent in the peptides of the invention.

The term "amino acid" is used interchangeably with "amino acid residue," and may refer to a free amino acid and to an amino acid residue of a peptide. It will

be apparent from the context in which the term is used whether it refers to a free amino acid or a residue of a peptide.

Amino acids may be classified into seven groups on the basis of the side chain R: (1) aliphatic side chains; (2) side chains containing a hydroxyl (OH) group; (3) side chains containing sulfur atoms; (4) side chains containing an acidic or amide group; (5) side chains containing a basic group; (6) side chains containing an aromatic ring; and (7) proline, an imino acid in which the side chain is fused to the amino group.

As used herein, the term "conservative amino acid substitution" is defined herein as exchanges within one of the following five groups:

I. Small aliphatic, nonpolar or slightly polar residues:

Ala, Ser, Thr, Pro, Gly;

II. Polar, negatively charged residues and their amides:

Asp, Asn, Glu, Gln;

III. Polar, positively charged residues:

His, Arg, Lys;

IV. Large, aliphatic, nonpolar residues:

Met Leu, Ile, Val, Cys

V. Large, aromatic residues:

Phe, Tyr, Trp

As used herein, the term "nucleic acid" encompasses RNA as well as single, double and triple stranded DNA and cDNA. Furthermore, the terms, "nucleic acid," "DNA," "RNA" and similar terms also include nucleic acid analogs, i.e. analogs having other than a phosphodiester backbone. For example, the so called "peptide nucleic acids," which are known in the art and have peptide bonds instead of phosphodiester bonds in the backbone, are considered within the scope of the present invention. By "nucleic acid" is also meant any nucleic acid, whether composed of deoxyribonucleosides or ribonucleosides, and whether composed of phosphodiester linkages or modified linkages such as phosphotriester, phosphoramidate, siloxane, carbonate, carboxymethylester, acetamidate, carbamate, thioether, bridged phosphoramidate, bridged methylene phosphonate, bridged phosphoramidate, bridged phosphoramidate, bridged methylene phosphonate, phosphorothioate, methylphosphonate, phosphorodithioate, bridged

phosphorothioate or sulfone linkages, and combinations of such linkages. The term nucleic acid also specifically includes nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine and uracil). Conventional notation is used herein to describe polynucleotide sequences: the left-hand end of a single-stranded polynucleotide sequence is the 5'-end; the left-hand direction of a double-stranded polynucleotide sequence is referred to as the 5'-direction. The direction of 5' to 3' addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction. The DNA strand having the same sequence as an mRNA is referred to as the "coding strand"; sequences on the DNA strand which are located 5' to a reference point on the DNA are referred to as "upstream sequences"; sequences on the DNA strand which are 3' to a reference point on the DNA are referred to as "downstream sequences."

Unless otherwise specified, a "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

"Homologous" as used herein, refers to the subunit sequence similarity between two polymeric molecules, e.g., between two nucleic acid molecules, e.g., two DNA molecules or two RNA molecules, or between two polypeptide molecules. When a subunit position in both of the two molecules is occupied by the same monomeric subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then they are homologous at that position. The homology between two sequences is a direct function of the number of matching or homologous positions, e.g., if half (e.g., five positions in a polymer ten subunits in length) of the positions in two compound sequences are homologous then the two sequences are 50% homologous, if 90% of the positions, e.g., 9 of 10, are matched or homologous, the two sequences share 90% homology. By way of example, the DNA sequences 3'ATTGCC5' and 3'TATGGC share 50% homology.

As used herein, "homology" is used synonymously with "identity."

The determination of percent identity between two nucleotide or amino acid sequences can be accomplished using a mathematical algorithm. For example, a mathematical algorithm useful for comparing two sequences is the algorithm of Karlin and Altschul (1990, Proc. Natl. Acad. Sci. USA 87:2264-2268), modified as

in Karlin and Altschul (1993, Proc. Natl. Acad. Sci. USA 90:5873-5877). This algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990, J. Mol. Biol. 215:403-410), and can be accessed, for example at the National Center for Biotechnology Information (NCBI) world wide web site. BLAST nucleotide searches can be performed with the NBLAST program (designated "blastn" at the NCBI web site), using, for example, the following parameters: gap penalty = 5; gap extension penalty = 2; mismatch penalty = 3; match reward = 1; expectation value 10.0; and word size = 11 to obtain nucleotide sequences homologous to a nucleic acid described herein. BLAST protein searches can be performed with the XBLAST program (designated "blastn" at the NCBI web site) or the NCBI "blastp" program, using the following parameters: expectation value 10.0, BLOSUM62 scoring matrix to obtain amino acid sequences homologous to a protein molecule described herein. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997, Nucleic Acids Res. 25:3389-3402). Alternatively, PSI-Blast or PHI-Blast can be used to perform an iterated search which detects distant relationships between molecules (Id.) and relationships between molecules which share a common pattern. When utilizing BLAST, Gapped BLAST, PSI-Blast, and PHI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

The terms "comprises," "comprising," and the like can have the meaning ascribed to them in U.S. Patent Law and can mean "includes," "including" and the like. As used herein, "including" or "includes" or the like means including, without limitation.

TALENS

Transcription Activator-Like Effector Nucleases (TALENs) are artificial restriction enzymes generated by fusing the TAL effector DNA binding domain to a DNA cleavage domain. These reagents enable efficient, programmable, and specific DNA cleavage and represent powerful tools for genome editing *in situ*. Transcription activator-like effectors (TALEs) can be quickly engineered to bind practically any

DNA sequence. The term TALEN, as used herein, is broad and includes a monomeric TALEN that can cleave double stranded DNA without assistance from another TALEN. The term TALEN is also used to refer to one or both members of a pair of TALENs that are engineered to work together to cleave DNA at the same site. TALENs that work together may be referred to as a left-TALEN and a right-TALEN, which references the handedness of DNA. See USSN 12/965,590; USSN 13/426,991 (US 8,450,471); USSN 13/427,040 (US 8,440,431); USSN 13/427,137 (US 8, 440,432); and USSN 13/738,381, all of which are incorporated by reference herein in their entirety.

TAL effectors are proteins secreted by *Xanthomonas* bacteria. The DNA binding domain contains a highly conserved 33-34 amino acid sequence with the exception of the 12th and 13th amino acids. These two locations are highly variable (Repeat Variable Diresidue (RVD)) and show a strong correlation with specific nucleotide recognition. This simple relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA binding domains by selecting a combination of repeat segments containing the appropriate RVDs.

The non-specific DNA cleavage domain from the end of the FokI endonuclease can be used to construct hybrid nucleases that are active in a yeast assay. These reagents are also active in plant cells and in animal cells. Initial TALEN studies used the wild-type FokI cleavage domain, but some subsequent TALEN studies also used FokI cleavage domain variants with mutations designed to improve cleavage specificity and cleavage activity. The FokI domain functions as a dimer, requiring two constructs with unique DNA binding domains for sites in the target genome with proper orientation and spacing. Both the number of amino acid residues between the TALEN DNA binding domain and the FokI cleavage domain and the number of bases between the two individual TALEN binding sites are parameters for achieving high levels of activity. The number of amino acid residues between the TALEN DNA binding domain and the FokI cleavage domain may be modified by introduction of a spacer (distinct from the spacer sequence) between the plurality of TAL effector repeat sequences and the FokI endonuclease domain. The spacer sequence may be 12 to 30 nucleotides.

The relationship between amino acid sequence and DNA recognition of the TALEN binding domain allows for designable proteins. In this case artificial gene

synthesis is problematic because of improper annealing of the repetitive sequence found in the TALE binding domain. One solution to this is to use a publicly available software program (DNAWorks) to calculate oligonucleotides suitable for assembly in a two step PCR; oligonucleotide assembly followed by whole gene amplification. A number of modular assembly schemes for generating engineered TALE constructs have also been reported. Both methods offer a systematic approach to engineering DNA binding domains that is conceptually similar to the modular assembly method for generating zinc finger DNA recognition domains.

Once the TALEN genes have been assembled they are inserted into plasmids; the plasmids are then used to transfect the target cell where the gene products are expressed and enter the nucleus to access the genome. TALENs can be used to edit genomes by inducing double-strand breaks (DSB), which cells respond to with repair mechanisms. In this manner, they can be used to correct mutations in the genome which, for example, cause disease.

Vectors and Nucleic Acids

A variety of nucleic acids may be introduced into cells to obtain expression of a gene. As used herein, the term nucleic acid includes DNA, RNA, and nucleic acid analogs, and nucleic acids that are double-stranded or single-stranded (i.e., a sense or an antisense single strand). Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the nucleic acid. Modifications at the base moiety include deoxyuridine for deoxythymidine, and 5-methyl-2'-deoxycytidine and 5-bromo-2'-doxycytidine for deoxycytidine. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a six membered, morpholino ring, or peptide nucleic acids, in which the deoxyphosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. See, Summerton and Weller (1997) Antisense Nucleic Acid Drug Dev. 7(3):187; and Hyrup et al. (1996) Bioorgan. Med. Chem. 4:5. In addition, the deoxyphosphate backbone can be replaced with, for example, a phosphorothioate or phosphorodithioate backbone, a phosphoroamidite, or an alkyl phosphotriester backbone.

Nucleic acid sequences can be operably linked to a regulatory region such as a promoter. Regulatory regions can be from any species. As used herein, operably linked refers to positioning of a regulatory region relative to a nucleic acid sequence in such a way as to permit or facilitate transcription of the target nucleic acid. Any type of promoter can be operably linked to a nucleic acid sequence. Examples of promoters include, without limitation, tissue-specific promoters, constitutive promoters, and promoters responsive or unresponsive to a particular stimulus (e.g., inducible promoters).

Additional regulatory regions that may be useful in nucleic acid constructs, include, but are not limited to, polyadenylation sequences, translation control sequences (e.g., an internal ribosome entry segment, IRES), enhancers, inducible elements, or introns. Such regulatory regions may not be necessary, although they may increase expression by affecting transcription, stability of the mRNA, translational efficiency, or the like. Such regulatory regions can be included in a nucleic acid construct as desired to obtain optimal expression of the nucleic acids in the cell(s). Sufficient expression, however, can sometimes be obtained without such additional elements.

A nucleic acid construct may be used that encodes signal peptides or selectable markers. Signal peptides can be used such that an encoded polypeptide is directed to a particular cellular location (e.g., the cell surface). Non-limiting examples of selectable markers include puromycin, ganciclovir, adenosine deaminase (ADA), aminoglycoside phosphotransferase (neo, G418, APH), dihydrofolate reductase (DHFR), hygromycin-B-phosphtransferase, thymidine kinase (TK), and xanthin-guanine phosphoribosyltransferase (XGPRT). Such markers are useful for selecting stable transformants in culture. Other selectable markers include fluorescent polypeptides, such as green fluorescent protein or yellow fluorescent protein.

Nucleic acid constructs can be introduced into cells of any type using a variety of techniques. Non-limiting examples of techniques include the use of transposon systems, recombinant viruses that can infect cells, or liposomes or other non-viral methods such as electroporation, microinjection, or calcium phosphate precipitation, that are capable of delivering nucleic acids to cells.

Nucleic acids can be incorporated into vectors. A vector is a broad term that

includes any specific DNA segment that is designed to move from a carrier into a target DNA. A vector may be referred to as an expression vector, or a vector system, which is a set of components needed to bring about DNA insertion into a genome or other targeted DNA sequence such as an episome, plasmid, or even virus/phage DNA segment. Vectors most often contain one or more expression cassettes that comprise one or more expression control sequences, wherein an expression control sequence is a DNA sequence that controls and regulates the transcription and/or translation of another DNA sequence or mRNA, respectively.

Many different types of vectors are known. For example, plasmids and viral vectors, e.g., retroviral vectors, are known. Mammalian expression plasmids typically have an origin of replication, a suitable promoter and optional enhancer, and also any necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences. Examples of vectors include: plasmids (which may also be a carrier of another type of vector), adenovirus, adeno-associated virus (AAV), lentivirus (e.g., modified HIV-1, SIV or FIV), retrovirus (e.g., ASV, ALV or MoMLV), and transposons (e.g., Sleeping Beauty, P-elements, Tol-2, Frog Prince, piggyBac).

Therapeutic Uses

TALEN-based gene correction has many clinical and preclinical (e.g., research) applications. For example, TALEN-based gene correction can used to correct genes in which mutations lead to disease. For example, any disease characterized by small base alterations including insertions and deletions such as, but not restricted to, epidermolysis bullosa, osteogenesis imperfecta, dyskeratosis congenital, the mucopolysaccharidoses, muscular dystrophy, cystic fibrosis (CFTR), fanconi anemia, the sphingolipidoses, the lipofuscinoses,, adrenoleukodystrophy, severe combined immunodeficiency, sickle-cell anemia, thalassemia, and the like.

In one embodiment, the disease is Epidermolysis Bullosa. Recessive dystrophic epidermolysis bullosa (RDEB) is characterized by a functional deficit of the type VII collagen protein due to gene defects in the type VII collagen (COL7A1) gene. This gene encodes the alpha chain of type VII collagen. The type VII collagen fibril, composed of three identical alpha collagen chains, is restricted to the basement zone beneath stratified squamous epithelia. It functions as an anchoring

fibril between the external epithelia and the underlying stroma. Mutations in this gene are associated with all forms of dystrophic epidermolysis bullosa.

COL7A1 is located on the short arm of human chromosome 3, in the chromosomal region denoted 3p21.31 (Ensembl No: ENSG00000114270). The gene is approximately 31,000 base pairs in size and its coding sequence is fragmented into 118 exons, see SEQ ID NO: 32.

COL7A1 is transcribed into an mRNA of 9,287 base pairs (Accession Nos. for human mRNA and protein are NM_000094 and NP_000085, respectively). In the skin, the type VII collagen protein is synthesized by keratinocytes and dermal fibroblasts. The symbol for the orthologous gene in the mouse is Col7a1 (Accession No for Mouse mRNA and protein are NM_00738 and NP_031764, respectively).

People with RDEB exhibit incurable, often fatal skin blistering and are at increased risk for aggressive squamous cell carcinoma[1]. Gene augmentation therapies are promising, but run the risk of insertional mutagenesis. It is therefore described herein engineered transcription activator like effector nucleases (TALENs) for precision genome-editing in cells of patients with RDEB. It is described herein the ability of TALENs to induce site-specific double-stranded DNA breaks (DSB) leading to homology-directed repair (HDR) from an exogenous donor template. This process resulted in COL7A1 gene mutation correction and restoration of normal gene and protein expression. This study provides proof-of-concept for personalized genomic medicine and is the first TALEN-mediated *in situ* correction of an endogenous human gene in fibroblasts.

Cells to be modified by TALEN-based gene correction can be obtained from the patient or from a donor. The cells can be of any type, such as fibroblast cells, keratinocytes, inducible pluripotent-, hematopoietic-, mesenchymal-, and embryonic stem cells, hematopoietic progeny cells, such as T-cells, B-cells, glia and neurons, neuroglial progenitor and stem cells, muscle cells, lung cells, pancreatic and liver cells and/or cells of the reticular endothelial system). Once modified by TALEN-based gene correction, the cells can be expanded and/or administered to a patient to treat the disease.

Matrices can be used to deliver cells of the present invention to specific anatomic sites, where particular growth factors may or may not be incorporated into the matrix, or encoded on plasmids incorporated into the matrix for uptake by the

cells, can be used to direct the growth of the initial cell population. Plasmid DNA encoding cytokines, growth factors, or hormones can be trapped within a polymer gene-activated matrix carrier. The biodegradable polymer is then implanted near the site where treatment is desired.

For the purposes described herein, either autologous, allogeneic or xeongenic cells of the present invention can be administered to a patient by direct injection to a preselected site, systemically, on or around the surface of an acceptable matrix, or in combination with a pharmaceutically acceptable carrier.

Additionally, nucleic acid constructs or proteins can be injected locally or systemically into a subject, with, for example, a pharmaceutically acceptable carrier.

Growth/Expansion of Cells

Cells to be modified by TALEN-based gene correction can be obtained from the patient or from a donor. The cells can be of any type, such as fibroblast cells. Once modified by TALEN-based gene correction, the cells can be expanded and/or administered to a patient to treat the disease.

The cells can be cultured in culture medium that is established in the art and commercially available from the American Type Culture Collection (ATCC), Invitrogen and other companies. Such media include, but are not limited to, Dulbecco's Modified Eagle's Medium (DMEM), DMEM F12 medium, Eagle's Minimum Essential Medium, F-12K medium, Iscove's Modified Dulbecco's Medium, Knockout D-MEM, or RPMI-1640 medium. It is within the skill of one in the art to modify or modulate concentrations of media and/or media supplements as needed for the cells used. It will also be apparent that many media are available as low-glucose formulations, with or without sodium pyruvate.

Also contemplated is supplementation of cell culture medium with mammalian sera. Sera often contain cellular factors and components that are needed for viability and expansion. Examples of sera include fetal bovine serum (FBS), bovine serum (BS), calf serum (CS), fetal calf serum (FCS), newborn calf serum (NCS), goat serum (GS), horse serum (HS), human serum, chicken serum, porcine serum, sheep serum, rabbit serum, rat serum (RS), serum replacements (including, but not limited to, KnockOut Serum Replacement (KSR, Invitrogen)), and bovine embryonic fluid. It is understood that sera can be heat-inactivated at 55-65°C if deemed needed to inactivate components of the complement cascade. Modulation

of serum concentrations, or withdrawal of serum from the culture medium can also be used to promote survival of one or more desired cell types. In one embodiment, the cells are cultured in the presence of FBS /or serum specific for the species cell type. For example, cells can be isolated and/or expanded with total serum (e.g., FBS) or serum replacement concentrations of about 0.5% to about 5% or greater including about 5% to about 15% or greater, such as about 20%, about 25% or about 30%. Concentrations of serum can be determined empirically.

Additional supplements can also be used to supply the cells with trace elements for optimal growth and expansion. Such supplements include insulin, transferrin, sodium selenium, and combinations thereof. These components can be included in a salt solution such as, but not limited to, Hanks' Balanced Salt Solution™ (HBSS), Earle's Salt Solution™, antioxidant supplements, MCDB-201™ supplements, phosphate buffered saline (PBS), N-2-hydroxyethylpiperazine-N'-ethanesulfonic acid (HEPES), nicotinamide, ascorbic acid and/or ascorbic acid-2-phosphate, as well as additional amino acids. Many cell culture media already contain amino acids; however some require supplementation prior to culturing cells. Such amino acids include, but are not limited to, L-alanine, L-arginine, L-aspartic acid, L-asparagine, L-cysteine, L-cystine, L-glutamic acid, L-glutamine, L-glycine, L-histidine, L-inositol, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, and L-valine.

Antibiotics are also typically used in cell culture to mitigate bacterial, mycoplasmal, and fungal contamination. Typically, antibiotics or anti-mycotic compounds used are mixtures of penicillin/streptomycin, but can also include, but are not limited to, amphotericin (Fungizone™), ampicillin, gentamicin, bleomycin, hygromycin, kanamycin, mitomycin, mycophenolic acid, nalidixic acid, neomycin, nystatin, paromomycin, polymyxin, puromycin, rifampicin, spectinomycin, tetracycline, tylosin, and zeocin.

Hormones can also be advantageously used in cell culture and include, but are not limited to, D-aldosterone, diethylstilbestrol (DES), dexamethasone, $\beta$-estradiol, hydrocortisone, insulin, prolactin, progesterone, somatostatin/human growth hormone (HGH), thyrotropin, thyroxine, and L-thyronine. $\beta$-mercaptoethanol can also be supplemented in cell culture media.

Lipids and lipid carriers can also be used to supplement cell culture media, depending on the type of cell and the fate of the differentiated cell. Such lipids and carriers can include, but are not limited to cyclodextrin ( $\alpha$ , $\beta$ , $\gamma$ ), cholesterol, linoleic acid conjugated to albumin, linoleic acid and oleic acid conjugated to albumin, unconjugated linoleic acid, linoleic-oleic-arachidonic acid conjugated to albumin, oleic acid unconjugated and conjugated to albumin, among others. Albumin can similarly be used in fatty-acid free formulation.

Cells in culture can be maintained either in suspension or attached to a solid support, such as extracellular matrix components and synthetic or biopolymers. Cells often require additional factors that encourage their attachment to a solid support (e.g., attachment factors) such as type I, type II, and type IV collagen, concanavalin A, chondroitin sulfate, fibronectin, "superfibronectin" and/or fibronectin-like polymers, gelatin, laminin, poly-D and poly-L-lysine, Matrigel™, thrombospondin, and/or vitronectin.

Cells can be cultured at different densities, e.g., cells can be seeded or maintained in the culture dish at different densities. For example, at densities, including, but not limited to, densities of less than about 2000 cells/well of a 12-well plate (for example, 12-well flat-bottom growth area: 3.8cm2 well volume: 6.0 ml or well ID x depth (mm) 22.1x17.5; well capacity (ml) 6.5, growth area (cm2) 3.8), including less than about 1500 cells/well of a 12-well plate, less than about 1,000 cells/well of a 12-well plate, less than about 500 cells/well of a 12-well plate, or less than about 200 cells/well of a 12-well plate. The cells can also be seeded or maintained at higher densities, for example, great than about 2,000 cells/well of a 12-well plate, greater than about 2,500 cells/well of a 12-well plate, greater than about 3,000 cells/well of a 12-well plate, greater than about 3,500 cells/well of a 12-well plate, greater than about 4,000 cells/well of a 12-well plate, greater than about 4,500 cells/well of a 12-well plate, greater than about 5,000 cells/well of a 12-well plate, greater than about 5,500 cells/well of a 12-well plate, greater than about 6,000 cells/well of a 12-well plate, greater than about 6,500 cells/well of a 12-well plate, greater than about 7,000 cells/well of a 12-well plate, greater than about 7,500 cells/well of a 12-well plate or greater than about 8,000 cells/well of a 12-well plate.

## Examples

The following example is provided in order to demonstrate and further

illustrate certain embodiments and aspects of the present invention and is not to be construed as limiting the scope thereof.

Example 1

Materials and Methods.

Research subject and cell line derivation.

After obtaining informed parental consent we obtained a punch biopsy from the skin of a male RDEB patient with a homozygous c.1837 C>T premature termination codon mutation. Approval for research on human subjects was obtained from the University of Minnesota Institutional Review Board. A primary fibroblast cell line was derived and maintained in low oxygen concentration conditions.

TALEN and donor construction.

The TALEN candidate described in Fig. 1A was generated via the Golden Gate Assembly method and inserted into a homodimeric form of a CAGGs promoter driven *Fok*I endonuclease as described [1, 2]. The left donor arm was amplified with the LAF and LAR primers shown in Table 1. The right arm was synthesized in two fragments (inner and outer) using an overlapping oligonucleotide assembly strategy as described [3, 4]. All primer sets are shown in Table 1; the left and right arms were cloned into a floxed PGK puromycin cassette.

**Table 1. (SEQ ID NOs: 1-21)**

TALEN correction for RDEB

| | | |
|---|---|---|
| C06 | outer fragment 1-12 | TCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCTTAGGAGAGAAGCGGAGGAATC |
| C07 | C7GT1 | Atcgtcccacatccctgtctctt |
| C08 | C7APAF | CAAAGGGACCAATGAGGGTA |
| C09 | C7GT2 | tctagtggggagaggcaatg |
| C10 | RT1 | TCGACTTGGATGACGTTCAG |
| C11 | RT2 | GTTCGAGCCACGATGACTG |
| C12 | Surveyor F | tttcagccatatcccagctc |
| D01 | Surveyor R | tgctccagctaatcegaaat |
| D02 | Oligo Duplex Top | G*T*CCGTACGGATCCAAGCTTCGTCGACCTAGCC |
| D03 | Oligo Duplex Bottom | CATGCCTAGGTTCGAAGCAGCTGGATCGGGG*A*C |
| D04 | Linker F | GGATCCAAGCTTCGTCGACCTAGCC |
| D05 | ssODN donor (PAGE purified) | tctgcgtccc tgtccatcac tgccatcgtc ccacatcccc gtctctttct gacccctgcccacct |
| D06 | | agtagtgtgtgcccgtctgtt gt gtgactctggtaa ctagag atcctcagacccttttagtc agtggaigac gttcaggctg ggcttagcta cactgtgagg gtgtctgctc gagtgggttc cogtgagggc a |
| D07 | Off target surveyor primers | |
| D08 | 1q23.3 FWD | TCTCAGGCAAGAAAATTGGA |
| D09 | 1q23.3 REV | TGTGCATTTATTCTGTGTCTTGTT |
| D10 | 5q33.1 FWD | GAGTTCCCTTGGGCCTATTC |
| D11 | 5q33.1 REV | GGCTGCAGTGAGCTATGATG |
| D12 | 7q21.3 FWD | ACTCCAAGTCACAGGGGATG |
| E01 | 7q21.3 REV | CAGCTCTGACTGCTGTTTGC |
| E02 | 16p13.3 FWD | TTGCTCACAGAAGGACCACA |
| E03 | 16p13.3 REV | ACGTGGGTGTGACGGTTATT |

Gene transfer.

All TALEN treatments consisted of delivery of 2.5 μg of each TALEN and

10 μg amount of donor via the Neon Transfection System (Life Sciences) with the following instrument settings: 1500 V, 20ms pulse width, and a single pulse. For 48 hours post gene transfer the cells were incubated at 31 C[5].

Cell Culture.

Cells were maintained in growth media comprised of DMEM supplemented with 20 % FBS, 100 U/mL nonessential amino acids, and 0.1 mg/ml each of penicillin and streptomycin, respectively (Invitrogen) and cultured at 2% $O_2$, 5% $CO_2$, and 37 C.

Surveyor nuclease.

Genomic DNA was isolated 48 hours post TALEN gene transfer and amplified for 30 cycles with Surveyor F and Surveyor R primers and subjected to Surveyor nuclease treatment as described [6]. Products were resolved on a 10% TBE PAGE gel (Invitrogen). For off target amplicons the PCR reaction proceeded for 35 cycles and all primers are listed in Table 1.

Homology directed repair analysis.

For quantification of HDR, TALENs and 5 μl of a 40 μM single stranded oligonucleotide donor were transfected into cells and screened by PCR at 48 hours using three primers: Surveyor F, Surveyor R, and linker forward primers. Densitometry was performed as described [6]. For gene correction, 10 μg of the donor plasmid was introduced along with the 2.5 μg each of TALEN DNA and selection was performed as described subsequently.

Selection.

Cells were selected in bulk in 0.2 μg/mL puromycin, segregated into sub-pools, screened for HDR, and then plated at low density (250-750 total cells) in a 10 $cm^2$ dish. A cloning disk with silicone grease (all from Corning) was placed over single cells in the presence of base media supplemented with 10 ng/mL epidermal growth factor and 0.5 ng/mL fibroblast growth factor. Cells were expanded to sequentially larger vessels. An adenoviral cre recombinase was added at an MOI of 20 to remove the PGK puromycin cassette (Vector BioLabs).

Cell correction molecular screening.

C7GT1 and C7GT2 primer pairs were employed to amplify a junction from the donor into the endogenous locus (upstream SPMP screening). The *Apa*I SPMP region was assessed on genomic DNA treated with *Apa*I pre- and post- PCR

amplification with C7APAF and C7GT2. Messenger RNA from clonal isolates was converted to cDNA and screened with RT1 and RT2 and then digested with *Apa*I. *Apa*I-resistant amplicons were cloned and Sanger sequenced.

Cell Expansion Analysis.

Gene corrected fibroblasts were expanded in T150 flasks and trypsinized to obtain single cell suspensions. Cells were then resuspended in 100ul PBS + 0.5% BSA + propidium iodide (eBiosciences), followed by addition of an equal volume of PKH26 reference microbeads (SIGMA). Five thousand bead events were collected and absolute viable cell number was calculated as per manufacturer protocol (SIGMA).

iPSC Generation and teratoma assay.

Gene corrected fibroblasts (or un-corrected cells as a control) were reprogrammed to iPSCs as described [7, 8] and then placed in the flank of a SCID mouse until a visible mass formed. The mass was excised for embedding and staining.

Immunofluorescence.

Gene corrected cells were plated on a chamber slide and were fixed 24 hours later with 4% paraformaldehyde, permeabilized with 0.2% Triton X, blocked with 1% BSA and stained with a polyclonal anti-type VII collagen antibody (1:1500; generously provided by Drs David Woodley and Mei Chen). Secondary antibody staining was performed with donkey anti-rabbit IgG Cy3 (1:500; Jackson Immunoresearch). Isotype control staining was done using whole molecule rabbit IgG (Jackson Immunoresearch). Nuclei were stained with 4', 6-diamidino-2-phenylindole (Vector Laboratories). Images were taken using a PMT voltage of 745 on an Olympus BX61 FV500 confocal microscope (Olympus Optical Co LTD) and analyzed using the Fluoview software version 4.3. Light microscopy was performed on a Leica microscope.

IDLV and LAM-PCR/nrLAM PCR.

Integrase-defective lentiviral (IDLV) particles were produced in 293T cells via lipid based co-transfection (Lipofectamine 2000, Invitrogen) of the CMV-GFP transfer vector, the pCMV-ΔR8.2 packaging plasmid harboring the D64V integrase mutation [9, 10], and the pMD2.VSV-G envelope-encoding plasmid. Gene tagging was performed by nucleofection of HEK 293 cells with the TALENs followed 24

hours later by a transduction of GFP IDLV at an MOI of 7. 100ng of genomic DNA was analyzed in duplicate by LAM-PCR [11] using enzymes MseI and Tsp509I and nrLAM-PCR [12] to ensure genome–wide recovery of IDLV integration sites. (nr)LAM-PCR amplicons were sequenced by the Roche/454 pyrosequencing platform and integration site data were analyzed using the HISAP pipeline [13, 14],[15]. Genomic position harboring >1 IS in close distance were scanned for potential TALEN off-target binding sites using the pattern matcher scan-for-matches [13].

Results/Discussion

Lack of type VII collagen protein at the dermal-epidermal junction (DEJ) results in loss of the structural integrity of the skin. Restoration of deposition of the type VII collagen at the DEJ by allogeneic systemic hematopoietic cell or localized fibroblast transplantation can alleviate symptoms [16-18]. However, suboptimal efficacy of allogeneic cell transplantation due to risks of toxicity, infection, and graft failure provides impetus to develop new autologous cell-based therapies. Therefore, a genome-editing strategy for COL7A1 correction based on TALEN technology is described herein. Fibroblasts are an ideal cell type due to their ease of derivation and low susceptibility to growth arrest in culture as well as their ability to deposit type VII collagen at the DEJ [18, 19]. TALENs are engineered nucleases that can induce a double-stranded DNA break at a user-defined genomic locus, thus stimulating HDR, and are superior to other nucleases in their targeting capacity and ease of generation [20, 21].

The TAL Effector-Nucleotide Targeter software [22, 23]identified 68 potential TALEN sites for the human COL7A1 locus and support recent experimental data on a large series of human genes [21] emphasize the high targeting capacity for TALENs, a consideration for RDEB and other diseases that exhibit heterogeneity in the location and number of mutated sequences. The Golden Gate cloning methodology was used to generate a patient-specific nuclease proximal to a premature termination codon in exon 14 of the COL7A1 gene (Fig. 1A). A TALEN is composed of an engineered TALE repeat array fused to the FokI nuclease domain (Fig. 1B); the binding specificities of TALE repeats in the array are dictated by the identities of two hypervariable residues within each repeat (Fig. 1C). TALEN-treated RDEB fibroblasts were analyzed for evidence of repair by the two major

DNA repair pathways: error-prone non-homologous end-joining (NHEJ) and HDR. Surveyor nuclease assay and Sanger sequencing that showed 11 mutated alleles out of 75 total analyzed were consistent with NHEJ (Figs. 2A and 2E). TALEN cleavage also resulted in the capture of an oligonucleotide duplex at the DNA break site (Figs. 2B-F)[24]. These data established that the nuclease is active at the target site. It was next ascertained whether RDEB cells could undergo HDR following co-delivery of TALENs and an oligonucleotide donor (ODN) containing a unique primer sequence flanked by short donor arms (Fig. 1F). RDEB fibroblasts transfected with TALEN plasmids and the ODN were then analyzed with a three-primer PCR approach that simultaneously detects the modified and unmodified alleles. This assay showed that TALENs in RDEB cells can stimulate HDR to incorporate an exogenous sequence from the ODN donor (Fig. 1G) and the 14.6% rate of NHEJ and 2.1% rate of HDR show the efficacy of TALEN use for high-level modification of human fibroblasts.

To determine whether a COL7A1 mutation causing RDEB could be corrected and a population of genetically corrected cells subsequently expanded, an exogenous donor plasmid was generated that would allow for selective detection and expansion of gene-corrected cells. This donor consisted of homology arms that spanned ~1 kb of the COL7A1 locus between exons 12 and 16 (Fig. 3A). Within the donor was a floxed-PGK-puromycin cassette oriented so that it would be inserted into the intron between exons 12 and 13. The flanking loxP sites allow for removal of the selectable marker with Cre recombinase, leaving a small loxP "footprint" in the intron (Fig. 4). Within the right donor arm, five single base pair alterations were engineered: the normal base at the site of the mutation that restores a normal genotype and four silent point mutation polymorphisms (SPMPs) that allowed for delineation of HDR-modified alleles versus unmodified ones (Fig. 3A). Three of these SPMPs are upstream of the target base and the one downstream removes an ApaI restriction site (alterations hereafter referred to as upstream or downstream SPMPs).

Of the nine clones analyzed, four were obtained that showed evidence of HDR. In one clone, the presence of the upstream SPMPs was evident; however, the RDEB-pathogenic COL7A1 mutation persisted and the downstream SPMP was not found (Fig. 5). These data suggest that an HDR crossover event occurred within the donor arm upstream of the region that restores a normal genotype (Fig. 6). For the

remaining three clones, however, the downstream donor-inserted SPMP was detectable, indicating that one allele underwent HDR and the other did not, resulting in a heterozygous COL7A1 locus (Figs. 3B and 3C).

HDR should revert the mutant base and restore normal gene expression. Accordingly, this was assessed with an RT-PCR strategy for the detection of the normal base and the downstream SPMPs in the same transcript following splicing out of the intervening intron (Fig. 7). Interestingly, direct sequencing of the cDNA in one clone showed a deletion of sequences at the TALEN target site (Fig. 8). These data indicate that the TALEN was active after HDR and induced an additional NHEJ-mediated mutation. Previous studies with zinc finger endonucleases (ZFNs) show that silent mutations in the donor sequence can reduce the frequency of this undesired event[12]; however, this was not possible in this experiment because the TALEN site was at an intron/exon boundary and it was opted to leave the donor TALEN sequence unperturbed so as not to disrupt splicing. This negatively impacted the recovery of one clone; however, two clones exhibited the desired HDR-based, donor-derived, normal transcripts (Fig. 9A). It was next ascertained whether TALEN treatment restored type VII collagen protein expression compared to untreated RDEB mutant or wild-type cells bearing abnormal or normal transcripts, respectively (Figs. 9C and 9E). Immunofluorescence-based detection of type VII collagen revealed a rescue of type VII collagen production in TALEN-treated cells and a complete absence in untreated control RDEB fibroblasts (Figs. 9B and 9D). These results confirm the ability of TALENs to mediate a genetic modification at a disease-specific target site with restoration of normal mRNA and protein production.

The risk of off-target effects is a consideration in the clinical use of genome-editing reagents. Options for mapping off-target sites of gene-editing nucleases include: (i) performing *in vitro* Systematic Evolution of Ligands by Exponential Enrichment (SELEX) with monomeric DNA-binding proteins of each nuclease in a pair and then using this data to predict potential off target sites[25], (ii) performing an *in vitro* cleavage site selection using dimeric nucleases and then interrogating sites from this selection that occur in the genome of cells of interest for nuclease-induced mutations, (iii) utilizing the propensity of an integration-defective lentivirus (IDLV) to integrate into nuclease-induced DSBs and then identifying points of insertion by LAM-PCR[9]. Although methods (ii) and (iii) appear to be better at

identifying nuclease off-target sites than method (i), the former methods fail to identify off-target sites predicted by the other, suggesting that no method is comprehensive in its detection of off-target events. Method (iii) was utilized with an IDLV with green fluorescent protein (GFP) gene that can be trapped into a nuclease-generated DSB (Fig. 11A)[9, 26]. Human embryonic kidney (293) cells were used due to their accelerated proliferative capacity, which should promote rapid dilution of non-integrated IDLV and minimize random integration. In addition, it was hypothesized that, due to the open chromatin structure of 293 cells, any off-target effects will manifest to a greater degree than in primary cells and will allow for a more sensitive mapping of off-target events. Introduction of the GFP IDLV alone resulted in a rapid loss of GFP expression in 293 cells (Fig. 12). The co-introduction of IDLV and TALENs resulted in a stable population of GFP cells (Fig. 11B), which were used for mapping the integration sites with nonrestrictive linear amplification-mediated PCR ((nr)LAM-PCR) (Fig. 11C). Five sites were recovered that showed a junction between the IDLV and adjacent genomic sequence (Fig. 11D). These events are not unexpected, as even nucleases used in clinical trials show off-target effects [9] and the non-coding regions recovered suggest that this TALEN possesses a safety profile that is not predicted to negatively impact gene expression.

At the resolution of the LAM-PCR methodology, the TALEN described herein shows a high rate of on-target activity. In addition, these studies, like others, show that a potential target for engineered nucleases is the donor construct itself and they highlight the benefits of the inclusion of a marker sequence that can aid in selection of the desired HDR event [27].

In summary, skin cells from an RDEB patient were obtained and the donor and TALEN reagents (sequences are included below) were designed and rapidly constructed to specifically target this unique mutation. The application of the gene editing tools resulted in correction of the RDEB mutation in diploid human fibroblasts—cells that are suitable for therapeutic use after direct expansion or reprogramming into pluripotency followed by expansion [7, 8] - and provide the first-ever demonstration of TALEN-mediated correction of a disease gene in the human genome. These studies provide the proof that TALENs can be used in the development of clinically relevant individualized therapies.

Example 2

An example of a Donor Plasmid Sequence is set forth in SEQ ID NO: 22. An example of the Left Arm of the Donor Sequence is set forth in SEQ ID NO:31. An example of the Loxp site of Donor is set forth in SEQ ID NO:23. An example of the PGK Promoter of Donor is set forth in SEQ ID NO:24. An example of the Puromycin Gene of the Donor sequence is set forth in SEQ ID NO:25. An example of the Bovine Growth Hormone polyadenylation signal of Donor is set forth in SEQ ID NO:26. An example of the Loxp Site Of Donor is set forth in SEQ ID NO:27. An example of the Right Arm of Donor is set forth in SEQ ID NO:28. An example of TALEN Left (pTAL 286) is set forth in SEQ ID NO:29. An example of TALEN Right (pTAL 287) is set forth in SEQ ID NO:30.

Bibliography

1. Carlson, D. F., et al. Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences of the United States of America* 109: 17382-17387.
2. Cermak, T., et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* 39: e82.
3. Osborn, M. J., Defeo, A. P., Blazar, B. R., and Tolar, J. Synthetic Zinc Finger Nuclease Design and Rapid Assembly. *Human gene therapy.*
4. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6: 343-345.
5. Doyon, Y., Choi, V. M., Xia, D. F., Vo, T. D., Gregory, P. D., and Holmes, M. C. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nature methods* 7: 459-460.
6. Guschin, D. Y., Waite, A. J., Katibah, G. E., Miller, J. C., Holmes, M. C., and Rebar, E. J. A rapid and general assay for monitoring endogenous gene modification. *Methods in molecular biology (Clifton, N.J* 649: 247-256.
7. Tolar, J., et al. Keratinocytes from Induced Pluripotent Stem Cells in Junctional Epidermolysis Bullosa. *The Journal of investigative dermatology.*
8. Tolar, J., et al. Induced pluripotent stem cells from individuals with recessive dystrophic epidermolysis bullosa. *The Journal of investigative dermatology* 131: 848-856.
9. Gabriel, R., et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology* 29: 816-823.
10. Vargas, J., Jr., Gusella, G. L., Najfeld, V., Klotman, M. E., and Cara, A. (2004). Novel integrase-defective lentiviral episomal vectors for gene transfer. *Human gene therapy* 15: 361-372.
11. Schmidt, M., et al. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* 4: 1051-1057.
12. Paruzynski, A., et al. (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat Protoc* 5: 1379-1395.
13. Dsouza, M., Larsen, N., and Overbeek, R. (1997). Searching for patterns in genomic data. *Trends Genet* 13: 497-498.
14. Arens, A., et al. Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites. *Human gene therapy methods* 23: 111-118.
15. Arens, A., et al. (2012). Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites. *Hum Gene Ther Methods* 23: 111-118.

16. Wagner, J. E., et al. Bone marrow transplantation for recessive dystrophic epidermolysis bullosa. *The New England journal of medicine* 363: 629-639.

17. Tolar, J., et al. (2009). Amelioration of epidermolysis bullosa by transfer of wild-type bone marrow cells. *Blood* 113: 1167-1174.

18. Wong, T., et al. (2008). Potential of fibroblast cell therapy for recessive dystrophic epidermolysis bullosa. *The Journal of investigative dermatology* 128: 2179-2189.

19. Goto, M., et al. (2006). Fibroblasts show more potential as target cells than keratinocytes in COL7A1 gene therapy of dystrophic epidermolysis bullosa. *The Journal of investigative dermatology* 126: 766-772.

20. Cermak, T., et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research.*

21. Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D., and Joung, J. K. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology* 30: 460-465.

22. Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F., and Dobbs, D. (2007). Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic acids research* 35: W599-605.

23. Doyle, E. L., et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic acids research* 40: W117-122.

24. Orlando, S. J., et al. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic acids research* 38: e152.

25. Pattanayak, V., Ramirez, C. L., Joung, J. K., and Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods* 8: 765-770.

26. Paruzynski, A., et al. Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nature protocols* 5: 1379-1395.

27. Zou, J., et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell stem cell* 5: 97-110.

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

## CLAIMS

1. A method to treat a genetic disease or disorder caused by a genetic mutation in a target gene comprising contacting a cell with

one or more nucleic acids encoding a TALEN protein, the TALEN protein comprising a plurality of TAL effector repeat sequences and an endonuclease domain, and

a nucleic acid donor sequence,

wherein the TALEN protein is expressed in the cell and induces a site-specific double stranded DNA break in the target gene, wherein the donor sequence is a template for DNA repair resulting in a correction of the genetic mutation and provides correct gene expression, so as to treat the genetic disease or disorder.

2. The method of claim 1, wherein the cell is selected from the group consisting of a fibroblast, keratinocyte, inducible pluripotent stem cell, hematopoietic stem cell, mesenchymal stem cell, embryonic stem cell, hematopoietic progeny cell, T-cell, B-cell, glial cell, neural cell, neuroglial progenitor cell, neuroglial stem cell, muscle cell, lung cell, pancreatic cell, liver cell and a cell of the reticular endothelial system.

3. The method of claim 1, wherein the TALEN protein is a left TALEN and the one or more nucleic acids further comprise a nucleic acid encoding a right TALEN capable of cooperating with the left TALEN to make the double strand break in the target gene.

4. The method of claim 1, wherein the nucleic acid encoding the TALEN or the nucleic acid donor sequence is part of a vector or plasmid.

5. The method of claim 1, wherein the target gene is a gene with a genetic alteration or mutation.

6. The method of claim 5, wherein the target gene is COL7A1.

7. The method of claim 1, wherein the nucleic acid encoding the TALEN protein comprises a spacer between the plurality of TAL effector repeat sequences and the endonuclease domain.

8. The method of claim 7, wherein the spacer is 12 to 30 nucleotides in length.

9. The method of claim 1, wherein the genetic disease is selected from the group consisting of epidermolysis bullosa, recessive dystrophic epidermolysis bullosa (RDEB), osteogenesis imperfecta, dyskeratosis congenital, a mucopolysaccharidosis, muscular dystrophy, cystic fibrosis (CFTR), fanconi

anemia, a sphingolipidosis, a lipofuscinosis, adrenoleukodystrophy, severe combined immunodeficiency, sickle-cell anemia and thalassemia.

10. The method of claim 1, where in the genetic disease is epidermolysis bullosa.

11. A method to treat a genetic disease or disorder caused by a genetic mutation comprising a) introducing into a cell (i) a first nucleic acid encoding a first transcription activator-like (TAL) effector endonuclease monomer, (ii) a second nucleic acid encoding a second TAL effector endonuclease monomer, and (iii) and a donor sequence, wherein each of the first and second TAL effector endonuclease monomers comprise a plurality of TAL effector repeat sequences and a FokI endonuclease domain, wherein each of the plurality of TAL effector repeat sequences comprises a repeat-variable diresidue, wherein the first TAL effector endonuclease monomer comprises the ability to bind to a first half-site sequence of a target DNA within the cell and comprises the ability to cleave the target DNA when the second TAL effector endonuclease monomer is bound to a second half-site sequence of the target DNA, wherein the target DNA comprises the first half-site sequence and the second half-site sequence separated by a spacer sequence, and wherein the first and second half-sites have the same nucleotide sequence or different nucleotide sequences, wherein the donor sequence comprises homology to the target at least at the 5' and 3's ends of the target sequence and the preselected genetic alteration and is a template for DNA repair resulting in a correction of the genetic mutation; and (b) culturing the cell under conditions in which the first and second TAL effector endonuclease monomers are expressed, so as to correct the mutation and restores correct gene expression.

12. A nucleic acid comprising a donor sequence, wherein the donor sequence is a template for site specific DNA repair resulting in a correction of a genetic mutation, wherein the donor sequence comprises homology to at least the 5' and 3' ends of the target sequence, wherein a portion of the donor sequence comprises a repair sequence to correct the target sequence for use in conjunction with a TALEN protein.

13. The nucleic acid of claim 12, wherein the donor comprises SEQ ID NO: 22.

14. The nucleic acid of claim 12, wherein the target is COL7A1.

15. The nucleic acid of claim 12, wherein the 5' and 3' ends of the donor each have at least 100 bases of sequence identity to the target.

16. A nucleic acid coding for a TALEN protein, wherein the nucleic acid comprises SEQ ID NO: 29.

17. A nucleic acid coding for a TALEN protein, wherein the nucleic acid comprises SEQ ID NO: 30.

18. A protein coded or expressed by the nucleic acid of claim 16.

19. A protein coded for or expressed by the nucleic acid of claim 17.

20. A vector or plasmid comprising a donor sequence, wherein the donor sequence is a template for site specific DNA repair resulting in a correction of a genetic mutation, wherein the donor sequence comprises homology to at least the 5' and 3' ends of the target sequence, wherein a portion of the donor sequence comprises a repair sequence to correct the target sequence for use in conjunction with a TALEN protein.

21. The vector or plasmid of claim 20, wherein the donor comprises SEQ ID NO:22.

22. The vector or plasmid of claim 20, wherein the target is COL7A1.

23. The vector or plasmid of claim 20, wherein the 5' and 3' ends of the donor each have at least 100 bases of sequence identity to the target.

24. A vector or plasmid comprising one or more of SEQ ID NOs: 22, 31, 28, 29 or 30.

25. An isolated host cell comprising one or more of exogenous SEQ ID NOs: 22, 31, 28, 29 or 30 or the proteins expressed from such sequences.

26. A transfected cell line comprising SEQ ID NOs: 22, 31, 28, 29 or 30 or the proteins expressed from such sequences.

27. A composition comprising

a nucleic acid encoding a first TALEN protein, wherein the first TALEN protein is capable of inducing a site-specific double stranded DNA break in a target gene in a cell, and

a nucleic acid donor sequence,

wherein the donor sequence is a template for correction of a genetic mutation in the target gene.

28. The composition of claim 27, wherein the cell is selected from the group consisting of a fibroblast, keratinocyte, inducible pluripotent stem cell,

hematopoietic stem cell, mesenchymal stem cell, embryonic stem cell, hematopoietic progeny cell, T-cell, B-cell, glial cell, neural cell, neuroglial progenitor cell, neuroglial stem cell, muscle cell, lung cell, pancreatic cell, liver cell and a cell of the reticular endothelial system.

29. The composition of claim 27, wherein the first TALEN protein is a left TALEN and the nucleic acid further encodes a second TALEN which is a right TALEN that cooperates with the left TALEN to make the site-specific double stranded DNA break in the target gene.

30. The composition of claim 27, wherein the nucleic acid encoding the TALEN protein or the nucleic acid donor sequence is part of a vector or plasmid.

31. The composition of claim 27, wherein the target gene is COL7A1.

32. The composition of claim 29, wherein the first TALEN or the second TALEN comprise a plurality of TAL effector repeat sequences and the endonuclease domain and a spacer between the plurality of TAL effector repeat sequences and the endonuclease domain includes a spacer.

33. The composition of claim 32, wherein the spacer is 12 to 30 nucleotides in length.

34. The composition of claim 27, wherein the genetic disease is selected from the group consisting of epidermolysis bullosa, recessive dystrophic epidermolysis bullosa (RDEB), osteogenesis imperfecta, dyskeratosis congenital, a mucopolysaccharidosis, muscular dystrophy, cystic fibrosis (CFTR), fanconi anemia, a sphingolipidosis, a lipofuscinosis, adrenoleukodystrophy, severe combined immunodeficiency, sickle-cell anemia and thalassemia.

35. The composition of claim 34, where in the genetic disease is epidermolysis bullosa.

36. A composition comprising at least one nucleic acid comprising

(i) a first nucleic acid encoding a first transcription activator-like (TAL) effector endonuclease monomer,

(ii) a second nucleic acid encoding a second TAL effector endonuclease monomer, and (iii) and a donor sequence,

wherein each of the first and second TAL effector endonuclease monomers comprises a plurality of TAL effector repeat sequences and a FokI endonuclease domain, wherein each of the plurality of TAL effector repeat sequences comprises a

repeat-variable diresidue, wherein the first TAL effector endonuclease monomer is capable of binding to a first half-site sequence of a target DNA within the cell, the target DNA having a genetic mutation, and the first TAL effector endonuclease monomer is capable of cleaving the target DNA when the second TAL effector endonuclease monomer is bound to a second half-site sequence of the target DNA, wherein the target DNA comprises the first half-site sequence and the second half-site sequence separated by a spacer sequence, and wherein the first and second half-sites have the same nucleotide sequence or different nucleotide sequences, wherein the donor sequence comprises a region homologous to the target DNA at least at the 5' and 3' ends of the target DNA and is a template for DNA repair resulting in a correction of the genetic mutation in the target DNA.

37. A protein coded for or expressed by the nucleic acid encoding the first TALEN protein of claim 27.

38. A protein coded for or expressed by the first nucleic acid or the second nucleic acid of claim 37.

39. A vector comprising the nucleic acid encoding the first TALEN protein claim 27.

40. A vector comprising the first nucleic acid or the second nucleic acid of claim 37.

41. A composition comprising

a nucleic acid encoding a first TALEN protein, wherein the first TALEN protein is capable of inducing a site-specific double stranded DNA break in a COL7A1 gene in a cell, the COL7A1 gene having a genetic mutation capable of causing epidermolysis bullosa, and

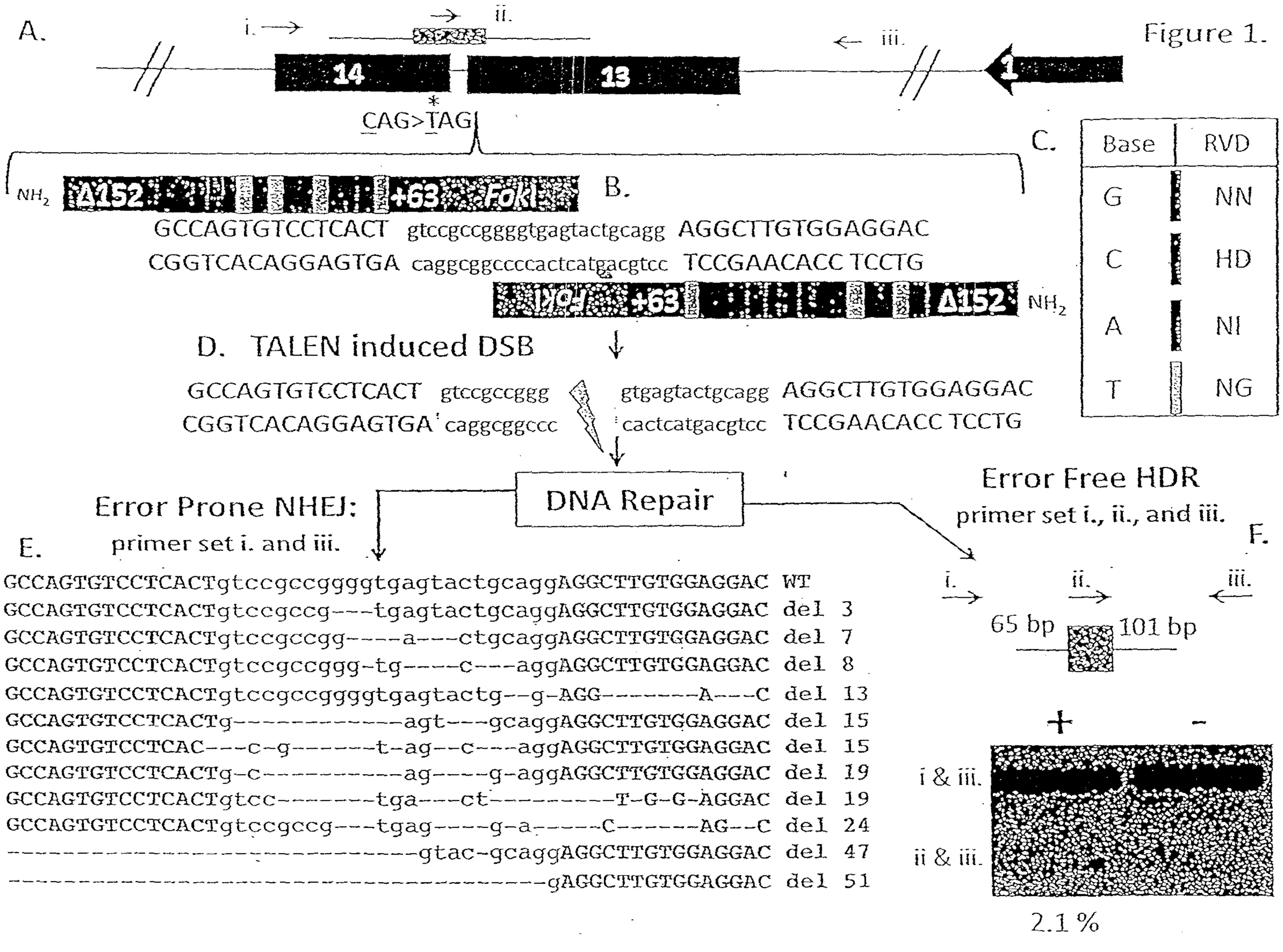
a nucleic acid donor sequence,

wherein the donor sequence is a template for correction of the genetic mutation in the COL7A1 gene.

CA 02902954 2015-08-27



Figure 1.




SUBSTITUTE SHEET (RULE 26)

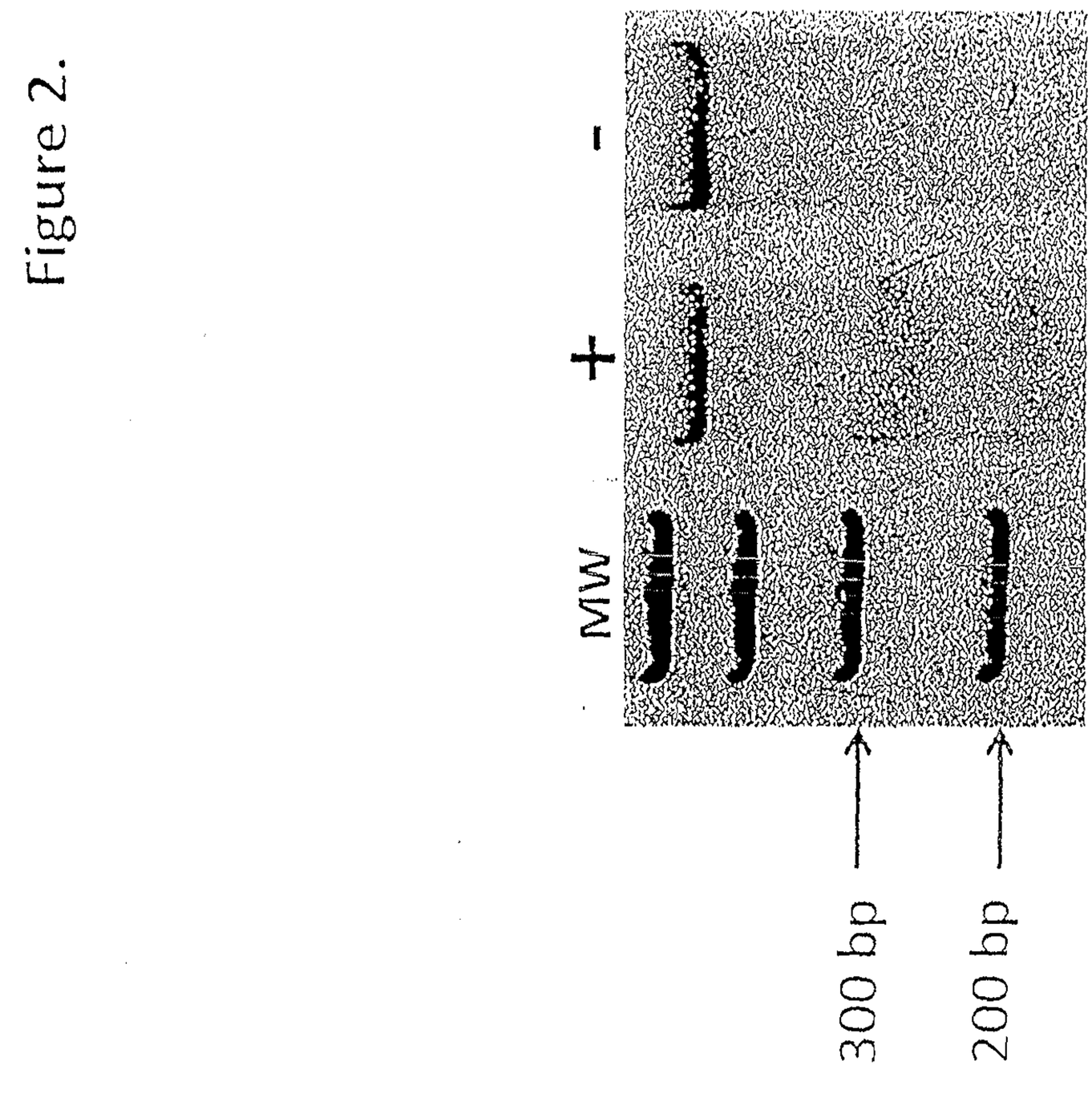



Figure 2.

Figure 3.

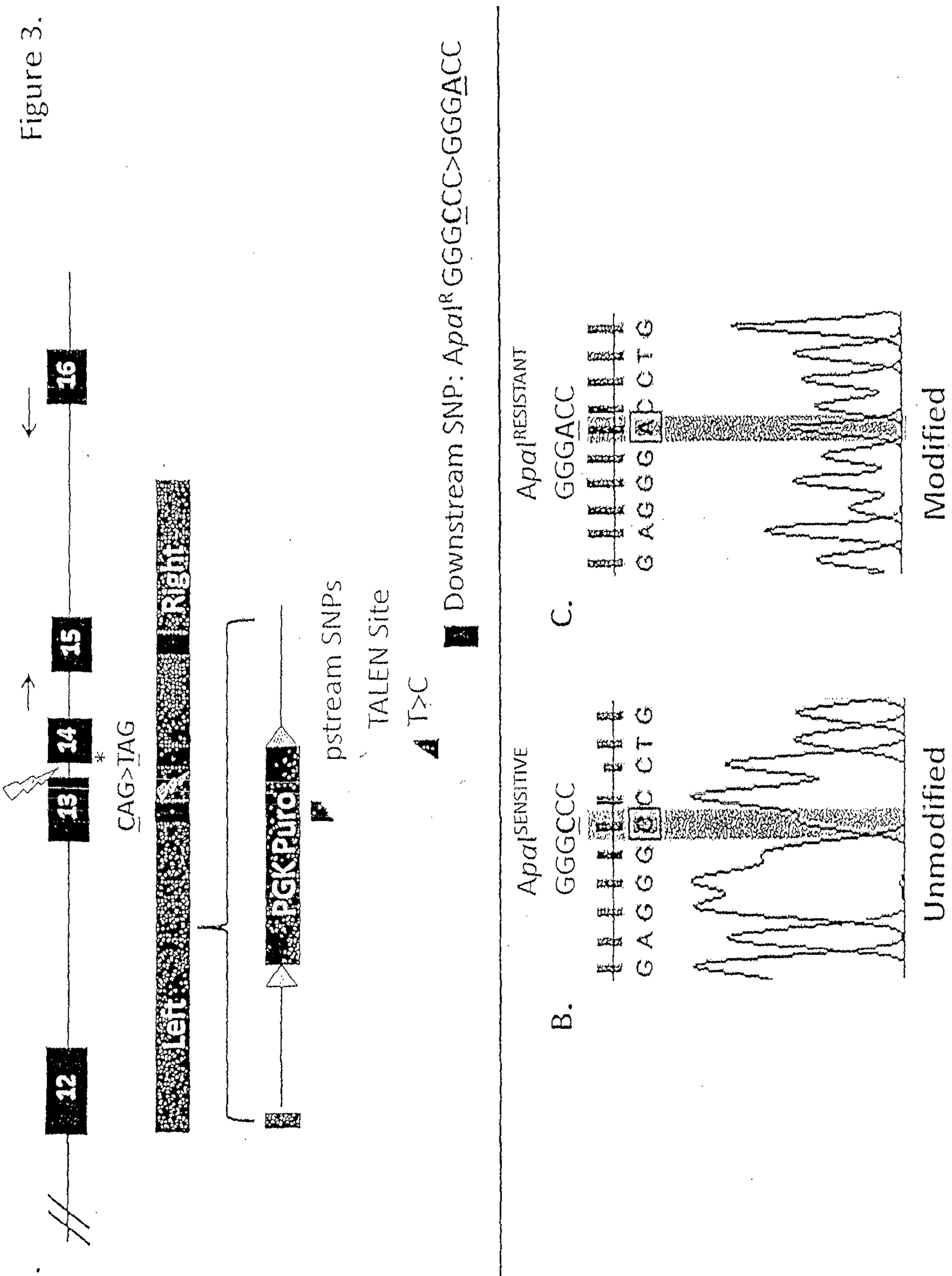

A.
12
13
14
15
16
CAG>TAG
Left
Right
PGK Puro
pstream SNPs
TALEN Site
T>C
Downstream SNP: ApaI^R GGGCCC>GGGACC
B.
ApaI^SENSITIVE
GGGCCC
C.
ApaI^RESISTANT
GGGACC
Unmodified
Modified

Figure 4

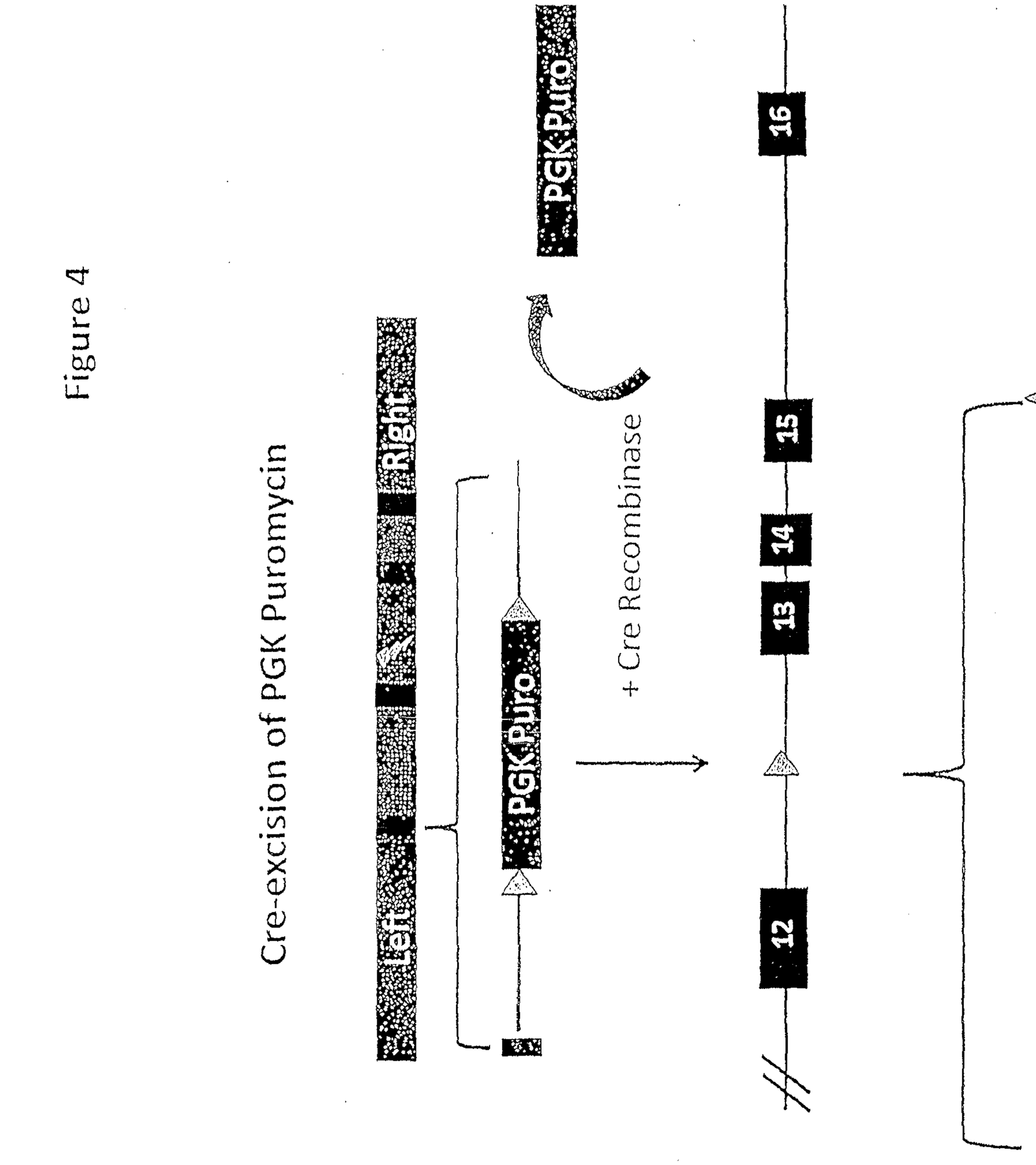
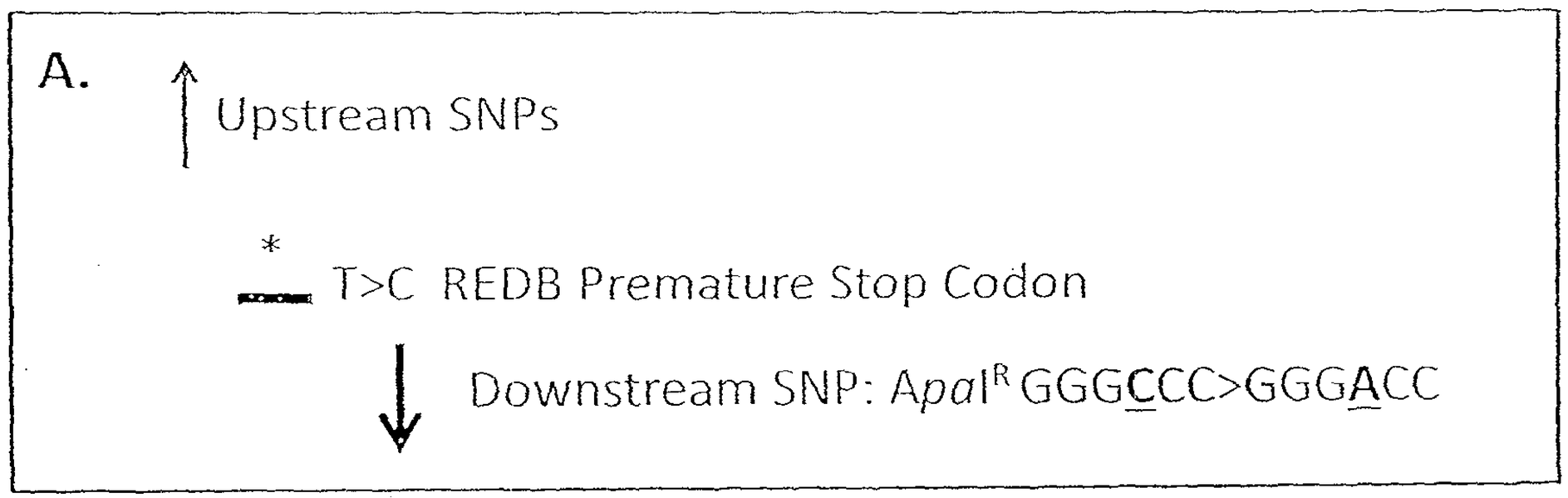

Cre-excision of PGK Puromycin
Left
PGK Puro
+ Cre Recombinase
Right
PGK Puro
12
13
14
15
16
TGGCCCTTCTCACTCTGCGTCCCTGTCCATCACTG
A
A.
B.

Figure 5





SUBSTITUTE SHEET (RULE 26)

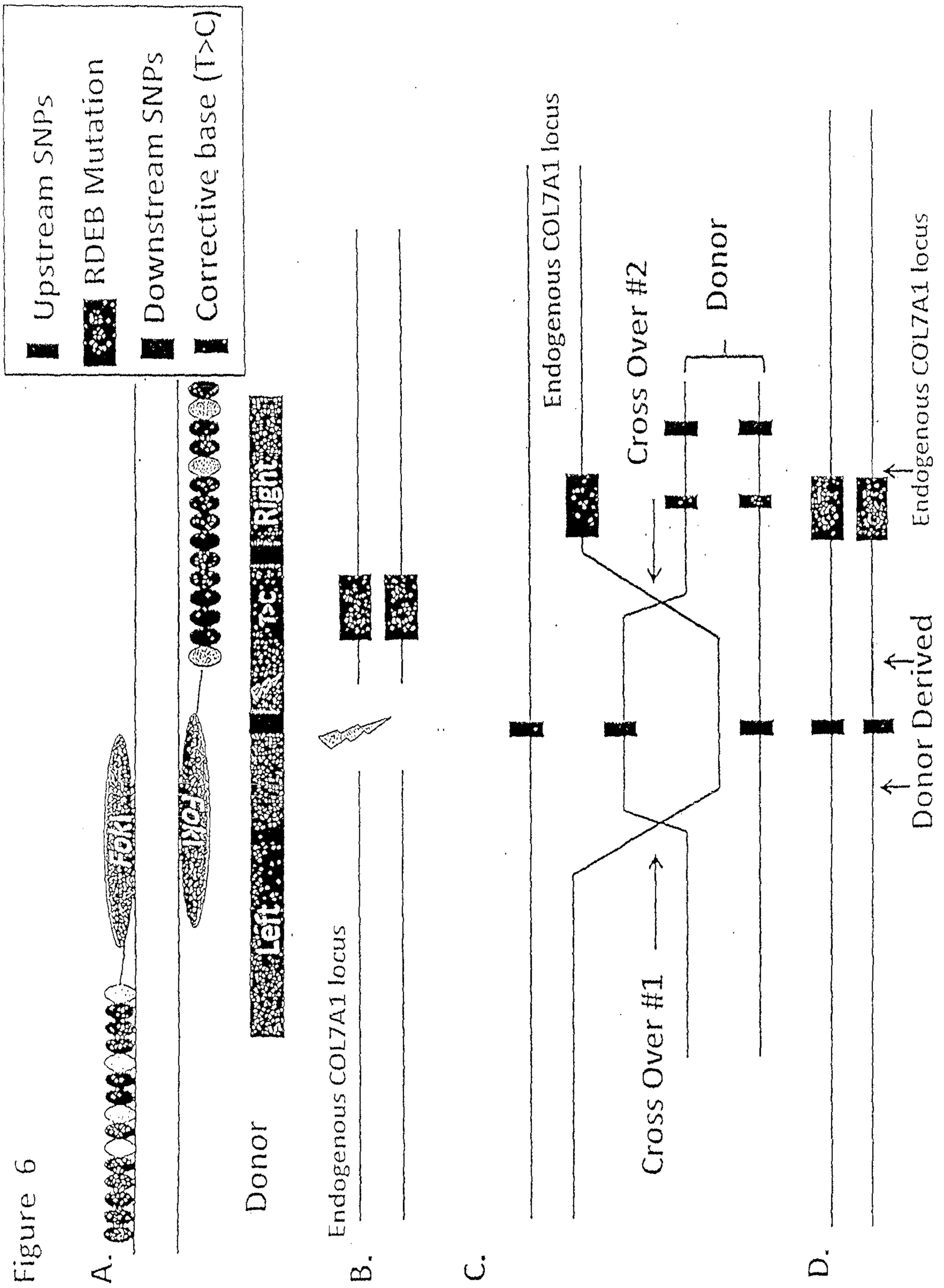

Upstream SNPs
RDEB Mutation
Downstream SNPs
Corrective base (T>C)
Left
Right
Endogenous COL7A1 locus
Cross Over #2
Donor
Endogenous COL7A1 locus
Derived
Cross Over #1
Donor
A.
B.
C.
D.

Figure 6

Figure 7

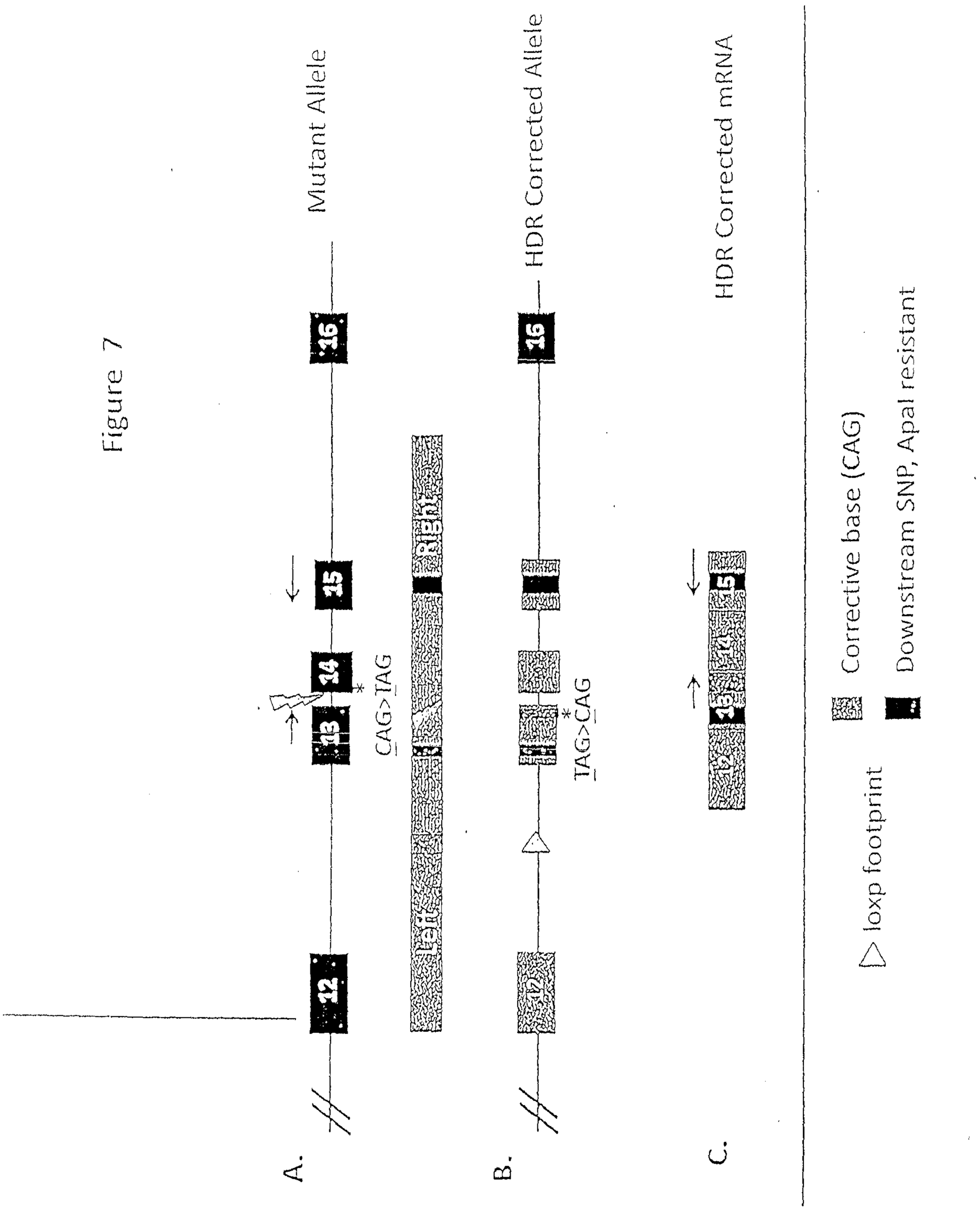

A.
B.
C.
Mutant Allele
HDR Corrected Allele
HDR Corrected mRNA
CAG>TAG
TAG>CAG
12
13
14
15
16
loxp footprint
Corrective base (CAG)
Downstream SNP, ApaI resistant

Figure 8

RDEB Mutation C>T

TALEN Target Site

GCCAGTGTCCTCAC GTCCGCCGGGGTGAGTACTGCAGG AGGCTTGTGGAGGAC

A.

Exon/Exon Boundary ↓

APA$^R$ GGGCCC>GGGACC ↑

```
TCGACTTGGATGACGTTCAGGCTGGGCTTAGCTACACTGTGCGGGTGTCTGCTCGAGTGG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TCGACTTGGATGACGTTCAGGCTGGGCTTAGCTACACTGTGCGGGTGTCTGCTCGAGTGG

GTCCCCGTGAGGGCAGTGCCAGTGTCCTCACTGTCCGCCGGGAGCCGGAAACTCCACTTG
|||||||||||  |  |  |||||  |  ||  ||||  |   |  |  ||    |  |  |
GTCCCCGTGA--G--G-G-CAGTG--C-CAGTGTCC-TC---A--CTG-----T-C-C--G
                                        *

CTGTTCCAGGGCTGCGGGTTGTGGTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGAC
|     |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
C----CGGGGGCTGCGGGTTGTGGTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGAC

CCTACCAGGTGGCTGTGTCGGTACTGCGAGGCAGAGAGGAGGGACCTGCTGCAGTCATCG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CCTACCAGGTGGCTGTGTCGGTACTGCGAGGCAGAGAGGAGGGACCTGCTGCAGTCATCG
```

SUBSTITUTE SHEET (RULE 26)

CA 02902954 2015-08-27

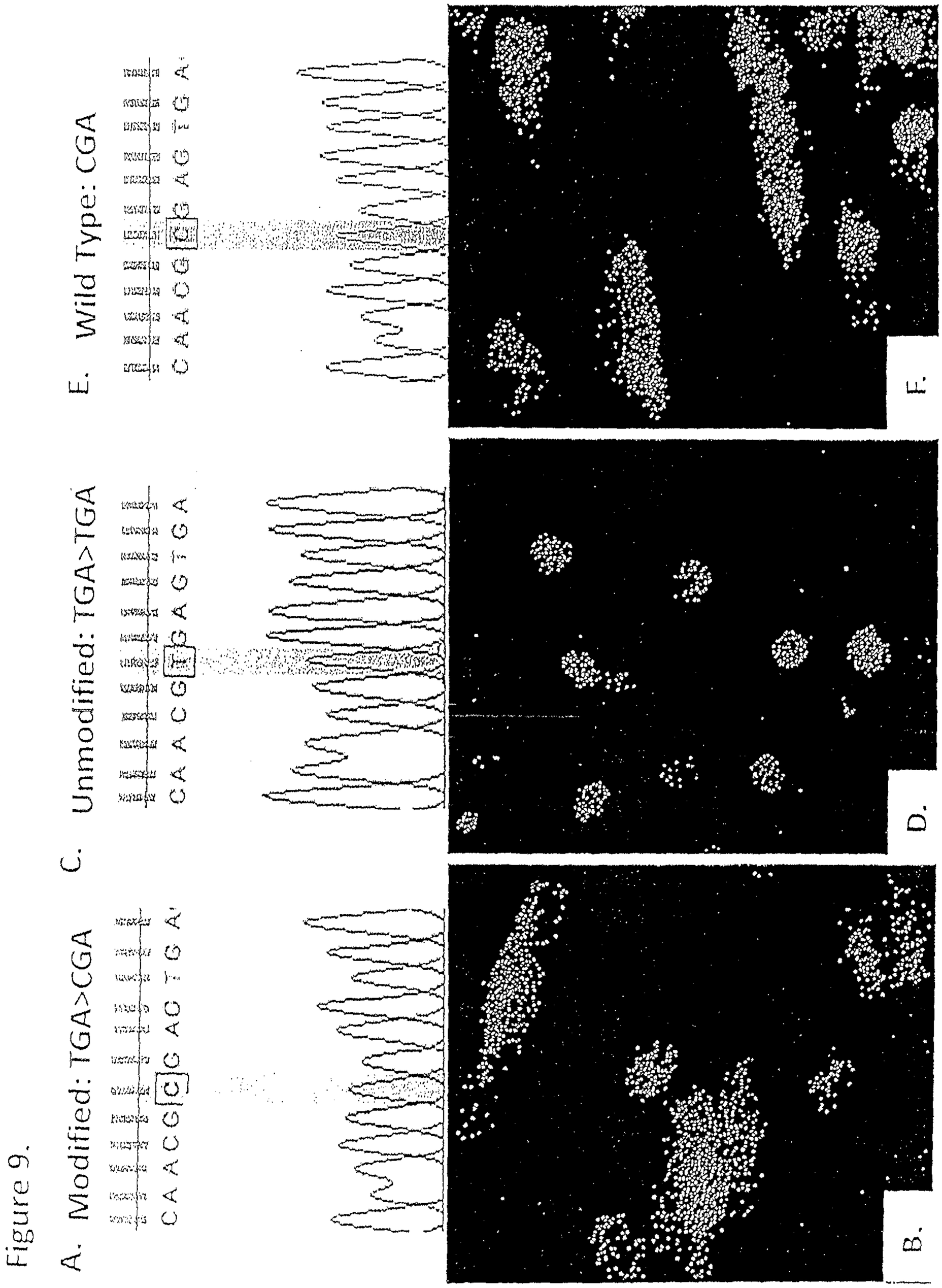

A. Modified: TGA>CGA
C A A C G C A G T G A
B.
C. Unmodified: TGA>TGA
C A A C G T A G T G A
D.
E. Wild Type: CGA
C A A C G C A G C G A
F.

Figure 9.

Figure 10

A.

```
GTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGACCCGTCCCTGGAGCCAGTGGATTT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGACCCGTCCCTGGAGCCAGTGGATTT

CGGATTAGCTGGAGCACAGGCAGTGGTCCGGAGTCCAGCCAGACACTGCCCCCAGACTCT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CGGATTAGCTGGAGCACAGGCAGTGGTCCGGAGTCCAGCCAGACACTGCCCCCAGACTCT

ACTGCCACAGACATCACAGGGCTGCAGCCTGGAACCACCTACCAGGTGGCTGTGTCGGTA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACTGCCACAGACATCACAGGGCTGCAGCCTGGAACCACCTACCAGGTGGCTGTGTCGGTA

CTGCGAGGCAGAGAGGAGGGACCTGCTGCAGTCATCGTGGCTCGAAC
||||||||||||||||||||||||||||||||||||||||||||||
CTGCGAGGCAGAGAGGAGGGACCTGCTGCAGTCATCGTGGCTCGAAC
```

Corrected Clone 1-19

---

B.

```
GTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGACCCGTCCCTGGAGCCAGTGGATTT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GTGTCAGATGCAACGCGAGTGAGGGTGGCCTGGGGACCCGTCCCTGGAGCCAGTGGATTT

CGGATTAGCTGGAGCACAGGCAGTGGTCCGGAGTCCAGCCAGACACTGCCCCCAGACTCT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CGGATTAGCTGGAGCACAGGCAGTGGTCCGGAGTCCAGCCAGACACTGCCCCCAGACTCT

ACTGCCACAGACATCACAGGGCTGCAGCCTGGAACCACCTACCAGGTGGCTGTGTCGGTA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACTGCCACAGACATCACAGGGCTGCAGCCTGGAACCACCTACCAGGTGGCTGTGTCGGTA

CTGCGAGGCAGAGAGGAGGGACCTGCTGCAGTCATCGTGGCTCGAAC
||||||||||||||||||||||||||||||||||||||||||||||
CTGCGAGGCAGAGAGGAGGGACCTGCTGCAGTCATCGTGGCTCGAAC
```

Corrected Clone 1-21

SUBSTITUTE SHEET (RULE 26)

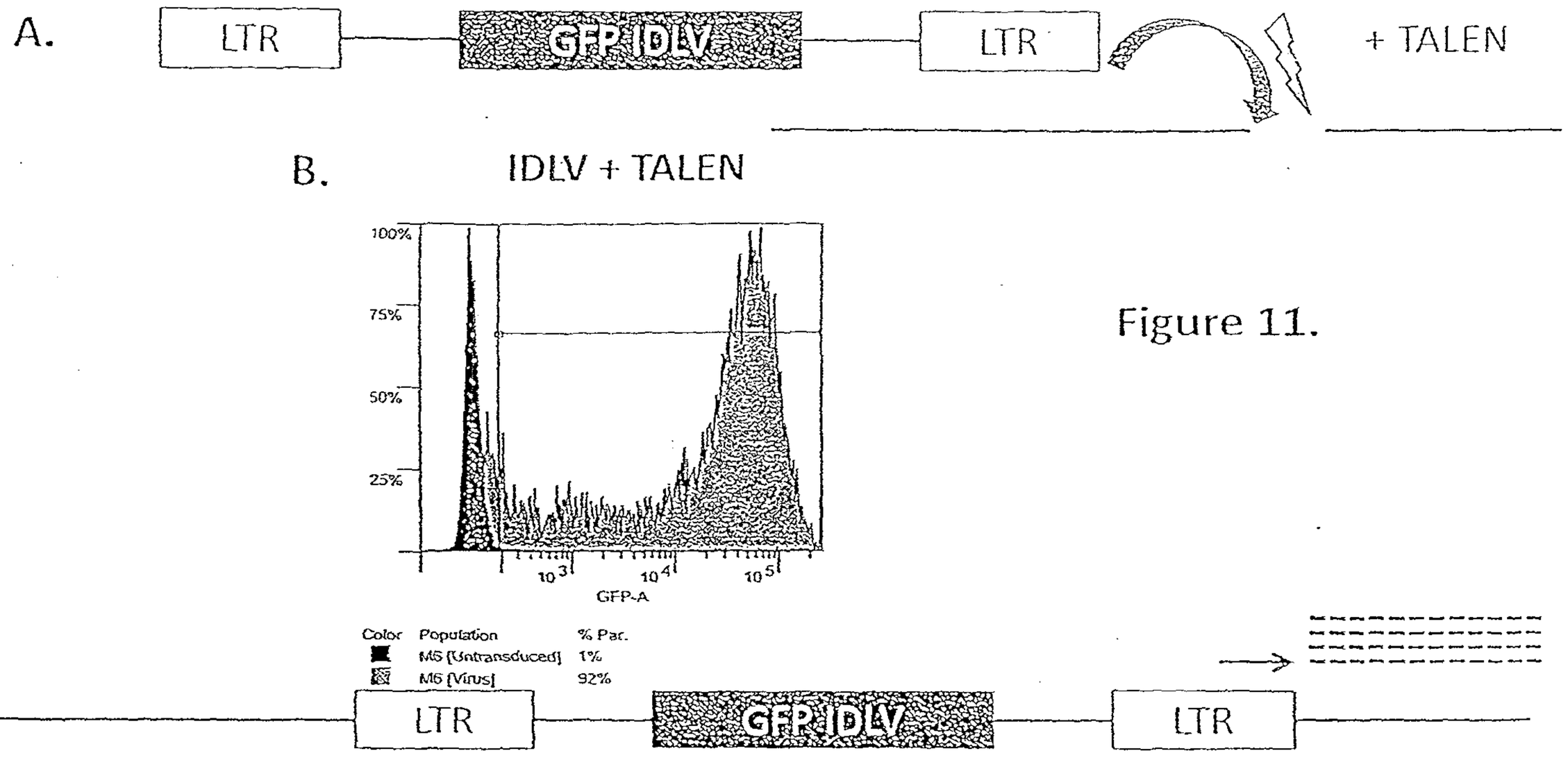



Figure 11.

D.

| Site | Sequence | Locus Information |
|---|---|---|
| 3p21.31 | AGTACTCAGGAGGCTTGTGGAGGACAGCTGC | TALEN Target |
| 7p22.3 | GATGCCGCCCCACCCCGCACCTCAGCCTTCTCA | Non-coding |
| 16p13.3 | GTGCAGGGTAATGGGAGGGCTTGCACGACG | Intron |
| 1q23.3 | CCAGATTTCACCTTTTAACCTTGGTGCTATATAG | Non-coding |
| 5q33.1 | CACCTTGGGAGTTAGGATTTTAACATGGATTTA | Non-coding |

SUBSTITUTE SHEET (RULE 26)

Figure 12

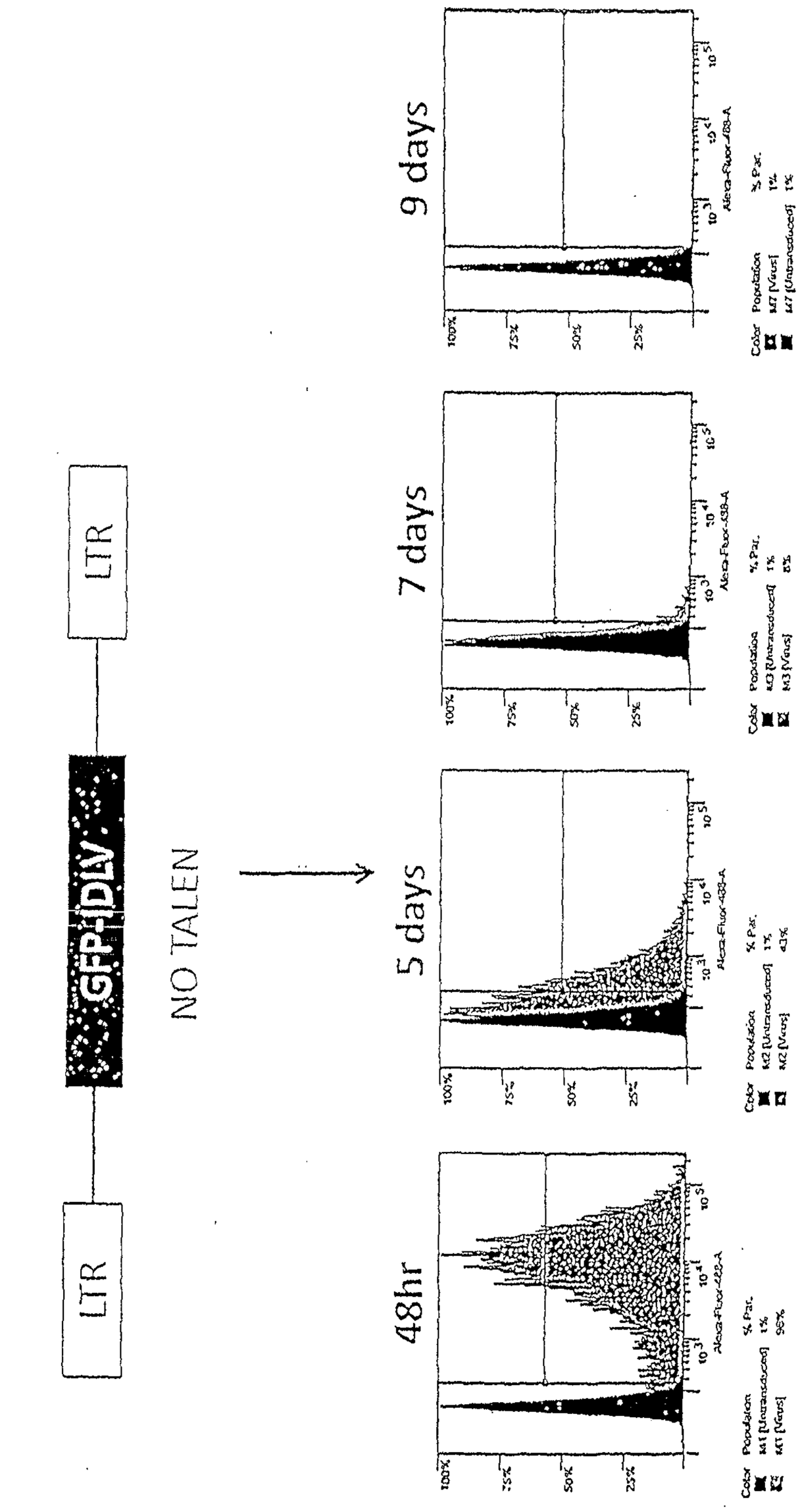


Figure 13

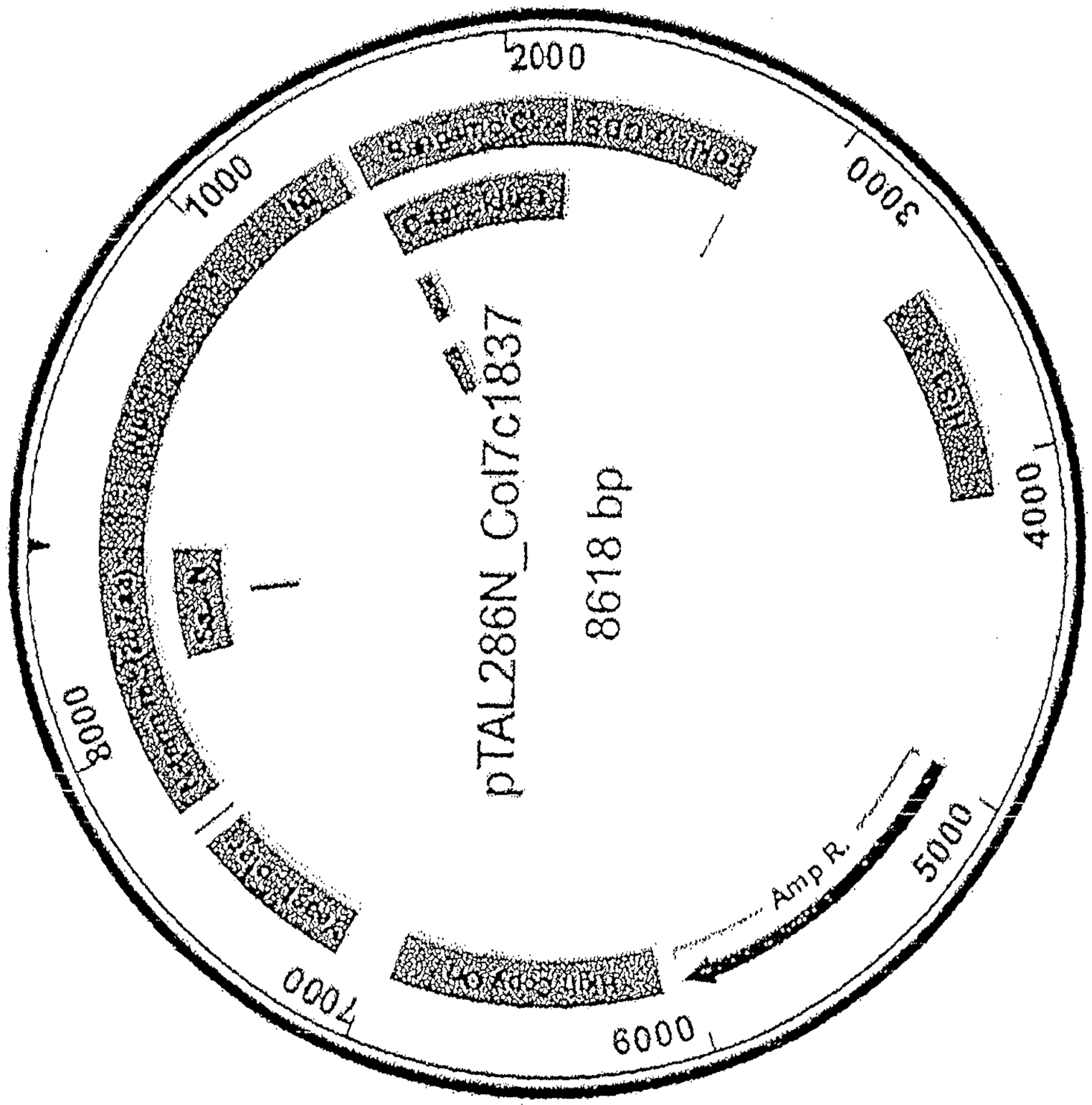
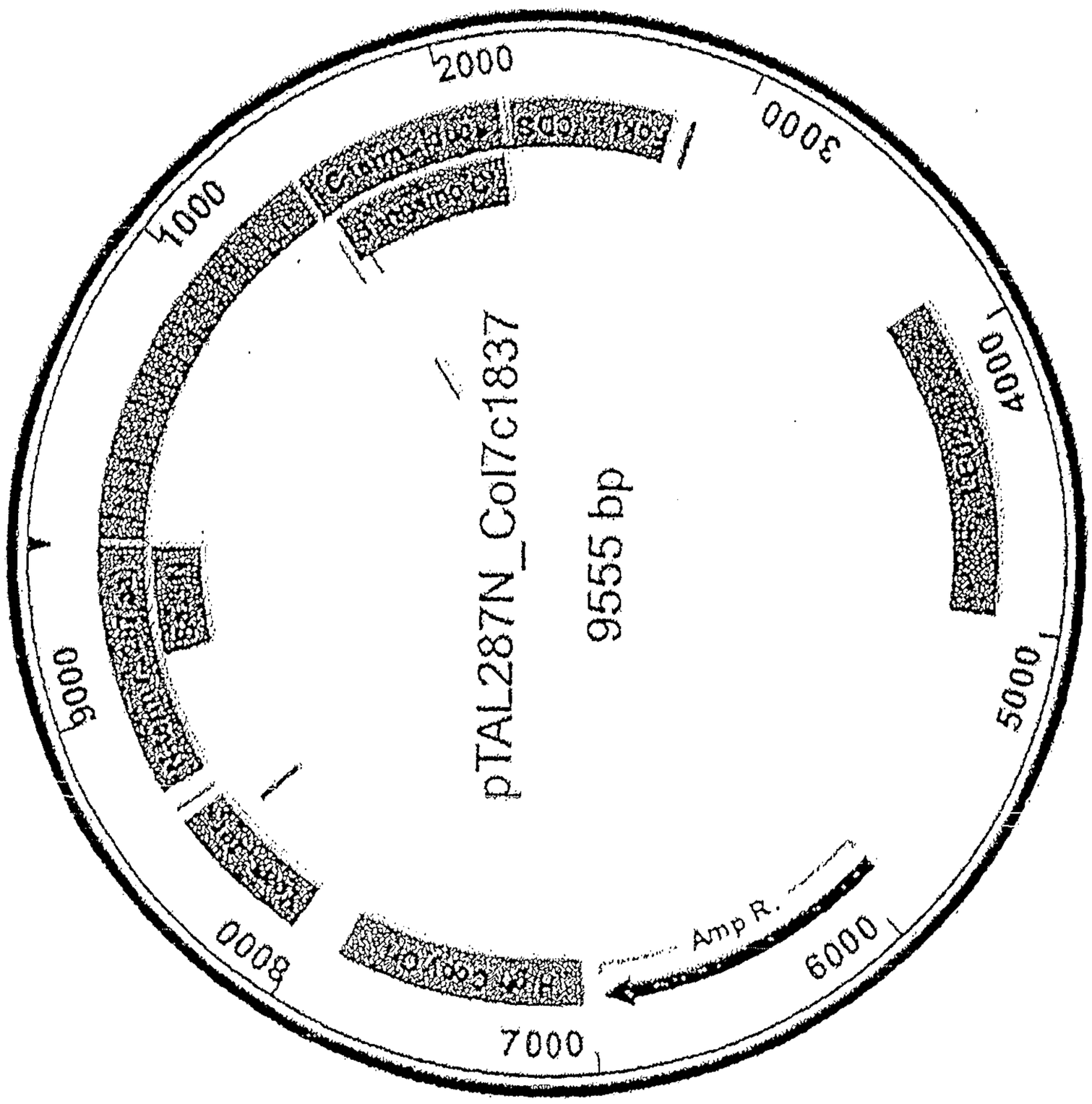

2000
3000
4000
5000
6000
7000
8000
1000
pTAL286N_Col7c1837
8618 bp
Amp R.

Figure 14


pTAL287N_Col7c1837
9555 bp
Amp R.
1000
2000
3000
4000
5000
6000
7000
8000
9000