Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3037635 A1 2018/03/29

(21) 3 037 635

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2017/09/21
(87) Date publication PCT/PCT Publication Date: 2018/03/29
(85) Entrée phase nationale/National Entry: 2019/03/20
(86) N° demande PCT/PCT Application No.: CA 2017/051109
(87) N° publication PCT/PCT Publication No.: 2018/053632
(30) Priorités/Priorities: 2016/09/23 (CAPCT/CA2016/051117); 2017/03/22 (US62/474,827)

(51) Cl.Int./Int.Cl. *C12N 15/12* (2006.01), *A61K 48/00* (2006.01), *A61P 21/00* (2006.01), *C07K 14/47* (2006.01), *C12N 15/09* (2006.01), *C12N 15/11* (2006.01), *C12N 15/86* (2006.01), *C12N 9/22* (2006.01)

(71) Demandeur/Applicant:
UNIVERSITE LAVAL, CA

(72) Inventeurs/Inventors:
TREMBLAY, JACQUES P., CA;
IYOMBE-ENGEMBE, JEAN-PAUL, CA;
CHAPDELAINE, PIERRE, CA;
AGUDELO, DANIEL, CA;
DUCHENE, BENJAMIN, FR

(74) Agent: LAVERY, DE BILLY, LLP

(54) Titre : PROCEDES DE MODIFICATION DU GENE DE LA DYSTROPHINE ET DE RESTAURATION DE L'EXPRESSION DE LA DYSTROPHINE ET LEURS UTILISATIONS

(54) Title: METHODS OF MODIFYING THE DYSTROPHIN GENE AND RESTORING DYSTROPHIN EXPRESSION AND USES THEREOF

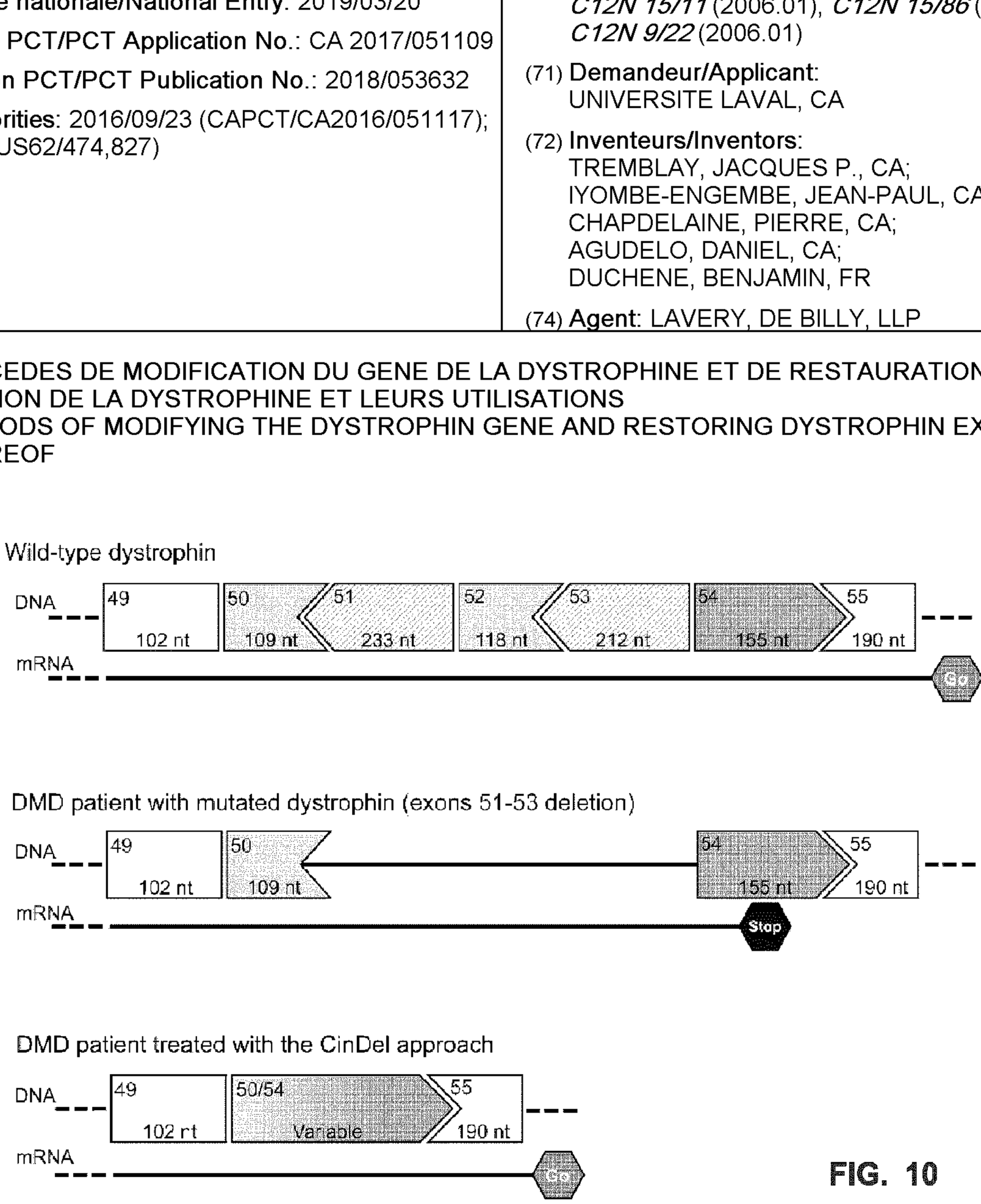


FIG. 10

(57) Abrégé/Abstract:
Methods for modifying a dystrophin gene are disclosed, for restoring dystrophin expression within a cell having an endogenous frameshift or nonsense mutation within the dystrophin gene. The methods comprise introducing a first cut within an exon or intron of the dystrophin gene creating a first exon end or intron end, wherein said first cut is located upstream of the endogenous frameshift or nonsense mutation; and introducing a second cut within an exon or intron of the dystrophin gene creating a second exon end or intron end, wherein said second cut is located downstream of the frameshift or nonsense mutation. Upon joining/ligation of said first and second exon ends or intron ends a hybrid exon or intron junction is created and dystrophin expression is restored, as the correct reading frame is restored. Reagents and uses of the method are also disclosed, for example to treat a subject suffering from muscular dystrophy.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2018 (29.03.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/053632 A1

(51) International Patent Classification:
C12N 15/12 (2006.01) *C12N 15/09* (2006.01)
A61K 48/00 (2006.01) *C12N 15/11* (2006.01)
A61P 21/00 (2006.01) *C12N 15/86* (2006.01)
C07K 14/47 (2006.01) *C12N 9/22* (2006.01)

(21) International Application Number:
PCT/CA2017/051109

(22) International Filing Date:
21 September 2017 (21.09.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CA2016/051117
23 September 2016 (23.09.2016) CA
62/474,827 22 March 2017 (22.03.2017) US

(71) Applicant: **UNIVERSITÉ LAVAL** [CA/CA]; Pavillon des Sciences de l'Éducation, 2320 rue des Bibliothèques, local 1434, Québec, Québec G1V 0A6 (CA).

(72) Inventors: **TREMBLAY, Jacques P.**; 1200 rue Sauvé, apt 837, Québec, Québec G1W 0G2 (CA). **LYOMBE-ENGEMBE, Jean-Paul**; 2255 rue de l'Université, Pavillon Parent, Chambre 3706, Québec, Québec G1V 0A7 (CA). **CHAPDELAINE, Pierre**; 180 rue Saint-Robert, Saint-Romuald, Québec G6W 3C4 (CA). **AGUDELO, Daniel**; 3178 rue de Namur, appartement 5, Québec, Québec G1W 2V2 (CA). **DUCHÊNE, Benjamin**; 3 rue du Champ Gayot, 35410 Chateaugiron (FR).

(74) Agent: **GOUDREAU GAGE DUBUC**; 2000, avenue McGill College, #2200, Montréal, Québec H3A 3H3 (CA).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

(54) Title: METHODS OF MODIFYING THE DYSTROPHIN GENE AND RESTORING DYSTROPHIN EXPRESSION AND USES THEREOF

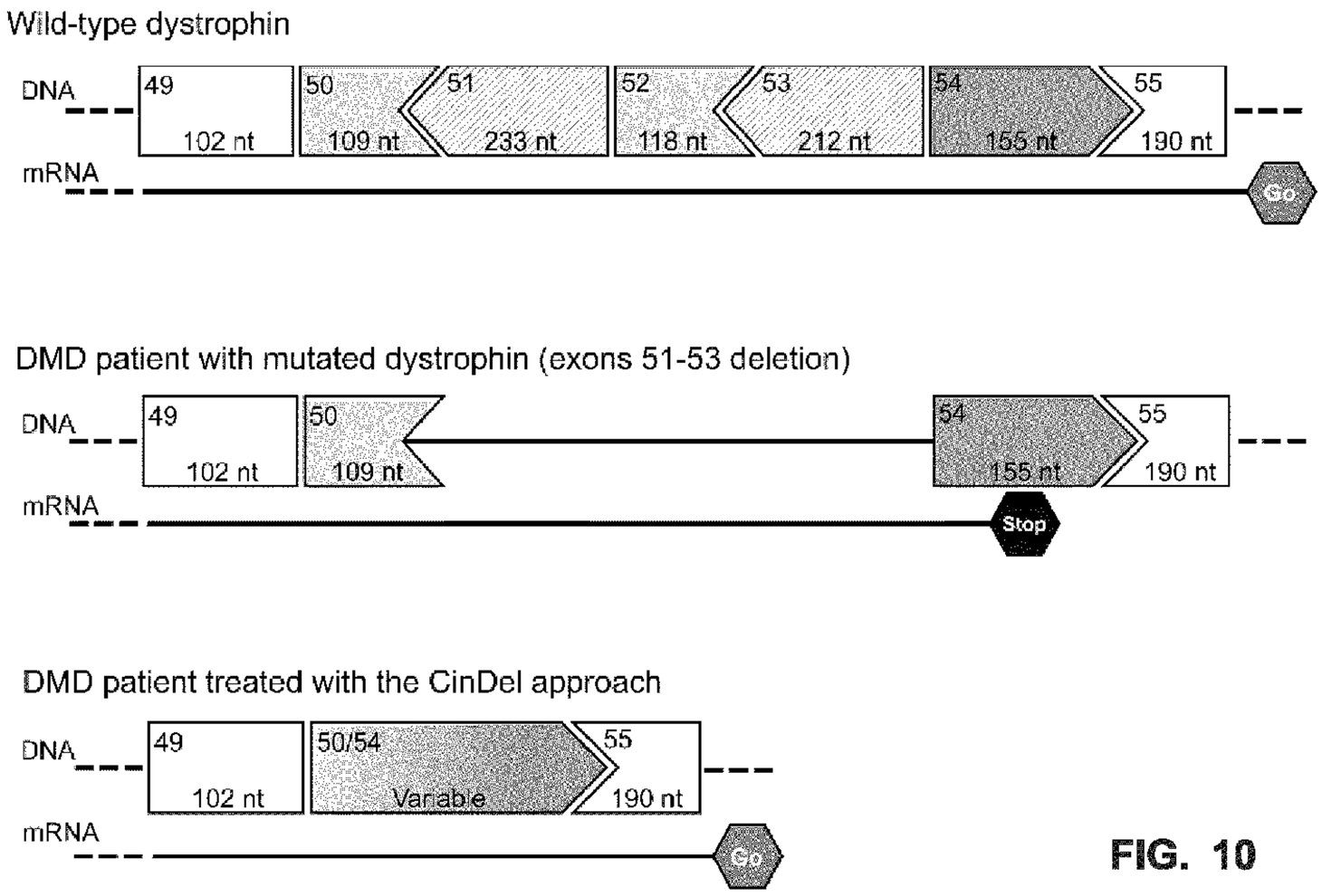


FIG. 10

(57) Abstract: Methods for modifying a dystrophin gene are disclosed, for restoring dystrophin expression within a cell having an endogenous frameshift or nonsense mutation within the dystrophin gene. The methods comprise introducing a first cut within an exon or intron of the dystrophin gene creating a first exon end or intron end, wherein said first cut is located upstream of the endogenous frameshift or nonsense mutation; and introducing a second cut within an exon or intron of the dystrophin gene creating a second exon end or intron end, wherein said second cut is located downstream of the frameshift or nonsense mutation. Upon joining/ligation of said first and second exon ends or intron ends a hybrid exon or intron junction is created and dystrophin expression is restored, as the correct reading frame is restored. Reagents and uses of the method are also disclosed, for example to treat a subject suffering from muscular dystrophy.

[Continued on next page]

WO 2018/053632 A1

WO 2018/053632 A1

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, 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))*

TITLE OF THE INVENTION

METHODS OF MODIFYING THE DYSTROPHIN GENE AND RESTORING DYSTROPHIN EXPRESSION AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority, of PCT Application No. PCT CA2016/051117 filed on September 23, 2016 and of US provisional application serial No. 62/474,827, filed on March 22, 2017, which are incorporated herein by reference in their entirety.

SEQUENCE LISTING

[0002] This application contains a Sequence Listing in computer readable form entitled "11229_375_SL_ST25.txt", created on September 19, 2017 and having a size of about 155 KB. The computer readable form is incorporated herein by reference.

FIELD OF THE INVENTION

[0003] The present invention relates to the targeted modification of an endogenous mutated dystrophin gene to restore dystrophin expression in mutated cells, such as cells of subjects suffering from Muscular Dystrophy (MD), such as Duchenne MD (DMD) and Becker MD (BMD). More specifically, the present invention is concerned with correcting the reading frame of a mutated dystrophin gene by targeting exon or intron sequences close to the endogenous mutation. The present invention also relates to such modified forms of dystrophin.

BACKGROUND OF THE INVENTION

[0004] Duchenne Muscular Dystrophy (DMD) is a monogenic hereditary disease linked to the X chromosome, which affects one in about 3500 male births [1]. The cause of the disease is the inability of the body to synthesize the dystrophin (DYS) protein, which plays a fundamental role in maintaining the integrity of the sarcolemma [2, 3]. The absence of this protein is secondary to a mutation of the *DYS* gene [4]. The most frequently encountered mutations, found in over 60% of DMD patients, are deletions of one or more exons in the region between exons 45 and 55, called the hot region of *DYS* gene [5]. Most of these deletions induce a codon frame-shift of the mRNA transcript leading to the production of a truncated DYS protein. Since the latter is rapidly degraded, the absence of DYS at the sarcolemma increases its fragility and leads to muscle weakness characteristic of DMD. In some cases deletions result in the milder Becker Muscular Dystrophy (BMD) phenotype [6]. For DMD patients, skeletal muscular weaknesses will unfortunately lead to death, between 18 and 30 years of age [7, 8], while some BMD patients can have a normal life expectancy [6]. To date, there is no cure for DMD and BMD.

[0005] The identification of the molecular basis for the DMD and BMD phenotypes established the foundation for DMD gene therapy [9-13]. Different strategies for DMD gene therapy are currently under development. Since the 2.4-Mb *DYS* gene contains 79 exons and encodes a 14 kb mRNA [14, 15], it is difficult to develop a gene therapy to

deliver efficiently the full-length gene or even its cDNA in muscle precursor cells *in vitro* or in muscle fibers *in vivo*.

[0006] An alternative to gene replacement is to modify the DYS mRNA or the DYS gene itself directly within cells. Correction of the reading frame of the mRNA can be obtained by exon skipping using a synthetic antisense oligonucleotide (AON) interacting in with the primary transcript with the splice donor or spice acceptor of the exon, which precedes or follows the patient deletion [20-28]. Unfortunately, this therapeutic approach is facing a number of difficulties associated with the lifetime use of AONs [29]. Further, the AONs act only on the mRNA, thus the DMD patients treated with this approach are required to receive this treatment for life, which is very expensive and increases the risks of complications.

[0007] Thus, there remains a need for novel therapeutic approaches for restoring dystrophin expression in cells.

[0008] The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

[0009] The present invention relates to restoring the correct reading frame of a mutant DYS gene, which may be used as a new therapeutic approach for treating muscular dystrophy (MD) (e.g., DMD). This can be done directly on the cells of a subject suffering from MD. This approach is based on the permanent restoration of the *DYS* reading frame by generating additional mutations (e.g., deletion) upstream and downstream of an endogenous frameshift or nonsense mutation. These engineered upstream and downstream mutations may be within an exon containing the endogenous frameshift or nonsense mutation, and/or may be within exons or introns flanking the endogenous frameshift or nonsense mutation (e.g., exons or introns upstream and downstream from the frameshift or nonsense mutation). In a first aspect, by targeting exons (as opposed to introns) as the sites to introduce these engineered mutations, it is possible to restore the reading frame of the *DYS* gene in cells to produce a modified dystrophin protein having the smallest possible deletion while still retaining sufficient dystrophin protein function. Alternatively, an entire exon comprising an endogenous frameshift or nonsense mutation the may be deleted to restore the dystrophin reading frame. Notably, by specifically selecting the deletion to be introduced into the dystrophin gene comprising the endogenous frameshift or nonsense mutation, applicants were able to maintain the configuration of normal spectrin-like repeats in the modified dystrophin protein where hydrophobic amino acids are localized in position “a” and “d” of the heptad motif, thereby generating a functional (albeit shorter) dystrophin protein.

[0010] For the treatment of MD such as DMD, agents for introducing gene modifications need to be delivered effectively into cells for the treatment to be efficient. AAV vectors are currently the vehicle of choice for delivery of gene modifying components into cells. However, sustained expression of the CRISPR nuclease (e.g., Cas9 nuclease) may increase off-target mutations. In an effort to identify effective alternative delivery methods, Applicants explored various methods/agents for carrying sgRNAs/CRISPR nuclease combination into cells. It was found that ribonucleoprotein sgRNA/Cas9 complex can be effectively transduced *in vitro* in myoblasts and

in vivo in muscle fibers using different delivery agents, including new cell penetrating peptides called Feldan Shuttles (FS). This delivery method reduces the life-time of the CRISPR nuclease in cells and may thus be used to reduce off-target mutations making the treatment more secure. Furthermore, it was found that among the polypeptide-based shuttles tested, some carriers were surprisingly more effective than others.

[0011] Accordingly, in accordance with the present invention, there is provided a method of modifying a dystrophin gene and restoring the correct reading frame for dystrophin expression within a cell having an endogenous frameshift or nonsense mutation within the dystrophin (DYS) gene, the method comprising:

a) introducing a first cut within an exon or intron of the DYS gene creating a first exon end or intron end, wherein said first cut is located upstream of the endogenous frameshift or nonsense mutation;

b) introducing a second cut within an exon or an intron of the DYS gene creating a second exon end or intron end, wherein said second cut is located downstream of the frameshift or nonsense mutation;

[0012] wherein upon ligation of said first and second exon ends or said first and second intron ends, a modified dystrophin gene comprising a hybrid exon or a hybrid intron is created and dystrophin expression is restored.

[0013] In an embodiment, the first and second cuts are within one or more exons, and are not within an intron, of the dystrophin gene (although a gRNA or a portion thereof may bind to an intron, in particular in an intronic region flanking an exon, as long as the resulting cut is in an exon). As a result, following the introduction of the first and second cuts, the first exon end is ultimately joined or ligated to the second exon end, creating a hybrid, fusion exon and at the same time restoring the correct reading frame, allowing transcription to the end of the dystrophin gene, producing a truncated dystrophin protein (at least lacking the portion comprising the endogenous frameshift or nonsense mutation) due to the removal of a portion of the gene by the first and second cuts (e.g., a first cut in the first exon upstream of the endogenous mutation and a second cut in the second exon downstream of the endogenous mutation). Depending on the site of the first and second cuts, ligation of exon ends may lead to the introduction of a new codon in the amino acid sequence of the dystrophin protein. Preferably, the location of the cuts to be introduced are specifically selected such that the configuration of spectrin-like repeats in the modified dystrophin protein is maintained (i.e., hydrophobic amino acids in the spectrin-like repeats are localized in position “a” and “d” of the heptad motif, as known in the art; see for example FIG. 16).

[0014] In another embodiment, the first and second cuts are within introns flanking the endogenous frameshift or nonsense mutation of the dystrophin gene (although a gRNA or a portion thereof may bind to an exon, in particular in an exonic region flanking an intron, as long as the resulting cut is in an intron). As a result, following the introduction of the first and second cuts, the first intron end is ultimately joined or ligated to the second intron end, creating a hybrid, fusion intron, deleting the exon(s) located between the introns and at the

same time restoring the correct reading frame, allowing transcription to the end of the dystrophin gene, producing a truncated dystrophin protein (at least lacking the portion comprising the endogenous frameshift or nonsense mutation) due to the removal of a portion of the gene by the first and second cuts (e.g., a first cut in the first intron upstream of the endogenous mutation and a second cut in the second intron downstream of the endogenous mutation). Preferably, the location of the cuts to be introduced are specifically selected such that the configuration of spectrin-like repeats in the modified dystrophin protein is maintained (i.e., hydrophobic amino acids in the spectrin-like repeats are localized in position “a” and “d” of the heptad motif, as known in the art; see for example FIG. 16). In embodiments, one or more complete spectrin-like repeats (R1, R2, R3, R4….R15, R16, R17’ R18, R19, R20, R21, R22, R23, R24 or any combination thereof) are removed by the deletion of one or more exons.

[0015] In an embodiment, said first and second cuts are introduced by providing a cell with i) a CRISPR nuclease (e.g., Cas9 nuclease); and ii) a pair of gRNAs consisting of a) a first gRNA which binds to an exon or intron sequence of the DYS gene located upstream of the endogenous frameshift or nonsense mutation for introducing a first cut; b) a second gRNA which binds to an exon or intron sequence of the DYS gene located downstream of the endogenous frameshift or nonsense mutation for introducing the second cut.

[0016] In an embodiment, the endogenous frameshift or nonsense mutation is located in one or more exons selected from exons 45-58 of the dystrophin gene.

[0017] In embodiments, the first cut is within exon 45, 46, 47, 48, 49 or 50 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0018] In embodiments, the first cut is within exon 45 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0019] In embodiments, the first cut is within exon 46 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0020] In embodiments, the first cut is within exon 47 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0021] In embodiments, the first cut is within exon 48 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0022] In embodiments, the first cut is within exon 49 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0023] In embodiments, the second cut is within exon 51 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0024] In embodiments, the second cut is within exon 52 and the first cut is within exon 45, 46, 47, 48 or 49,

of the dystrophin gene.

[0025] In embodiments, the second cut is within exon 53 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0026] In embodiments, the second cut is within exon 54 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0027] In embodiments, the second cut is within exon 55 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0028] In embodiments, the second cut is within exon 56 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0029] In embodiments, the second cut is within exon 57 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0030] In embodiments, the second cut is within exon 58 and the first cut is within exon 45, 46, 47, 48 or 49, of the dystrophin gene.

[0031] In an embodiment, the first cut is within exon 50 and the second cut is within exon 54, of the dystrophin gene.

[0032] In an embodiment, the first cut is within exon 46 and the second cut is within exon 51, of the dystrophin gene.

[0033] In an embodiment, the first cut is within exon 46 and the second cut is within exon 53, of the dystrophin gene.

[0034] In an embodiment, the first cut is within exon 47 and the second cut is within exon 52, of the dystrophin gene.

[0035] In an embodiment, the first cut is within exon 49 and the second cut is within exon 52, of the dystrophin gene.

[0036] In an embodiment, the first cut is within exon 49 and the second cut is within exon 53, of the dystrophin gene.

[0037] In an embodiment, the first cut is within exon 47 and the second cut is within exon 58, of the dystrophin gene.

[0038] In embodiments, the first cut is within intron 45-58 and the second cut is within intron 46-59, of the dystrophin gene.

[0039] In embodiments, the first cut is within intron 45, 46, 47, 48, 49 or 50 and the second cut is within intron 51, 52, 53, 54, 55, 56, 57, 58 or 59, of the dystrophin gene.

[0040] In embodiments, the first cut is within intron 45 and the second cut is within intron 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0041] In embodiments, the first cut is within intron 46 and the second cut is within intron 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0042] In embodiments, the first cut is within intron 47 and the second cut is within intron 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0043] In embodiments, the first cut is within intron 48 and the second cut is within intron 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0044] In embodiments, the first cut is within intron 49 and the second cut is within intron 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene.

[0045] In embodiments, the second cut is within intron 51 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0046] In embodiments, the second cut is within intron 52 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0047] In embodiments, the second cut is within intron 53 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0048] In embodiments, the second cut is within intron 54 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0049] In embodiments, the second cut is within intron 55 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0050] In embodiments, the second cut is within intron 56 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0051] In embodiments, the second cut is within intron 57 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0052] In embodiments, the second cut is within intron 58 and the first cut is within intron 45, 46, 47, 48 or 49, of the dystrophin gene.

[0053] In an embodiment, the first cut is within intron 22 and the second cut is within intron 23, of the dystrophin gene.

[0054] In embodiments, the first and second cuts generate a hybrid exon. In embodiments. the first and second cuts generate a hybrid intron.

[0055] In an embodiment, the hybrid exon generated by the method of the present invention has a nucleic

acid sequence as set forth in FIG. 5a, FIG. 9a(ii) or FIG. 17. In an embodiment, the first and second gRNAs (the gRNA pair), together with a CRISPR nuclease (e.g., SaCas9 or SpCas9), allow to generate a hybrid exon from (i) exon 46 and exon 51 (hybrid exon 46-51); (ii) exon 46 and exon 53 (hybrid exon 46-53); (iii) exon 47 and exon 52 (hybrid exon 47-52); (iv) exon 49 and exon 52 (hybrid exon 49-52); (v) exon 49 and exon 53 (hybrid exon 49-53); or (vi) exon 47 and exon 58 (hybrid exon 47-58).

[0056] In an embodiment, the above-noted method generates a modified dystrophin gene encoding a modified dystrophin protein comprising a hybrid spectrin-like repeat (SLR) comprising a portion of a first SLR and a portion of second SLR and a hybrid SLR junction. In the hybrid SLR, the normal (wild-type) configuration of the SLR is maintained where hydrophobic amino acids are localized in position “a” and “d” of the heptad motif.

[0057] In an embodiment, the modified dystrophin protein generated by the method of the present invention comprises the following hybrid spectrin-like repeat (SLR): (i) a hybrid SLR comprising a portion of SLR 17 and a portion of SLR 19 (SLR 17-19); (ii) a hybrid SLR comprising a portion of SLR 17 and a portion of SLR 20 (SLR 17-20); (iii) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 20 (SLR 18-20); (iv) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 21 (SLR 18-21); (v) a hybrid SLR comprising a portion of SLR 19 and a portion of SLR 20 (SLR 19-20); or (vi) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 23 (SLR 18-23).

[0058] In an embodiment, the hybrid SLR has a hybrid SLR junction selected from the hybrid SLR junctions set forth in FIGs. 5a, 5b and 16b.

[0059] CRISPR nucleases and gRNAs of the present invention may be introduced into the cells using any useful methods known in the art.

[0060] In an embodiment, the first gRNA, the second gRNA and/or the CRISPR nuclease are delivered into the cell using one or more adeno-associated virus (AAV) vectors.

[0061] In another embodiment the first gRNA, the second gRNA and/or the CRISPR nuclease are delivered into the cell using one or more polypeptide-based shuttles (e.g., Feldan Shuttle).

[0062] In an embodiment, the one or more polypeptide-based shuttle comprises: (i) a polypeptide-based shuttle having amino acid sequence HHHHHHKWKLFKKIGAVLKVLTTGYARAAARQARA (FSA, SEQ ID NO: 196); (ii) a polypeptide-based shuttle having amino acid sequence HHHHHHKWKLFKKIGAVLKVLTTGYARAAARQARAHHHHHH (FSB, SEQ ID NO: 197); (iii) a polypeptide-based shuttle having amino acid sequence HHHHHHLLKLWSRLLKLWTQGRRLKAKRAKAHHHHHH (FSC, SEQ ID NO: 198); or (iv) any combination of at least two of (i), (ii) and (iii).

[0063] In an embodiment, the CRISPR nuclease used in accordance with the present invention is a Cas9 nuclease derived from *Staphylococcus aureus* (saCas9) or from *Streptococcus pyogenes* (SpCas9).

[0064] Also provided is a gRNA pair for use in the method of the present invention (e.g., for restoring dystrophin expression in a cell comprising an endogenous frameshift or nonsense mutation within the dystrophin (DYS) gene, or for treating muscular dystrophy). In an embodiment, the pair consists of a first gRNA and a second gRNA, wherein the first gRNA comprises a first target sequence upstream of the endogenous frameshift or nonsense mutation (i.e., the gRNA binds to the complementary strand (the opposite strand) of the first target sequence) and can direct a nuclease-mediated first cut in an exon or intron sequence of the DYS gene located upstream of the endogenous frameshift or nonsense mutation and wherein the second gRNA comprises a second target sequence downstream of the endogenous frameshift or nonsense mutation (i.e., it binds to the complementary sequence (the opposite strand) of the second target sequence) and can direct a nuclease-mediated second cut in an exon or intron sequence of the DYS gene located downstream of the endogenous frameshift or nonsense mutation.

[0065] In an embodiment, each of the first target sequence and the second target sequence of the first and second gRNAs is adjoining a PAM sequence in the DYS gene set forth in Table 3, Table 6, Table 8, FIG. 1 or FIG. 11. In embodiments, the target sequence does not consist of the target sequence of gRNAs 20, 22, 23 or 26 set forth in Table 8.

[0066] In embodiments, the target sequence of the first gRNA and/or second gRNA spans an intron-exon junction (i.e., the target sequence comprises both intronic and exonic sequences).

[0067] In an embodiment, each of the first target sequence and the second target sequence of the first and second gRNAs comprises or consists of a target sequence selected from the target sequences set forth in listed in Table 3, 6 or 8 or shown in FIG. 11.

[0068] In an embodiment, the first and second target sequences are each independently 10-40 nucleotides in length.

[0069] In an embodiment, the gRNA pair is selected from a gRNA pair set forth in FIG. 4, 13, 15, 17, 21 or 22 or Table 4 or 7. In embodiments, the gRNA pair is not selected from gRNAs 20 and 23 or 20 and 22 set forth in Table 8.

[0070] In an embodiment, the first gRNA and the second gRNA in the gRNA pair are selected from the gRNAs listed in Table 3, 6 or 8 or shown in FIG. 11. In an embodiment, the first gRNA or the second gRNA in the gRNA pair does not have a target sequence consisting of the target sequence of gRNA 20, 22, 23 or 26 set forth in Table 8. In an embodiment, the first gRNA of the gRNA pair has the target sequence which comprises prises or consists of the sequence AGATCTGAGCTCTGAGTGGA (SEQ ID NO: 83).

[0071] In an embodiment, the first gRNA of the gRNA pair has the target sequence which comprises prises or consists of the sequence GTCTGTTTCAGTTACTGGTGG (SEQ ID NO: 108).

[0072] In an embodiment, the first gRNA of the gRNA pair has the target sequence which comprises prises or

consists of the sequence CTTATGGGAGCACTTACAAGC (SEQ ID NO: 110)

[0073] In an embodiment, the second gRNA of the gRNA pair has the target sequence which comprises prises or consists of the sequence GTGGCAGACAAATGTAGATG (SEQ ID NO: 93).

[0074] In an embodiment, the second gRNA of the gRNA pair has the target sequence which comprises prises or consists of the sequence TCATTTCACAGGCCTTCAAGA (SEQ ID NO: 121).

[0075] In an embodiment, the second gRNA of the gRNA pair has the target sequence which comprises prises or consists of the sequence CAATTACCTCTGGGCTCCTGG (SEQ ID NO: 123).

[0076] In an embodiment, the gRNA pair consists of a first gRNA having a first target sequence which is GTCTGTTTCAGTTACTGGTGG (SEQ ID NO: 108) and a second gRNA having a second target sequence which is TCATTTCACAGGCCTTCAAGA (SEQ ID NO: 121).

[0077] In an embodiment, the gRNA pair consists of a first gRNA having a first target sequence which is CTTATGGGAGCACTTACAAGC (SEQ ID NO: 110) and a second gRNA having a second target sequence which is CAATTACCTCTGGGCTCCTGG (SEQ ID NO: 123).

[0078] In an embodiment, the first gRNA of the gRNA pair has the target sequence which comprises or consists of the sequence ATTTCAGGTAAGCCGAGGTT (SEQ ID NO: 207).

[0079] In an embodiment, the first gRNA of the gRNA pair has the target sequence which comprises or consists of the sequence TCTTAATAATGTTTCACTGT (SEQ ID NO: 208).

[0080] In an embodiment, the second gRNA of the gRNA pair has the target sequence which comprises or consists of the sequence ATAGTTTAAAGGCCAAACCT (SEQ ID NO: 211).

[0081] In an embodiment, the second gRNA of the gRNA pair has the target sequence which comprises or consists of the sequence ATAATTTCTATTATATTACA (SEQ ID NO:209).

[0082] In an embodiment, the second gRNA of the gRNA pair has the target sequence which comprises or consists of the sequence TTTCATTCATATCAAGAAGA (SEQ ID NO: 210).

[0083] In an embodiment, the gRNA pair consists of a first gRNA having a first target sequence which is ATTTCAGGTAAGCCGAGGTT (SEQ ID NO: 207) and a second gRNA having a second target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO:209).

[0084] In an embodiment, the gRNA pair consists of a first gRNA having a first target sequence which is ATTTCAGGTAAGCCGAGGTT (SEQ ID NO: 207) and a second gRNA having a second target sequence which is TTTCATTCATATCAAGAAGA (SEQ ID NO: 210).

[0085] In an embodiment, the gRNA pair consists of the gRNA pair consists of a first gRNA having a first target sequence which is TCTTAATAATGTTTCACTGT (SEQ ID NO: 208) and a second gRNA having a

second target sequence which is TTTCATTCATATCAAGAAGA (SEQ ID NO: 210).

[0086] In an embodiment, the gRNA pair consists of the gRNA pair consists of a first gRNA having a first target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO: 209) and a second gRNA having a second target sequence which is ATAGTTTAAAGGCCAAACCT (SEQ ID NO: 211).

[0087] In an embodiment, the gRNA pair consists of the gRNA pair consists of a first gRNA having a first target sequence which is TTTCATTCATATCAAGAAGA (SEQ ID NO: 210) and a second gRNA having a second target sequence which is ATAGTTTAAAGGCCAAACCT (SEQ ID NO: 211).

[0088] In an embodiment, the gRNA pair consists of the gRNA pair consists of a first gRNA having a first target sequence which is TCTTAATAATGTTTCACTGT (SEQ ID NO: 208) and a second gRNA having a second target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO: 209).

[0089] Also provided is a nucleic acid comprising one or more sequences encoding one or both members of a gRNA pair described herein. In an embodiment, the nucleic acid further comprises a sequence encoding a CRISPR nuclease.

[0090] Also provided is a nucleic acid comprising a modified dystrophin gene comprising ligated first and second exon ends or intron ends as described herein. In an embodiment, the nucleic acid comprises a hybrid exon sequence set forth in FIG. 5a, 9a(ii) or 17. In an embodiment, the nucleic acid comprises a polynucleotide sequence encoding a hybrid spectrin-like repeat (SLR) having a polypeptide sequence set for in FIG. 16b.

[0091] In embodiments, the modified dystrophin gene comprises ligated first and second exon ends defined by the cut sites defined in Table 3or 6. In embodiments, the modified dystrophin gene comprises ligated first and second intron ends defined by the cut sites defined in Table 8. In a further embodiment, the first cut site is between nucleotides 6769 and 6770 of the DYS gene and the second cut site is between nucleotides 8554 and 8555 of the DYS gene. In a further embodiment, the first cut site is between nucleotides 6833 and 6834 of the DYS gene and the second cut site is between nucleotides 8657 and 8658 of the DYS gene. In a further embodiment, the first cut site is between nucleotides 7228 and 7229 of the DYS gene and the second cut site is between nucleotides 7912 and 7913 of the DYS gene.

[0092] Also provided is a modified dystrophin polypeptide encoded by the above-noted nucleic acid.

[0093] Also provided is a vector, comprising a nucleic acid described herein. In an embodiment, the vector is a viral vector (e.g. an AAV or a Sendai virus derived vector).

[0094] Also provided is a cell (e.g. a host cell,) comprising one or both members of a gRNA pair, nucleic acid, polypeptide and/or vector described herein. In embodiments, the host cell may be prokaryotic or eukaryotic. In an embodiment, the cell is a mammalian cell, in a further embodiment, a human cell. In an embodiment, the cell is a muscle cell (e.g. myoblast or myocyte). In an embodiment, the cell is a cell from a subject suffering from muscular dystrophy (e.g., DMD).

[0095] Also provided is a composition, comprising one or both members of a gRNA pair, nucleic acid, polypeptide, vector, and/or cell described herein. In an embodiment, the composition further comprises a CRISPR nuclease or a nucleic acid encoding a CRISPR nuclease. In an embodiment, the composition further comprises a biologically or pharmaceutically acceptable carrier. In an embodiment, the composition is for a use described herein.

[0096] Also provided is a kit, comprising one or both members of a gRNA pair, nucleic acid, polypeptide, vector, cell, composition, CRISPR nuclease and/or a nucleic acid encoding a CRISPR nuclease, described herein. In an embodiment, the kit further comprises instructions for performing a method described herein, or is for a use described herein.

[0097] In an embodiment, the kit is for use in treating muscular dystrophy in a subject in need thereof.

[0098] Also provided is a method for treating muscular dystrophy in a subject, comprising modifying a dystrophin gene and restoring the correct reading frame for dystrophin expression within a cell of said subject according to a method described herein.

[0099] Also provided is a method for treating muscular dystrophy in a subject, comprising contacting a cell of the subject with (i)(a) a gRNA pair described herein or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) a composition described herein.

[0100] Also provided is a use of (i)(a) a gRNA pair described herein or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) a composition described herein, for treating muscular dystrophy in a subject.

[0101] Also provided is a use of (i)(a) a gRNA pair described herein or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) a composition described herein, for the preparation of a medicament for treating muscular dystrophy in a subject.

[0102] Also provided is (i)(a) a gRNA pair described herein or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) a composition described herein, for use in treating muscular dystrophy in a subject.

[0103] Also provided is (i)(a) a gRNA pair described herein or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) a composition described herein, for use in the preparation of a medicament for treating muscular dystrophy in a subject.

[0104] In an embodiment, the muscular dystrophy is Duchenne muscular dystrophy.

[0105] Also provided is a reaction mixture comprising (a) a gRNA pair described herein or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR

nuclease polypeptide.

[0106] Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0107] In the appended drawings:

[0108] FIG. 1 shows a plasmid used in this study and protospacer adjacent motif (PAM) sites. (a) The expression vector pSpCas(BB)-2A-GFP contains 2 BbsI sites for the insertion of the protospacer sequence. The guide RNA is under the control of the U6 promoter. Guide RNAs were designed following the identification of PAMs (i.e., NGG sequence) in exons 50 (b) and 54(c) of the *DYS* gene. The FIG. illustrates the sequence of exons 50 (b) and 54 (c) of the human *DYS* gene. For exon 50, 10 different PAMs (numbered 1 to 10) were identified; six are in the sense strand and 4 in the antisense strand. For exon 54, 14 PAMs were identified, 5 in the sense strand and 9 in the antisense strand. The GG's of the PAM are shaded in the sense (upper) and antisense (lower) strands. The third nucleotide of the PAMs (i.e. adjacent to the GG's) is also shaded in both strands. See Tables 3, 6 and 8 for exemplary gRNAs targeting sequences adjoining these PAMs;

[0109] FIG. 2 shows Transfection efficiency of constructs prepared in accordance with an embodiment of the present invention. The eGFP expression was monitored in 293T (a) and in DMD myoblasts (b and c) after transfection of the pSpCas(BB)-2A-GFP with Lipofectamine 2000. Transfection efficiency was increased in DMD myoblasts following a modification of the transfection protocol with Lipofectamine 2000 (c vs b);

[0110] FIG. 3 shows a Surveyor assay for gRNA screening in 293T cells and in myoblasts. The assay was performed on genomic DNA extracted from 293T cells (a and b) or myoblasts (c and d) transfected individually with different gRNAs targeting exons 50 and 54. Screening was performed separately for exon 50 (a and c) and exon 54 (b and d). Genomic DNA of non-transfected cells was used for negative control (NC) for the Surveyor assay. The gRNA numbers correspond with the targeted sequences (Table 1). MW: molecular weight marker;

[0111] FIG. 4 shows that the CinDel approach can generate four possible *DYS* gene modifications. (a) Double-strand breaks created by the Cas9 and different gRNA pairs can theoretically modify the *DYS* gene four different ways: 1) in light grey (shaded cells of columns 1 and 5), correct junction of the normal codons of exons 50 and 54; 2) in darker grey (shaded cells of columns 2-4, 6-9, 11, 13 and 14, and shaded cells in rows 3 and 4 of columns 10 and 12) the junction of the nucleotides of exons 50 and 54 generates the codon for a new amino acid at the junction site but the remaining codons of exon 54 are normal; 3) in white (non-shaded cells), junction of the nucleotides of exons 50 and 54 results in an incorrect reading frame that changes the remaining codons of exon 54; and 4) in black (dark shaded cells in row 2 of columns 10 and 12), the junction of the nucleotides of exons 50 and 54 generates a new stop codon at the junction site. (b) Different gRNA combinations were experimentally tested in 293T cells and in myoblasts and PCR amplification generated amplicons of the expected sizes. The sequencing of the amplicons of these hybrid exons showed the expected modifications (first row corresponds to "light grey" above;

second row corresponds to “darker grey” above; third row corresponds to “white” above; fourth row corresponds to “black” above). MW: molecular weight markers;

[0112] FIG. 5 shows that gRNA pairs can induce deletions that restore the reading frame in the *DYS* gene in DMD myoblasts. Sequence (a) obtained from the amplification of the hybrid exon 50-54 following transfection of the gRNA2-50 and gRNA2-54 pair shows a newly formed codon TAT (coding for tyrosine) at the junction site. This new codon is formed by the nucleotide T from the remaining exon 50 and nucleotides AT from the remaining exon 54. Other in-frame and out-of-frame sequences were also found (b);

[0113] FIG. 6 shows that CinDel correction is effective *in vivo* in the hDMD/*mdx* mouse model. The *Tibialis anterior* (TA) of hDMD/mdx mice was electroporated with 2 plasmids coding for gRNA2-50 and gRNA2-54. The mice were sacrificed 7 days later. Surveyor assay (a) was performed on amplicons of exons 50 and 54. Two additional bands due to the cutting by the Surveyor enzyme were observed for amplicons of the muscles electroporated with the gRNAs but not in the control muscles (CTL) not electroporated with gRNAs. PCR amplifications (b) of exon 50, exon 54 and hybrid exon 50-54 from DNA extracted from hDMD/*mdx* muscles electroporated with the gRNA pair. MW: molecular weight markers;

[0114] FIG. 7 shows that CinDel correction in myoblasts restored the DYS protein expression in myotubes. (a) Normal wild-type myoblasts (CTL+), uncorrected DMD myoblasts with a deletion of exons 51–53 (CTL-) as well as CinDel-corrected DMD myoblasts (CinDel) were allowed to fuse to form abundant myotubes containing multiple nuclei. Proteins were extracted from these three types of myotubes. The DMD myoblasts (Δ51–53) were genetically corrected with (b) gRNA2–50 and gRNA2–54 and (c) with gRNA1–50 and gRNA5–54. In b and c, western blot detected no DYS protein in uncorrected DMD myotubes (CTL-), a 427 kDa DYS protein was detected in the wild-type myotubes (CTL+), and a truncated DYS protein (about 400 kDa) was detected in the CinDel-corrected DMD myotubes (CinDel);

[0115] FIG. 8 shows the *in vivo* and *in vitro* activity of the ribonucleoprotein-nucleic acid complex CRISPR/SpCas9. (A) *In vitro* cutting of the dystrophin exon 54 amplicon by the crRNA (5’ CAGAGAATATCAATGCCTCTGUUUUAGAGCUAUGCUGUUUUG), tracrRNA (Edit-R tracrRNA, Dharmacon) and SpCas9 protein complex. (B) Surveyor assay on amplicon of DMD exon 54 obtained following the transduction in Hela cells with the crRNA:tracrRNA or sgRNA 5-54 targeting exon 54 complexed with SpCas9 protein and delivered with Feldan Shuttle FSA (HHHHHHKWKLFKKIGAVLKVLTTGYARAAARQARA, SEQ ID NO: 196) The presence of expected additional bands confirms that exon 54 amplicons contained INDELs that have been introduced by DSBs in exon 54 by the complex. (C) Delivery of the SpCas9 protein and a pair of sgRNAs in hDMD mouse muscles using different transduction techniques (electroporation, Lipofectamine/RNAiMax™ and Feldan shuttles A (SEQ ID NO: 196) and B (HHHHHHKWKLFKKIGAVLKVLTTGYARAAARQARAHHHHHH, SEQ ID NO: 197). The muscles were longitudinally separated in 4 fragments for each transduction method and the result of each fragment is illustrated. The PCR amplification of the hybrid exon 50-54 produced a band at about 600 bp;

[0116] FIG. 9 shows the polynucleotide sequence of hybrid exon 50-54 obtained *in vivo* following intramuscular

delivery of two sgRNA:SpCas9 complexes targeting exons 50 and 54 using the Feldan Shuttle B. (A)(i) Expected (SEQ ID NO: 199) and (ii) obtained (SEQ ID NO: 200) sequences of the hybrid exon 50-54 formed by genome editing using two ribonucleic complexes sgRNA:SpCas9 (gRNAs 1-50 and 5-54) in hDMD/mdx mouse model. In this FIG., forward and reverse primers used for the PCR amplification are highlighted in grey. The remaining section of the exon 50 following the cut by Cas9 is in bold and italic. The junction site is underlined and the remaining part of exon 54 is in bold. (B) Analysis of the homology between the expected sequence (Query) and the sequence obtained (Sbjct) indicate a perfect match between the two sequences in 60% of the clones. This confirms the formation of the hybrid exon 50-54;

[0117] FIG. 10 shows a summary of the CinDel therapeutic approach according to embodiments of the present invention. The *DYS* gene of a DMD patient has a deletion of exons 51, 52 and 53 compared to the wild-type dystrophin. This produces a reading frame shift when the DNA is translated into mRNA resulting into a stop codon in exon 54 which aborts transcription. When exons 50 and 54 are cut by the CinDel treatment, a hybrid exon 50/54 is formed and the reading frame is restored, allowing the normal transcription of the mRNA;

[0118] FIG. 11 shows a plasmid used in this study and protospacer adjacent motif (PAM) sites of exemplary gRNAs shown in Tables 6 and 8. (a) The plasmid pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA (SaCas9/CRISPR system, Addgene plasmid # 61591; SEQ ID NO: 167) containing two BsaI restriction sites necessary for insertion of a protospacer (see below) under the control of the U6 promoter was used in our study..The pX601 plasmid also contains the Cas9 of *S. aureus* (SaCas9). The pSpCas(BB)-2A-Puro plasmid containing the Cas9 of *S. pyogenes* was alternatively used for gRNAs designed for SpCas9 (see example 19 for details). Guide RNAs were designed following the identification of PAMs of the *S. aureus* Cas9 (i.e., NNGRRT or NNGRR(N), see (b) to (h)) or the Streptococcus pyogenes Cas9 (NGG, see (i) and (j)). Panels (b) to (j) show the nucleic acid sequences of f (b) exon 46 (position: 1407236-1407383 in ENS00000198947), (c) exon 47 (position: 1409718-1409867 in ENS00000198947), (d) exon 49 (position: 1502665-1502766 in ENS00000198947), (e) exon 51 (position: 1565292-1565524 in ENS00000198947), (f) exon 52 (position: 1609736-1609853 in ENS0000019894), (g) exon 53 (position: 1659898-1660109 in ENS00000198947), (h) exon 58 (position: 1860391-1860501 in ENS00000198947); (i) part of intron 22; and (j) part of intron 23 of the human DYS gene. (j). The sequences targeted by the gRNA are in bold and the corresponding PAMs are underlined. For exon 46, 2 PAMs (numbered 1 and 2) were identified, 1 in the sense strand and 1 in the antisense strand. For exon 47, 3 PAMs (numbered 3 to 5) were identified, 1 in the sense strand and 2 in the antisense strand. For exon 49, 1 PAM (numbered 6) was identified in the antisense strand. For exon 51, 2 PAMs (numbered 7 and 8) were identified in the antisense strand. For exon 52, 2 PAMs (numbered 9 and 10) were identified, 1 in the sense strand and 1 in the antisense strand. For exon 53, 5 PAMs (numbered 11 to 15) were identified, 3 in the sense strand and 2 in the antisense strand. For exon 58, 3 PAMs (numbered 16 to 18) were identified, 1 in the sense strand and 2 in the antisense strand. For intron 22 (SEQ ID NO: 228), 4 PAMs (numbered 19, 21, 22 and 23) were selected, 2 in the sense strand and 2 in the antisense strand. For intron 23 (SEQ ID NO: 229), 4 PAMs were selected (numbered 20, 24, 25 and 26), 2 in the sense strand and 2 in the antisense strand. The portion of exon 23 downstream of intron 22 and upstream of intron 23 is

highlighted in grey. See Tables 6 and 8 for gRNAs targeting sequences adjoining these PAMs;

[0119] FIG. 12 shows Surveyor assays for gRNA's screening (SaCas9/CRISPR system) in 293T cells. The assay was performed on genomic DNA extracted from 293T cells (a to g) transfected individually with different gRNAs. Screening was performed separately for exon 46 (a), exon 47 (b), exon 49 (c), exon 51 (d), exon 52 (e), exon 53 (f), exon 58 (g). Genomic DNA of non-transfected cells was used as a control (Ct) for the Surveyor assay. The gRNA numbers correspond to the targeted sequences (Table 6). MW: molecular weight marker. Among the gRNAs we tested, gRNAs 9,11 and 15 show no detectable activity under the conditions tested while gRNAs 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 14, 16, 17 and 18 exhibited good efficiency as demonstrated by the surveyor enzyme assay;

[0120] FIG. 13 shows the activity of different gRNA combinations **(SaCas9/CRISPR system)** in 293T cells for which PCR amplification generated amplicons of the expected sizes. (a) The combination of gRNAs 1 and 7 and the combination of gRNAs 1 and 8 generated a hybrid exon 46-51. (b) The combination of gRNAs 1 and 12, combination of gRNAs 1 and 13, combination of gRNAs 2 and 14, and the combination of gRNAs 2 and 15 generated the hybrid exon 46-53. (c) A hybrid exon 47-52 can be generated by the combination of gRNAs 5 and 9. (d) A hybrid exon 49-52 can be generated by the combination of gRNAs 6 and 10. (e) A hybrid exon 49-53 can be generated by the combination of gRNAs 6 and 11. The combination of gRNA 3 and 16, combination of gRNA 4 and 17, and the combination of gRNAs 5 and 18 can generate a hybrid exon 47-58;

[0121] FIG. 14 shows structural representations of integral spectrin-like repeat R19 and of various hybrid spectrin-like repeats. (a) Primary structure alignments for spectrin-like repeats R19, R20 and R21. Exons associated with these spectrin repeats are identified in gray (below the sequences). The secondary structure for spectrin repeats is represented above the sequences, H for alpha helices and C for the loop segments. Residues between pairs of arrows of the same color are deleted in the resulting hybrid spectrin-like repeats R19–R21. For a patient with a deletion of exons 51–53, the reading frame may be restored by skipping exon 50, thus linking directly exon 49–54. Linking points of deletion of exons 49–54 are highlighted in red. The hybrid exons 2–50/2–54 linking points are highlighted blue and those of hybrid exons 1–50/4–54 in green. (b) Homology models for integral spectrin repeat R19 was obtained from eDystrophin Website. (c) The homology model for the deletion of exons 50–53 (obtained by skipping of exon 50 in a patient with a deletion of exons 51–53). The homology models for (d) hybrid exon 2–50/2–54 and (e) hybrid exon 1–50/4–54 are also illustrated. Structural motifs, as identified in the primary sequence alignment, are colored as follows: helix A is in green, helix B is in orange, and helix C is in blue. Loops AB and BC are in light gray. Colors are darker for spectrin repeat R19 and lighter for spectrin repeat R21;

[0122] FIG. 15 shows the activity of different gRNA combinations used in combination with SaCas9. These combinations were experimentally tested in three different myoblast cells from DMD patient with different out-of-frame deletions (delta 49-50, delta 51-53 and delta 51-56) and for which PCR amplification generated amplicons of the expected sizes. The combination of sgRNAs 3 and 16 and the combination of sgRNAs 5 and 18 generated a hybrid exon 47-58 that we amplified by PCR using the forward primer Sense 47' (5'-CAATAGAAGCAAAGACAAGGTAGTTG) (SEQ ID NO: 194) and the reverse primer Antisense 58' (5'-GCACAAACTGATTTATGCATGGTAG) (SEQ ID NO: 195);

[0123] FIG. 16 shows the localization of the cutting site of 18 exemplary sgRNAs (SaCas9/CRISPR) identified in spectrin like repeats and tested (A) and hybrid spectrin-like repeats 17-19, 17-20, 18-20, 18-21, 19-20 and 18-23 generated using various combinations of gRNAs (B) and (C). The hybrid spectrin like repeats 17-19 was generated from the combination of gRNAs 1 [TTCTCCAGGCTAGAAGAACAA] (SEQ ID NO: 106) and 7 [TTGTGTCACCAGAGTAACAGT] (SEQ ID NO: 112) and from combination of gRNAs 1 [TTCTCCAGGCTAGAAGAACAA] (SEQ ID NO: 106) and 8 [AGTAACCACAGGTTGTGTCAC] (SEQ ID NO: 113). The hybrid spectrin like repeats 17-20 was generated from the combination of sgRNAs 1 [TTCTCCAGGCTAGAAGAACAA] (SEQ ID NO: 106) and 12 [CTTCAGAACCGGAGGCAACAG] (SEQ ID NO: 117) and from the combination of gRNAs 1 [TTCTCCAGGCTAGAAGAACAA] (SEQ ID NO: 106) and 13 [CAACAGTTGAATGAAATGTTA] (SEQ ID NO: 118). The hybrid spectrin like repeats 18-20 was generated from the combination of sgRNAs 5 [CTTATGGGAGCACTTACAAGC] (SEQ ID NO: 110) and 9 [TTCAAATTTTGGGCAGCGGTA] (SEQ ID NO: 114). The hybrid spectrin like repeats 18-21 was generated from the combination of sgRNAs 2 [CTGCTCTTTTCCAGGTTCAAG] (SEQ ID NO: 107) and 14 [GCCAAGCTTGAGTCATGGAAG] (SEQ ID NO: 119) and from the combination of gRNAs 2 [CTGCTCTTTTCCAGGTTCAAG] (SEQ ID NO: 107) and 15 [CTTGGTTTCTGTGATTTTCTT] (SEQ ID NO: 120). The hybrid spectrin like repeats 19-20 was generated from the combinations of sgRNAs 6 [TTGCTTCATTACCTTCACTGG] (SEQ ID NO: 111) and 10 [CAAGAGGCTAGAACAATCATT] (SEQ ID NO: 115) and from the combination of gRNAs 6 [TTGCTTCATTACCTTCACTGG] (SEQ ID NO: 111) and 11 [TTGTACTTCATCCCACTGATT] (SEQ ID NO: 116). The hybrid spectrin-like repeats 18-23 was generated from combination of gRNAs 3 [GTCTGTTTCAGTTACTGGTGG] (SEQ ID NO: 108) and 16 [TCATTTCACAGGCCTTCAAGA] (SEQ ID NO: 121) and 5 [CTTATGGGAGCACTTACAAGC] (SEQ ID NO: 110) and 18 [CAATTACCTCTGGGCTCCTGG] (SEQ ID NO: 123). (A) Arrows indicate cut sites which may be induced by gRNAs. (B) Arrows indicate the hybrid junctions which can be obtained with the listed combinations of sgRNAs;

[0124] FIG. 17 shows the DNA sequences of eight hybrid exons obtained from different combinations of gRNAs for use with the SaCas9/CRISPR system. In light grey is represented the first part of the hybrid exon corresponding to the exon (upstream exon) targeted by the first gRNA while in bold is represented the last part of the hybrid exon corresponding to the exon targeted by the second gRNA. The slash (/) indicates the location of the junction;

[0125] FIG. 18 illustrates the results of the sequencing of the hybrid exons generated from several gRNAs combinations (SaCas9/CRISPR system) following cloning of PCR product into pMiniT plasmid vector. (a) overall number of clones presenting the precise nucleotide sequences of the expected hybrid exons (identified in FIG. 17) in comparison to the overall number of sequenced clones obtained in 293T cells. (b) shows the sequencing results of hybrid exons 47-58 generated from the combination of gRNAs 3 and 16 and from the combination of gRNAs 5 and 18 in myoblasts from three different DMD patients harbouring a deletion of exons 49-50, or a deletion of exons 51 to 53, or a deletion of exons 51 to 56. Using the combination of gRNAs 3 and 16, we restored a correct reading frame in 70% (7/10), 40% (4/10) and 88.8% (8/9) of the generated hybrid exons in myoblasts harbouring the

deletion 49-50, 51-53 and 51- 56 respectively. Using the combination of gRNAs 5 and 18, we restored a correct reading frame in 70% (7/10), 50% (5/10) and 63.6% (7/11) of the generated hybrid exons in myoblasts harbouring the deletion 49-50, 51-53 and 51-56 respectively;

[0126] FIG. 19 shows the cDNA sequence (SEQ ID NO: 1) of the human *DYS* gene and the encoded amino acid sequence (SEQ ID NO: 2) of human dystrophin (transcript DMD-001 (ENST00000357033.8) of ENSG00000198947). Exons are shown in the first line via alternating upper and lower case sequence regions;

[0127] FIG. 20 shows the cDNA sequence of the human *DYS* gene (transcript DMD-001 (ENST00000357033.8) of ENSG00000198947). cDNA sequence (SEQ ID NO: 1) is shown in uppercase, grouped by exons. Flanking intronic sequences (25 bases on either side of a given exon) are shown in lowercase, not bold. 25 nts of 5' UTR are shown in lowercase bold at beginning; 25 nts of 3' UTR are shown in lowercase bold at end. 25 nts of 5' UTR + cDNA sequence of exon 1 + 25 nts of intron sequence at 3' correspond to SEQ ID NO: 3; cDNA sequences of exons 2 to 78 with flanking 25 nts of intron sequences on each side (5' and 3' correspond to SEQ ID NOs: 4-80, respectively; 25 nts of intron sequence at 3'+ cDNA sequence of exon 79 + 25 nts of 3' UTR correspond to SEQ ID NO: 81;

[0128] FIG. 21 shows different gRNA combinations **(SaCas9/CRISPR system)** that were experimentally tested *in vivo* in the mouse model hDMD/mdx using two AAV9 vectors. (a) shows that the intravenous (IV) or the intraperitoneal (IP) injections of a combination of $7,5.10^{11}$ vg of AAV9 encoding the SaCas9 along with $7,5.10^{11}$vg of AAV9 encoding the gRNA 3 and the gRNA 16 or the gRNA 5 and the gRNA 18 permit to generate the hybrid exons 47-58 *in vivo* in the heart of a mouse model as demonstrated by the generations of PCR amplicons of expected sizes. (b) shows that intravenous (IV) or intraperitoneal (IP) injections of a combination of $7,5.10^{11}$ vg of AAV9 encoding the SaCas9 along with $7,5.10^{11}$vg of AAV9 encoding gRNA 3 and gRNA 16 or gRNA 5 and gRNA 18 permit to generate hybrid exons 47-58 *in vivo* in the diaphragm of a mouse model as demonstrated by the generation of PCR amplicons of expected sizes. (c) shows that intravenous (IV) injections of a combination of $7,5.10^{11}$vg of AAV9 encoding the SaCas9 along with $7,5.10^{11}$vg of AAV9 encoding gRNA 3 and gRNA 16 permits to generate hybrid exons 47-58 *in vivo* in the *Tibialis anterior* of a mouse model as demonstrated by the generation PCR amplicons of expected size. However, we were not able to detect the hybrid exon in the *Tibialis anterior* through intraperitoneal injections and for the intravenous injection of AAV9 encoding the gRNA 5 and the gRNA 18;

[0129] FIG. 22 shows the activity of gRNAs targeting intron 22 or intron 23 in C2C12 myoblasts. (a) shows a surveyor enzyme assay of the 8 gRNAs selected that target the surrounding introns of the exon 23. gRNA 19, gRNA 21, gRNA 22 and gRNA 23 target intron 22 while gRNA 20, gRNA 24, gRNA 25 and gRNA 26 target intron 23. Target sequences for these gRNAs are shown in Table 8. Genomic DNA of non-transfected and non-selected C2C12 cells was used as control (NT) for the Surveyor assay. (b) Shows the different gRNA combinations (SaCas9/CRISPR system) that were experimentally tested and for which PCR amplification generated two products: the upper PCR amplification product corresponds to wild-type genomic DNA while the second amplification product of lower size corresponds to the deletion of the exon 23;

[0130] FIG. 23 shows immunohistochemistry analysis of muscle cuts labelled with an anti-dystrophin antibody in 4 months old mice following intramuscular injection of SpCas9/gRNAs complexed with FSA, FSB and FSC. We tested the intramuscular injection of the combination of the gRNA 20 and 22 and the combination of gRNAs 20 and 23 along with the SpCas9 nuclease complexed with the FSA (SEQ ID NO: 196), FSB (SEQ ID NO: 197) and FSC (HHHHHHLLKLWSRLLKLWTQGRRLKAKRAKAHHHHHH, SEQ ID NO: 198). Three weeks after the intramuscular injection, mice were sacrificed. Following *Tibialis anterior* harvesting, samples were submitted to the staining of the dystrophin using the primary mouse NCL-Dys2 anti-dystrophin antibody and the secondary Alexa Fluor 546 goat anti-mouse antibody (H+L). (a) shows the expression of dystrophin in a non-treated mouse. Thus, dystrophin positive fibers identify dystrophin revertant fibers. (b) shows the expression of the dystrophin protein in mouse injected with the FSA complexed with the SpCas9 and the combination of gRNAs 20 and 22. (c) shows the expression of the dystrophin protein in mouse injected with the FSB complexed with the SpCas9 and the combination of gRNAs 20 and 22. (d) shows the expression of the dystrophin protein in mouse injected with the FSC complexed with the SpCas9 and the combination of gRNAs 20 and 22. (e) shows the expression of the dystrophin protein in mouse injected with the FSA complexed with the SpCas9 and the combination of gRNAs 20 and 23. (f) shows the expression of the dystrophin protein in mouse injected with the FSB complexed with the SpCas9 and the combination of gRNAs 20 and 23. (g) shows the expression of the dystrophin protein in mouse injected with the FSC complexed with the SpCas9 and the combination of gRNAs 20 and 23;

[0131] FIG. 24 shows the number of positive dystrophin fibers identified by immunohistochemistry following intramuscular injection in Rag/mdx mice of gRNAs and SpCas9 complexed with various polypeptide-based shuttles (Feldan Shuttles). FSA, FSB and FSC Feldan Shuttles were used for the delivery of the SpCas9 protein complexed with a combination of gRNAs 22 and 20 or with a combination of gRNAs 23 and 20. In the control muscle (NT) we reported near 37 dystrophin positive fibers (SD +/- 4.68), which correspond to revertant fibers. For the combination of gRNAs 4 and 2, we counted 83 (SD +/- 18.8), 49.5 (SD +-/ 8.6) and 186 (SD +/- 52) dystrophin positive fibers with Feldan Shuttles FSA, FSB and FSC respectively. For the combination of gRNAs 5 and 2, we counted 39 (SD +/- 16.0), 99 (SD +-/ 15.6) and 199 (SD +/- 37.2) dystrophin positive fibers with Feldan Shuttles FSA, FSB and FSC respectively;

[0132] FIG. 25 shows the results of the PCR amplification for the detection of the hybrid intron (i.e., deletion of the exon 23) from genomic DNA of Rag/mdx mice following intramuscular injection of gRNAs and SpCas9 complexed with various polypeptide-based shuttles (Feldan Shuttles). (a) shows the amplicons generated from the first PCR using the primers Fw1-i22 and Rev1-i23 along with the poison primer Fw-P. The amplicons detected correspond to the amplification from Fw-P and Rev1-i22 of the wild type genomic DNA, where no deletion of the exon 23 is detected. (b) shows the results of the Nested-PCR amplifications performed on 1 µL of the previous PCR reaction with internal primers Fw2-i22 and Rev2-i22. NT only exhibited amplification of a wild type sequence, as expected. The delivery of the combination of gRNAs 22 and 20 complexed with the Cas9 and the FSA, and FSB, and FSC, allowed the amplification of a truncated PCR product of expected size for deletion of exon 23. Besides, delivery of the combination of gRNAs 23 and 20 complexed with SpCas9 and the FSA, FSB or FSC,

allowed the amplification of a truncated PCR product of expected size for deletion of the exon 23. However, the combination of gRNAs 23 and 20 along with the FSA shuttle did not allow the detection of a truncated PCR product. Sequencing of the truncated PCR products are underway to confirm the deletion of the exon 23;

[0133] FIG. 26 shows immunohistochemistry analysis of muscle cuts labelled with an anti-dystrophin antibody in 6 weeks old mice following intramuscular injection of SpCas9/gRNAs complexed with FSC or not. In this experiment, we tested the intramuscular injection of the combination of gRNAs 20 and 22 and the combination of gRNAs 20 and 23 with the protein SpCas9 complexed, or not, with the FSC. 6 weeks old mice were injected and *Tibialis anterior* were collected 3 weeks after the injection. Samples were submitted to the staining of the dystrophin using the primary mouse NCL-Dys2 anti-dystrophin antibody and the secondary Alexa Fluor 546 goat anti-mouse antibody (H+L). (a) and (b) show the expression of the dystrophin protein in muscles injected with gRNAs 20 and 23 complexed with the SpCas9 protein and with no shuttle. (c) shows the expression of the dystrophin protein in muscles injected with gRNAs 20 and 23 complexed with the SpCas9 protein and with the Feldan Shuttle FSC. (d) and (e) show the expression of the dystrophin protein in muscles injected with gRNAs 20 and 22 complexed with the SpCas9 protein and with no shuttle. (f) and (g) show the expression of the dystrophin protein in muscles injected with gRNAs 20 and 22 complexed with the SpCas9 protein and with the Feldan Shuttle FSC;

[0134] FIG. 27 shows the number of positive dystrophin fibers identified by immunohistochemistry following intramuscular injection in 6 weeks old Rag/mdx mice of gRNAs and SpCas9 complexed or not with the Feldan Shuttle FSC. Here we compared the delivery of the SpCas9 protein complexed with gRNAs 20 and 22 or gRNAs 20 and 23 supplemented, or not, with the FSC which previously exhibited the most promising results. In muscles injected with and SpCas9 complexed with gRNAs 20 and 22 or gRNAs 20 and 23 without shuttle, we counted 101 (SD +/- 6.2) and 120 (SD +/-38.9) dystrophin positive fibers, respectively. When the FSC was used in combination to the SpCas9 and gRNAs 20 and 22 or gRNAs 20 and 23 we detected a higher amount of dystrophin positive fibers as we respectively reported 151 (SD +/- 93.3) and 166 (SD +/- 25.6) positive fibers; and

[0135] FIG. 28 shows the PCR detection of the hybrid intron (i.e., deletion of exon 23) in 6 weeks old mice using the Feldan Shuttle FSC. (a) shows the amplicons generated from the first PCR using the primers Fw1-i22 and Rev1-i23 along with the poison primer Fw-P. The amplicons detected correspond to the amplification from Fw-P and Rev1-i22 of the wild type genomic DNA, where no deletion of the exon 23 is detected. (b) shows the results of the Nested-PCR amplifications performed on 1 µL of the previous PCR reaction with internal primers Fw2-i22 and Rev2-i22. NT only exhibited amplification of a wild type sequence, as expected. The delivery of the combination of gRNAs 22 and 20 complexed with the Cas9 with or without the shuttle FSC, allowed the amplification of a truncated PCR product of expected size for deletion of exon 23. Besides, delivery of the combination of gRNAs 23 and 20 complexed with SpCas9 with or without FSC, allowed the amplification of a truncated PCR product of expected size for deletion of the exon 23 except for the first lane regarding the gRNAs 20 and 23 with the FSC which corresponds to a failed injection. Sequencing of the truncated PCR products are underway to confirm the deletion of the exon 23.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0136] In a first aspect, the present invention is based on Applicants' finding that by introducing mutations within exon sequences located up-stream and downstream of an endogenous frameshift or nonsense mutation in the *DYS* gene of a cell, it is possible to restore the correct reading frame and in turn restore dystrophin expression within the cell. Preferably, the mutations correcting the reading frame are introduced as close as possible to the endogenous frameshift or nonsense mutation, but within an exon. Given that the sites of the engineered mutations are within one or more exons, the corrected gene has a fusion of two exon portions (i.e. which are normally not contiguous with one another), and at the same time restoring the correct reading frame of the *DYS* gene. Using this approach, Applicants have found that it is possible to restore dystrophin expression within the cell to produce a dystrophin protein having smaller deletions and being functionally closer to the wild-type dystrophin protein.

[0137] Several approaches can be used to introduce one or more mutations within one or more exons of the dystrophin gene and restore dystrophin expression. For example, sequence-specific nucleases such as meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the CRISPR/Cas9 system can be used to introduce one or more targeted mutations within one or more exons of the *DYS* gene to restore dystrophin expression. Depending on the endogenous mutation already present in *DYS* gene within the cell, the method of the present invention may or may not lead to the expression of a wild-type dystrophin protein. However, it has been found that by targeting exon sequences (as opposed to introns) which are close to the endogenous mutation(s), the cell will advantageously express a dystrophin protein having a function which is closer to that of the wild-type dystrophin protein.

[0138] In a particular embodiment, the present invention uses the CRISPR system to introduce further mutations within exons of a mutated dystrophin gene within a cell. The CRISPR system is a defense mechanism identified in bacterial species [37-42]. It has been modified to allow gene editing in mammalian cells. The modified system still uses a Cas9 nuclease to generate double-strand breaks (DSB) at a specific DNA target sequence [43, 44]. The recognition of the cleavage site is determined by base pairing of the gRNA with the target DNA and the presence of a trinucleotide called PAM (protospacer adjacent motif) juxtaposed to the targeted DNA sequence [45]. This PAM is NGG for the Cas9 of *S. pyogenes*, the most commonly used enzyme [46, 47]. This PAM is NNGRRT or NNGRR(N) for the high efficiency Cas9 of *Staphylococcus aureus*, a smaller Cas9 which can advantageously be used in the context of adeno-associated virus delivery and paired nickase (in case of the D10A and N580A variant) applications.

[0139] In a further aspect, the present invention is based on Applicant's finding that delivery of a combination of gRNAs and CRISPR nuclease (e.g., Cas9) in cells is particularly effective using polypeptide-based shuttles (Feldan Shuttles, from Feldan Therapeutics).

[0140] The shuttles deliver proteins or nucleic acids (e.g., gRNAs) directly inside cells. Once delivered, these proteins/gRNAs dislodge from the carrier and are free to induce genetic modifications. The shuttles are 100% protein-based and are rapidly degraded in the cell, leaving no toxic residues. They are capable of delivering proteins/nucleic acids in multiple cell types without altering cell viability. The Shuttles are virus-free, DNA- and RNA-free, thereby eliminating potential mutagenic risks.

[0141] Useful polypeptide-based shuttles and methods of preparing and using polypeptide-based (Feldan) shuttles are described in US Application No. 15/094,365 (Polypeptide-based shuttle agents for improving the transduction efficiency of polypeptide cargos to the cytosol of target eukaryotic cells, uses thereof, methods and kits relating to same), filed on April 8th, 2016; in International (PCT) application No. PCT/CA2016/050403 (Polypeptide-based shuttle agents for improving the transduction efficiency of polypeptide cargos to the cytosol of target eukaryotic cells, uses thereof, methods and kits relating to same), filed on April 8th 2016; in US provisional application No 62/320,065 (Peptide shuttle based gene disruption), filed on April 8th, 2016; and in U.S. Provisional Patent Application N° 62/407,232 (Rationally-designed synthetic peptide shuttle agents for delivering polypeptide cargos from an extracellular space to the cytosol and/or nucleus of a target eukaryotic cell, uses thereof, methods and kits relating to same), filed on October 12th, 2016, which are incorporated by reference in their entirety.

DEFINITIONS

[0142] In order to provide clear and consistent understanding of the terms in the instant application, the following definitions are provided.

[0143] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[0144] The articles "a," "an" and "the" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article.

[0145] As used in this specification and claim(s), the words “comprising” (and any form of comprising, such as “comprise” and “comprises”), “having” (and any form of having, such as “have” and “has”), “including” (and any form of including, such as “includes” and “include”) or “containing” (and any form of containing, such as “contains” and “contain”) are inclusive or open-ended and do not exclude additional, un-recited elements or method steps and are used interchangeably with, the phrases "including but not limited to" and "comprising but not limited to".

[0146] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 18-20, the numbers 18, 19 and 20 are explicitly contemplated, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated. The terms "such as" are used herein to mean, and is used interchangeably with, the phrase "such as but not limited to".

[0147] Practice of the methods, as well as preparation and use of the products and compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure

and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. See, for example, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P. M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P. B. Becker, ed.) Humana Press, Totowa, 1999.

[0148] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (e.g., phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; i.e., an analogue of A will base-pair with T.

[0149] Various genes and nucleic acid sequences of the invention may be recombinant sequences. The term "recombinant" means that something has been recombined, so that when made in reference to a nucleic acid construct the term refers to a molecule that is comprised of nucleic acid sequences that are joined together or produced by means of molecular biological techniques. The term "recombinant" when made in reference to a protein or a polypeptide refers to a protein or polypeptide molecule, which is expressed using a recombinant nucleic acid construct created by means of molecular biological techniques. The term "recombinant" when made in reference to genetic composition refers to a gamete or progeny or cell or genome with new combinations of alleles that did not occur in the parental genomes. Recombinant nucleic acid constructs may include a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Referring to a nucleic acid construct as "recombinant" therefore indicates that the nucleic acid molecule has been manipulated using genetic engineering, i.e. by human intervention. Recombinant nucleic acid constructs may for example be introduced into a host cell by transformation. Such recombinant nucleic acid constructs may include sequences derived from the same host cell species or from different host cell species, which have been isolated and reintroduced into cells of the host species. Recombinant nucleic acid construct sequences may become integrated into a host cell genome, either as a result of the original transformation of the host cells, or as the result of subsequent recombination and/or repair events.

[0150] The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

[0151] "Coding sequence" or "encoding nucleic acid" as used herein means the nucleic acids (RNA or DNA

molecule) that comprise a nucleotide sequence which encodes a protein or gRNA. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The coding sequence may be codon optimized.

[0152] "Complement" or "complementary" as used herein refers to Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[0153] "Subject" and "patient" as used herein interchangeably refers to any vertebrate, including, but not limited to, a mammal (e.g., cow, pig, camel, llama, horse, goat, rabbit, sheep, hamsters, guinea pig, cat, dog, rat, and mouse, a non-human primate (for example, a monkey, such as a cynomolgous or rhesus monkey, chimpanzee, etc.) and a human). In some embodiments, the subject may be a human or a non-human. In an embodiment, the subject or patient may suffer from DMA and has a mutated dystrophin gene. The subject or patient may be undergoing other forms of treatment.

[0154] “Vector" as used herein means a nucleic acid sequence containing an origin of replication. A "vector" as described herein refers to a vehicle that carries a nucleic acid sequence and serves to introduce the nucleic acid sequence into a host cell. In an embodiment, the vector will comprise transcriptional regulatory sequences or a promoter operably-linked to a nucleic acid comprising a sequence capable of encoding a gRNA, nuclease or polypeptide described herein. In embodiments, the promoter is a U6 or CBh promoter. A first nucleic acid sequence is "operably-linked" with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably-linked to a coding sequence if the promoter affects the transcription or expression of the coding sequences. Generally, operably-linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in reading frame. However, since, for example, enhancers generally function when separated from the promoters by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably-linked but not contiguous. "Transcriptional regulatory element" is a generic term that refers to DNA sequences, such as initiation and termination signals, enhancers, and promoters, splicing signals, polyadenylation signals which induce or control transcription of protein coding sequences with which they are operably-linked. A vector may be a viral vector (e.g., AAV), bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self- replicating extrachromosomal vector, and preferably, is a DNA plasmid. For example, the vector may comprise nucleic acid sequence(s) that/which encode(s) at least one gRNA and/or CRISPR nuclease (e.g. Cas9) described herein. Alternatively, the vector may comprise nucleic acid sequence(s) that/which encode(s) one or more CRISPR nucleases (Cas9 or Cpf1) at least one (preferably at least 2) gRNA nucleotide sequence of the present invention. A vector for expressing one or more gRNA will comprise a “DNA” sequence of the gRNA.

[0155] "Adeno-associated virus" or "AAV" as used interchangeably herein refers to a small virus belonging to the

genus Dependovirus of the Parvoviridae family that infects humans and some other primate species. AAV is not known to cause disease and consequently the virus causes a very mild immune response.

Sequence similarity

[0156] "Homology" and “homologous” refers to sequence similarity between two peptides or two nucleic acid molecules. Homology can be determined by comparing each position in the aligned sequences. A degree of homology between nucleic acid or between amino acid sequences is a function of the number of identical or matching nucleotides or amino acids at positions shared by the sequences. As the term is used herein, a nucleic acid sequence is "substantially homologous" to another sequence if the two sequences are substantially identical and the functional activity of the sequences is conserved (as used herein, the term “homologous” does not infer evolutionary relatedness, but rather refers to substantial sequence identity, and thus is interchangeable with the terms “identity”/”identical”). Two nucleic acid sequences are considered substantially identical if, when optimally aligned (with gaps permitted), they share at least about 50% sequence similarity or identity, or if the sequences share defined functional motifs. In alternative embodiments, sequence similarity in optimally aligned substantially identical sequences may be at least 60%, 70%, 75%, 80%, 85%, 90% or 95%. For the sake of brevity, the units (e.g., 66, 67...81, 82,...91, 92%....) have not systematically been recited but are considered, nevertheless, within the scope of the present invention.

[0157] Substantially complementary nucleic acids are nucleic acids in which the complement of one molecule is substantially identical to the other molecule. Two nucleic acid or protein sequences are considered substantially identical if, when optimally aligned, they share at least about 70% sequence identity. In alternative embodiments, sequence identity may for example be at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 98% or at least 99%. Optimal alignment of sequences for comparisons of identity may be conducted using a variety of algorithms, such as the local homology algorithm of Smith and Waterman, 1981, Adv. Appl. Math 2: 482, the homology alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48:443, the search for similarity method of Pearson and Lipman (Pearson and Lipman 1988), and the computerized implementations of these algorithms (such as GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, WI, U.S.A.). Sequence identity may also be determined using the BLAST algorithm, described in Altschul et al. (Altschul et al. 1990) 1990 (using the published default settings). Software for performing BLAST analysis may be available through the National Center for Biotechnology Information (through the internet at http://www.ncbi.nlm.nih.gov/). The BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold. Initial neighbourhood word hits act as seeds for initiating searches to find longer HSPs. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction is halted when the following parameters are met: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue

alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. One measure of the statistical similarity between two sequences using the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. In alternative embodiments of the invention, nucleotide or amino acid sequences are considered substantially identical if the smallest sum probability in a comparison of the test sequences is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0158] An alternative indication that two nucleic acid sequences are substantially complementary is that the two sequences hybridize to each other under moderately stringent, or preferably stringent, conditions. Hybridization to filter-bound sequences under moderately stringent conditions may, for example, be performed in 0.5 M NaHPO4, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.2 x SSC/0.1% SDS at 42°C (Ausubel 2010). Alternatively, hybridization to filter-bound sequences under stringent conditions may, for example, be performed in 0.5 M NaHPO4, 7% SDS, 1 mM EDTA at 65°C, and washing in 0.1 x SSC/0.1% SDS at 68°C (Ausubel 2010). Hybridization conditions may be modified in accordance with known methods depending on the sequence of interest (Tijssen 1993). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point for the specific sequence at a defined ionic strength and pH.

[0159] "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid or between a gRNA and a target polynucleotide or between a gRNA and a CRISPR nuclease (e.g., Cas9, Cpf1). Not all components of a binding interaction need to be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. “Affinity” refers to the strength of binding: increased binding affinity being correlated with a lower Kd. A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

[0160] As used herein, a “target gene”, “targeted gene”, “targeted polynucleotide” or “targeted gene sequence” corresponds to the polynucleotide within a cell that will be modified, in an embodiment by the introduction a gRNA pair and a CRISPR nuclease. It corresponds to an endogenous gene naturally present within a cell. In an embodiment, the targeted gene is a DYS gene comprising one or more frameshift or nonsense mutations associated with the development of MD (e.g., DMD or BMD). One or both alleles of a targeted gene may be modified within a cell in accordance with the present invention.

[0161] A “frameshift mutation” is a mutation within a polynucleotide sequence (e.g., a DNA gene sequence) caused by indels (insertions or deletions) of a number of nucleotides in a DNA sequence that is not divisible by three. Due to the triplet nature of gene expression by codons, a frameshift mutation caused by one or more insertion

or deletion changes the reading frame (the grouping of the codons), resulting in a different translation product (mutated polypeptide) from the original (wild type polypeptide). The frameshift mutation can result in a truncated polypeptide (by generating a premature stop codon), in a larger polypeptide (by removing a stop codon), and/or in a completely or partially different amino acid sequence. Mutated polypeptides caused by one or more frameshift mutations in the wild-type polynucleotide (gene) sequence are often non-functional, such as in the case of mutations in the DYS gene causing MD (e.g., DMD or BMD).

[0162] A “nonsense mutation” is a point mutation within a polynucleotide sequence (a DNA gene sequence) that results in a premature stop codon, or a nonsense codon in the transcribed mRNA, and in a truncated, incomplete, and usually nonfunctional protein product. “A nonsense mutation” within the context of the present invention encompasses the presence of at least one nonsense mutation within a polynucleotide (e.g. gene) sequence and thus includes the presence of a plurality of such mutations.

[0163] As used herein an “endogenous frameshift mutation” or an “endogenous nonsense mutation” is a mutation which is naturally present within a gene of a subject. For example, in the context of muscular dystrophy (DMD or BMD), an endogenous frameshift mutation or an endogenous nonsense mutation is a mutation found in the dystrophin gene (in one or both alleles) of cells of a subject suffering from muscular dystrophy. The presence of such endogenous mutation (frameshift or nonsense) is responsible for the development of the disease. The presence of such mutation results in a reduced level of dystrophin protein expression, in the production of an unstable protein (with a reduced half-life), in the production of a a truncated protein and/or in an inactive protein (or less active protein compared to the wild-type endogenous protein).

[0164] "Promoter" as used herein means a synthetic or naturally-derived nucleic acid molecule which is capable of conferring, modulating or controlling (e.g., activating, enhancing and/or repressing) expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance or repress expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the U6 promoter, bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter and the CMV IE promoter. In embodiments, the U6 promotor is used to express one or more gRNAs in a cell.

CRISPR system

[0165] CRISPR technology is a system for genome editing, e.g., for modification of the expression of a specific gene.

[0166] This system stems from findings in bacterial and archaea which have developed adaptive immune defenses termed clustered regularly interspaced short palindromic repeats (CRISPR) systems, which use CRISPR targeting RNAs (crRNAs) and Cas proteins to degrade complementary sequences present in invading viral and plasmid DNA. Jinek et al. (47) and Mali et al. (41) have engineered a type II bacterial CRISPR system using custom guide RNA (gRNA) to induce double strand break(s) in DNA. In one system, the Cas9 protein was directed to genomic target sites by a synthetically reconstituted "guide RNA" ("gRNA), which corresponds to a crRNA and tracrRNA which can be used separately or fused together, that obviates the need for RNase III and crRNA processing in general. It comprises a “gRNA guide sequence” or “gRNA target sequence” and an RNA sequence (Cas9 recognition sequence), which is necessary for Cas (e.g., Cas9) binding to the targeted gene. The gRNA guide sequence is the sequence which confers specificity. It hybridizes with (i.e., it is complementary to) the opposite strand of a target sequence (i.e., it corresponds to the RNA sequence of a DNA target sequence). Other CRISPR systems using difference CRISPR nucleases have been developed and are known in the art (e.g., using the Cpf1 nuclease instead of a Cas9 nuclease).

[0167] One may alternatively use in accordance with the present invention a pair of specifically designed gRNAs in combination with a Cas9 nickase or in combination with a dCas9-FokI nuclease to cut both strands of DNA.

[0168] In embodiments, provided herein are CRISPR/nuclease-based engineered systems for use in modifying the *DYS* gene and restoring its correct reading frame. The CRISPR/nuclease-based systems of the present invention include at least one nuclease (e.g. a Cas9 or Cpf1 nuclease) and at least one gRNA targeting the endogenous *DYS* gene in target cells.

[0169] Accordingly, in an aspect, the present invention involves the design and preparation of one or more gRNAs for inducing a DSB (or two single stranded breaks (SSB) in the case of a nickase) in a *DYS* gene. The gRNAs (targeting the *DYS* gene) and the nuclease are then used together to introduce the desired modification(s) (i.e., gene-editing events), e.g., by NHEJ or HDR, within the genome of one or more target cells.

[0170] The CRISPR/nuclease-based systems of the present invention include at least one CRISPR nuclease (e.g. a Cas9 or Cpf1 nuclease) and at least one gRNA targeting the endogenous DYS gene in target cells.

gRNAs

[0171] In order to cut DNA at a specific site, CRISPR nucleases require the presence of a gRNA and of a protospacer adjacent motif (PAM) on the targeted gene. The PAM, immediately follows (i.e., is adjacent to) the gRNA target sequence in the targeted polynucleotide gene sequence. The PAM is located at the 3’ end or 5’ end of the sgRNA target sequence (depending on the CRISPR nuclease used) but is not included in the sgRNA guide sequence. For example, the PAM for Cas9 CRSIPR nucleases is located at the 3’ end of the sgRNA target sequence on the target gene while the PAM for Cpf1 nucleases is located at the 5’ end of the sgRNA target sequence on the target gene. Different CRISPR nucleases also require a different PAM. Accordingly, selection of a specific polynucleotide gRNA target sequence (e.g., in the DYS gene nucleic acid sequence) is generally based on the CRISPR nuclease used. The PAM for the *Streptococcus pyogenes* Cas9 CRISPR system is 5'-NRG-3', where R

is either A or G, and characterizes the specificity of this system in human cells. The PAM of *S. aureus* Cas9 is NNGRR(T). The S. *pyogenes* Type II system naturally prefers to use an "NGG" sequence, where "N" can be any nucleotide, but also accepts other PAM sequences, such as "NAG" in engineered systems. Similarly, the Cas9 derived from *Neisseria meningitidis* (NmCas9) normally has a native PAM of NNNNGATT, but has activity across a variety of PAMs, including a highly degenerate NNNNGNNN PAM. The PAM for a AsCpf1 or LbCpf1 CRISPR nuclease is TTTN. In a preferred embodiment, the PAM for a Cas9 or Cpf1 protein is used in accordance with the present invention is a NGG trinucleotide-sequence (Cas9) or TTTN (AsCpf1 and LbCpf1). Table 1 below provides a list of non-limiting examples of CRISPR/nuclease systems with their respective PAM sequences.

Table 1: Non-exhaustive list of CRISPR-nuclease systems from different species (see. Mohanraju, P. et al. (60); Shmakov, S et al. (61); and Zetsche, B. et al. (62). Also included are engineered variants recognizing alternative PAM sequences (see Kleinstiver, BP. et al., (63)).

CRISPR nuclease	PAM Sequence
Streptococcus pyogenes (SP); SpCas9	NGG + NAG
SpCas9 D1135E variant	NGG (reduced NAG binding)
SpCas9 VRER variant	NGCG
SpCas9 EQR variant	NGAG
SpCas9 VQR variant	NGAN or NGNG
Staphylococcus aureus (SA); SaCas9	NNGRRT or NNGRR(N)
SaCas9 KKH variant	NNNRRT
Neisseria meningitides (NM)	NNNNGATT
Streptococcus thermophilus (ST)	NNAGAAW
Treponema denticola (TD)	NAAAAC
AsCpf1	TTTN
AsCpf1 S542R/K607R	TYCV
AsCpf1 S542R/K548V/N552R	TATV
LbCpf1	TTTN
LbCpf1 G532R/K595R	TYCV

[0172] As used herein, the expression “gRNA” or “sgRNA” refers to a guide RNA which works in combination with a CRISPR nuclease to introduce a cut into DNA. The sgRNA comprises a sgRNA guide sequence (corresponding to the target sequence) and a “CRISPR nuclease recognition sequence”.

[0173] As used herein, the expression “gRNA guide sequence” refers to the corresponding RNA sequence of the

"gRNA target sequence" (also known as the spacer sequence). Therefore, it is the RNA sequence equivalent of the protospacer on the target polynucleotide gene sequence. It does not include the corresponding PAM sequence in the genomic DNA. It is the sequence that confers target specificity. In embodiments, the gRNA guide sequence is linked to a CRISPR nuclease recognition sequence which binds to the nuclease (e.g., Cas9/Cpf1). The sgRNA guide sequence recognizes and binds to the targeted gene of interest. It hybridizes with (i.e., is complementary to) the opposite strand of a target gene sequence, which comprises the PAM (i.e., it hybridizes with the DNA strand opposite to the PAM). As noted above, the "PAM" is the nucleic acid sequence, that immediately follows (is contiguous to) the target sequence on the *DYS* gene or target polynucleotide but is not in the gRNA.

[0174] In embodiments, the gRNA is a fusion between the gRNA guide sequence and the CRISPR nuclease recognition sequence (CRISPR repeat and optionally tracrRNA). It provides both targeting specificity and scaffolding/binding ability for the CRSIPR nuclease of the present invention. gRNAs of the present invention do not exist in nature, i.e., they are non-naturally occurring nucleic acid(s).

[0175] A "target region", "target sequence" or "protospacer" in the context of gRNAs and CRISPR system of the present invention are used herein interchangeably and refers to the region of the target gene, which is targeted by the CRISPR/nuclease-based system, without the PAM. It refers to the sequence corresponding to the nucleotides that precede the PAM (i.e., in 5' or 3' of the PAM, depending of the CRISPR nuclease) in the genomic DNA. It is the sequence that is included into a gRNA expression construct (e.g., vector/plasmid/AVV). The CRISPR/nuclease-based system may include at least one (i.e., one or more, preferably two) gRNAs, wherein each gRNA targets a different DNA sequence on the target gene. The target DNA sequences may be overlapping. The target sequence or protospacer is followed or preceded by a PAM sequence at an end (3' or 5' depending on the CRISPR nuclease used) of the protospacer. Generally, the target sequence is immediately adjacent (i.e., is contiguous) to the PAM sequence (it is located on the 5' end of the PAM for SpCas9-like nuclease and at the 3' end for Cpf1-like nuclease).

[0176] A "CRISPR nuclease recognition sequence" as used herein refers broadly to one or more RNA sequences (or RNA motifs) required for the binding and/or activity (including activation) of the CRISPR nuclease on the target gene (such as "UAAUUUCUAC UCUUGUAGAU" (SEQ ID NO: 168) in 5' for Cpf1 nuclease). It encompasses the structural piece (the repeat sequence) that normally complements the tracrRNA and the tracrRNA sequence. Some CRISPR nucleases require longer RNA sequences than other to function. Also, some CRISPR nucleases require multiple RNA sequences (motifs) to function while others only require a single short RNA sequence/motif. For example, Cas9 proteins require a tracrRNA sequence in addition to a crRNA sequence (repeat) to function while Cpf1 only requires a crRNA sequence. Thus, unlike Cas9, which requires both crRNA sequence and a tracrRNA sequence (or a fusion or both crRNA and tracrRNA) to mediate interference, Cpf1 processes crRNA arrays independent of tracrRNA, and Cpf1-crRNA complexes alone cleave target DNA molecules, without the requirement for any additional RNA species (see Zetsche et al., PMID: 26422227).

[0177] The "CRISPR nuclease recognition sequence" included in the sgRNA described herein is thus selected based on the specific CRISPR nuclease used. It includes direct repeat sequences and any other RNA sequence known to be necessary for the selected CRISPR nuclease binding and/or activity. Various RNA sequences which

can be fused to an RNA guide sequence to enable proper functioning of CRISPR nucleases (referred to herein as CRISPR nuclease recognition sequence) are well known in the art and can be used in accordance with the present invention. The "CRISPR nuclease recognition sequence" may thus include a crRNA sequence only (e.g., for Cpf1 activity, such as the CRISPR nuclease recognition sequence UAAUUUCUAC UCUUGUAGAU set forth in SEQ ID NO: 168) or may include additional sequences (e.g., tracrRNA sequence necessary for Cas9 activity, such as the CRISPR nuclease recognition sequence set forth in SEQ ID NO: 166 which includes both crRNA and tracrRNA sequences). Furthermore, in accordance with the present invention and as well known in the art, RNA motifs necessary for CRISPR nuclease binding and activity may be provided separately (e.g., (i) RNA guide sequence-crRNA CRISPR recognition sequence" (also known as crRNA) in one RNA molecule and (ii) a tracrRNA CRISPR recognition sequence on another, separate RNA molecule. Alternatively, all necessary RNA sequences (motifs) may be fused together in a single RNA guide. . The CRISPR recognition sequence is preferably fused directly to the gRNA guide sequence (in 3' (e.g., Cas9) or 5' (Cpf1) depending on the CRISPR nuclease used) but may include a spacer sequence separating two RNA motifs.

[0178] In embodiments, the CRISPR nuclease recognition sequence is a Cas9 recognition sequence having at least 65 nucleotides. In embodiments, the CRISPR nuclease recognition sequence is a Cas9 CRISPR nuclease recognition sequence having at least 85 nucleotides. In embodiments, the CRISPR nuclease recognition sequence is a Cpf1 recognition sequence (5' direct repeat) having about 19 nucleotides. In a particular embodiment, the Cas9 recognition sequence comprises (or consists of) the sequence as set forth in SEQ ID NO: 166. In a particular embodiment, the AsCpf1 recognition sequence comprises (or consists of) the sequence UAAUUUCUAC UCUUGUAGAU (SEQ ID NO: 168). The gRNA of the present invention may comprise any variant of the above noted sequences, provided that it allows for the proper functioning of the selected CRISPR nuclease (e.g., binding of the CRISPR nuclease protein to the *DYS* gene and/or target polynucleotide sequence(s)).

[0179] Together, the RNA guide sequence and CRSIPR nuclease recognition sequence(s) provide both targeting specificity and scaffolding/binding ability for the CRSIPR nuclease of the present invention. sgRNAs of the present invention do not exist in nature, i.e., is a non-naturally occurring nucleic acid(s).

[0180] In embodiments, the CRISPR nuclease (e.g., Cas9 or Cpf1) recognition sequence is a CRISPR nuclease recognition sequence having at least 65 nucleotides. In embodiments, the CRISPR nuclease recognition sequence is a CRISPR nuclease recognition sequence having at least 85 nucleotides.

[0181] As noted above not all CRISPR nucleases require a tracrRNA to function. Cpf1 is a single crRNA-guided endonuclease. Unlike Cas9, which requires both an RNA guide sequence (crRNA) and a tracrRNA (or a fusion or both crRNA and tracrRNA) to mediate interference, Cpf1 processes crRNA arrays independent of tracrRNA, and Cpf1-crRNA complexes alone cleave target DNA molecules, without the requirement for any additional RNA species (see Zetsche et al. (62)).

[0182] In embodiments, the gRNA may comprise a "G" at the 5' end of its polynucleotide sequence. The presence of a "G" in 5' is preferred when the gRNA is expressed under the control of the U6 promoter (Koo T. et al. (65)). The

CRISPR/nuclease system of the present invention may use gRNAs of varying lengths. The gRNA may comprise a gRNA guide sequence of at least 10 nts, at least 11 nts, at least a 12 nts, at least a 13 nts, at least a 14 nts, at least a 15 nts, at least a 16 nts, at least a 17 nts, at least a 18 nts, at least a 19 nts, at least a 20 nts, at least a 21 nts, at least a 22 nts, at least a 23 nts, at least a 24 nts, at least a 25 nts, at least a 30 nts, or at least a 35 nts of a target sequence in the *DYS* gene (such target sequence is followed or preceded by a PAM in the *DYS* gene but is not part of the gRNA). In embodiments, the "gRNA guide sequence" or "gRNA target sequence" may be least 10 nucleotides long, preferably 10-40 nts long (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nts long), more preferably 17-30 nts long, more preferably 17-22 nucleotides long. In embodiments, the gRNA guide sequence is 10-40, 10-30, 12-30, 15-30, 18-30, or 10-22 nucleotides long. In embodiments, the PAM sequence is "NGG", where "N" can be any nucleotide. In embodiments, the PAM sequence is NNGRRT or NNGRR(N), where "N" can be any nucleotide and R is A or G. In embodiments, the PAM sequence is "TTTN", where "N" can be any nucleotide. gRNAs may target any region of a target gene (e.g., *DYS*) which is immediately adjacent (contiguous, adjoining, in 5' or 3') to a PAM (e.g., NGG/TTTN or CCN/NAAA for a PAM that would be located on the opposite strand) sequence. In embodiments, the gRNA of the present invention has a target sequence which is located in an exon (the gRNA guide sequence consists of the RNA sequence of the target (DNA) sequence which is located in an exon). In embodiments, the gRNA of the present invention has a target sequence which is located in an intron (the gRNA guide sequence consists of the RNA sequence of the target (DNA) sequence which is located in an intron). In embodiments, the gRNA may target any region (sequence) which is followed (or preceded, depending on the CRISPR nuclease used) by a PAM in the *DYS* gene which may be used to restore its correct reading frame.

[0183] The number of sgRNAs administered to or expressed in a target cell in accordance with the methods of the present invention may be at least 1 gRNA, preferably at least two gRNAs.

[0184] Although a perfect match between the gRNA guide sequence and the DNA sequence on the targeted gene is preferred, a mismatch between a gRNA guide sequence and target sequence on the gene sequence of interest is also permitted as along as it still allows hybridization of the gRNA with the complementary strand of the gRNA target polynucleotide sequence on the targeted gene. A seed sequence of between 8-12 consecutive nucleotides in the gRNA, which perfectly matches a corresponding portion of the gRNA target sequence is preferred for proper recognition of the target sequence. The remainder of the guide sequence may comprise one or more mismatches. In general, gRNA activity is inversely correlated with the number of mismatches. Preferably, the gRNA of the present invention comprises 7 mismatches, 6 mismatches, 5 mismatches, 4 mismatches, 3 mismatches, more preferably 2 mismatches, or less, and even more preferably no mismatch, with the corresponding gRNA target gene sequence (less the PAM). Preferably, the gRNA nucleic acid sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% and 99% identical to the gRNA target polynucleotide sequence in the gene of interest (e.g., *DYS*). Of course, the smaller the number of nucleotides in the gRNA guide sequence the smaller the number of mismatches tolerated. The binding affinity is thought to depend on the sum of matching gRNA-DNA combinations.

[0185] Any gRNA guide sequence can be selected in the target gene, as long as it allows introducing at the proper

location, the desired modification(s) (e.g., insertions/deletions or selected target modification(s)). Accordingly, the gRNA guide sequence or target sequence of the present invention may be in coding or non-coding regions of the *DYS* gene (i.e., exons or introns). Of course the complementary strand of the sequence may alternatively and equally be used to identify proper PAM and gRNA target/guide sequences.

CRISPR nucleases

[0186] Recently, Tsai et al. (64). have designed recombinant dCas9-FoKI dimeric nucleases (RFNs) that can recognize extended sequences and edit endogenous genes with high efficiency in human cells. These nucleases comprise a dimerization-dependent wild type FokI nuclease domain fused to a catalytically inactive Cas9 (dCas9) protein. Dimers of the fusion proteins mediate sequence specific DNA cleavage when bound to target sites composed of two half-sites (each bound to a dCas9 (i.e., a Cas9 nuclease devoid of nuclease activity) monomer domain) with a spacer sequence between them. The dCas9-FoKI dimeric nucleases require dimerization for efficient genome editing activity and thus, use two gRNAs for introducing a cut into DNA.

[0187] The recombinant CRISPR nuclease that may be used in accordance with the present invention is i) derived from a naturally occurring nuclease (e.g., Cas or Cpf1 nuclease); and ii) has a nuclease (or nickase) activity to introduce a DSB in cellular DNA when in the presence of appropriate gRNA(s). Thus, as used herein, the term "CRISPR nuclease" refers to a recombinant protein which is derived from a naturally occurring nuclease which has nuclease activity and which functions with the gRNAs of the present invention to introduce DSBs in the targets of interest, e.g., the *DYS* gene. In embodiments, the CRISPR nuclease is spCas9. In embodiments, the CRISPR nuclease is SaCas9. In embodiments, the CRISPR nuclease is Cpf1. Exemplary CRISPR nucleases that may be used in accordance with the present invention are provided in Table 1 above. A variant of Cas9 can be a Cas9 nuclease that is obtained by protein engineering or by random mutagenesis (i.e., is non-naturally occurring). Such Cas9 variants remain functional and may be obtained by mutations (deletions, insertions and/or substitutions) of the amino acid sequence of a naturally occurring Cas9, such as that of *S. pyogenes* or *S. aureus*.

[0188] CRISPR nucleases such as Cas9 nucleases cut 3-4bp upstream of the PAM sequence. CRISPR nucleases such as Cpf1 on the other hand, generate a 5' overhang. The cut occurs 19 bp after the PAM on the targeted (+) strand and 23 bp on the opposite strand (62). There can be some off-target DSBs using wildtype Cas9. The degree of off-target effects depends on a number of factors, including: how closely homologous the off-target sites are compared to the on-target site, the specific site sequence, and the concentration of nuclease and guide RNA (gRNA). These considerations only matter if the PAM sequence is immediately adjacent to the nearly homologous target sites. The mere presence of additional PAM sequences should not be sufficient to generate off target DSBs; there needs to be extensive homology of the protospacer followed or preceded by PAM.

Optimization of codon degeneracy

[0189] Because CRISPR nuclease proteins are (or are derived from) proteins normally expressed in bacteria, it may be advantageous to modify their nucleic acid sequences for optimal expression in eukaryotic cells (e.g., mammalian cells) when designing and preparing CRISPR nuclease recombinant proteins.

[0190] Accordingly, the following codon chart (Table 2) may be used, in a site-directed mutagenic scheme, to produce nucleic acids encoding the same or slightly different amino acid sequences of a given nucleic acid:

Table 2: Codons encoding the same amino acid

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

Dystrophin

[0191] As used herein, the term “dystrophin gene” refers to the gene encoding the dystrophin protein (see accessions HGNC:2928, Entrez Gene: 1756; OMIM: 300377; Ensembl: ENSG00000198947; UniProtKB: P11532 and GenBank: NC_000023.11, the contents of which are herein incorporated by reference). The term “dystrophin gene” includes allelic variants of the gene which are found in nature (i.e., within subjects normally expressing the dystrophin protein). The dystrophin gene contains at least eight independent, tissue-specific promoters and two polyA-addition sites. Further, dystrophin RNA is differentially spliced, producing a range of different transcripts, encoding a large set of protein isoforms. The terms "dystrophyn protein", “dystrophin polypeptide”, "dystrophin RNA" and "dystrophin transcript" refer to any and all isoforms expressed by the dystrophin gene within cells of a subject (including any allelic variants).

[0192] The human Dystrophin gene measures about 2.4 Mb, and was identified through a positional cloning approach, based on the isolation of the gene responsible for Duchenne (DMD) and Becker (BMD) Muscular Dystrophies. In general, DMD patients carry mutations which cause premature translation termination (nonsense or frame shift mutations), while BMD patients carry mutations resulting in a dystrophin that is reduced either in size (from in-frame deletions) or in expression level.

[0193] Mutations in the DYS gene have also been linked to X-linked dilated cardiomyopathy and familial dilated cardiomyopathy. More than 30 mutations in the DMD gene can cause an X-linked form of familial dilated cardiomyopathy. This heart condition enlarges and weakens the cardiac muscle, preventing the heart from pumping blood efficiently. Although dilated cardiomyopathy is a sign of Duchenne and Becker muscular dystrophy (described above), X-linked dilated cardiomyopathy is typically not associated with weakness and wasting of skeletal muscles. The mutations that cause X-linked dilated cardiomyopathy preferentially affect the activity of dystrophin in cardiac muscle cells. As a result of these mutations, affected individuals typically have little or no functional dystrophin in the heart. Without enough of this protein, cardiac muscle cells become damaged as the heart muscle repeatedly contracts and relaxes. The damaged muscle cells weaken and die over time, leading to the heart problems characteristic of X-linked dilated cardiomyopathy. Generally, in X-linked dilated cardiomyopathy enough of this protein is present to prevent weakness and wasting of the skeletal muscles.

[0194] Familial dilated cardiomyopathy is characterized by a heart muscle which becomes thin and weakened in at least one chamber of the heart, causing the open area of the chamber to become enlarged (dilated). As a result, the heart is unable to pump blood as efficiently as usual. To compensate, the heart attempts to increase the amount of blood being pumped through the heart, leading to further thinning and weakening of the cardiac muscle. Over time, this condition results in heart failure. It usually takes many years for symptoms of familial dilated cardiomyopathy to cause health problems. They typically begin in mid-adulthood, but can occur at any time from infancy to late adulthood. Signs and symptoms of familial dilated cardiomyopathy can include an irregular heartbeat (arrhythmia), shortness of breath (dyspnea), extreme tiredness (fatigue), fainting episodes (syncope), and swelling of the legs and feet. In some cases, the first sign of the disorder is sudden cardiac death. The severity of the condition varies among affected individuals, even in members of the same family.

[0195] As indicated above, in a particular embodiment, the present invention uses the CRISPR system to introduce further mutations within exons or introns of a mutated dystrophin gene within a cell. The CRISPR system is a defense mechanism identified in bacterial species [37-42]. It has been modified to allow gene editing in mammalian cells. The modified system still uses a Cas9 nuclease to generate double-strand breaks (DSB) at a specific DNA target sequence [43, 44]. The recognition of the cleavage site is determined by base pairing of the gRNA with the target DNA and the presence of a trinucleotide called PAM (protospacer adjacent motif) juxtaposed to the targeted DNA sequence [45]. This PAM is NGG for the Cas9 of *S. pyogenes*, the most commonly used enzyme [46, 47].

[0196] In a particular embodiment, Applicants have used various combinations of two gRNAs targeting exons 46, and 51, 46 and 53, 49 and 52, 49 and 53, 47 and 58 and 50 and 54 and introns 22 and 23, of the *DYS* gene both *in*

vitro and *in vivo*. The *in vitro* experiments were done in 293T cells or in myoblasts of DMD patients having an endogenous frameshift or nonsense mutation (e.g., a deletion of exons 49-50, 51-53 or 51-56 generating a stop codon or frameshift). The *in vivo* experiments were done in the hDMD/mdx mouse that contains a full length human *DYS* gene or in the RAG/MDX mouse model comprising a mutation in exon 23. Results show that *in vitro* and *in vivo*, the gRNA combinations allowed precise DSB at 3 nucleotides upstream of the PAM and induced a large deletion. The junction between the remaining DNA sequences was achieved exactly as predicted. Using specifically selected pairs of gRNAs, it was possible to restore the reading frame resulting in the synthesis of an internally deleted DYS protein by the myotubes formed by the corrected myoblasts of DMD patients with an out-of-frame deletion. Such a CRISPR induced Deletion (CinDel) therapeutic approach can be used to restore directly *in vivo* the reading frame for most deletions observed in DMD patients. This approach is summarized in FIG. 10 for hybrid exon 50-54. Importantly, the deletion in the dystophin gene generated by the method of the present invention restores the correct reading frame for the protein in the endogenously mutated DYS gene and generates a shorter but functional dystrophin protein. The deletion allows maintaining the configuration of a normal spectrin-like repeat where hydrophobic amino acids are localized in position “a” and “d” of the heptad motif.

[0197] As indicated above, polypeptides (e.g., CRISPR nucleases) and nucleic acids encoding gRNAs and nucleases or nickases (e.g., Cas9 or Cpf1) of the present invention may be delivered into cells using various methods. These methods may employ one or more various viral vectors or delivery agents such as the peptide based shuttles (Feldan shulltles-discussed above).

[0198] Accordingly, preferably, the above-mentioned vector is a viral vector for introducing the gRNA and/or nuclease of the present invention in a target cell. Non-limiting examples of viral vectors include retrovirus, lentivirus, Herpes virus, adenovirus or Adeno Associated Virus, as well known in the art.

[0199] Modified AAV vector which preferably targets one or more cell types affected in DMD subjects are preferably used in accordance with the present invention. In an embodiment, the cell type is a muscle cell, in a further embodiment, a myoblast. Accordingly, the modified AAV vector may have enhanced cardiac, skeletal muscle, neuronal, liver, and/or pancreatic tissue (Langerhans cells) tropism. The modified AAV vector may be capable of delivering and expressing the at least one gRNA and nuclease of the present invention in the cell of a mammal. For example, the modified AAV vector may be an AAV-SASTG vector (Piacentino et al. (2012) Human Gene Therapy 23:635-646). The modified AAV vector may deliver gRNAs and nucleases to neurons, skeletal and cardiac muscle, and/or pancreas (Langerhans cells) *in vivo*. The modified AAV vector may be based on one or more of several capsid types, including AAVl, AAV2, AAV5, AAV6, AAV8, and AAV9. The modified AAV vector may be based on AAV2 pseudotype with alternative muscle-tropic AAV capsids, such as AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5 and AAV/SASTG vectors that efficiently transduce skeletal muscle or cardiac muscle by systemic and local delivery. In an embodiment, the modified AAV vector is a AAV-DJ. In an embodiment, the modified AAV vector is a AAV-DJ8 vector. In an embodiment, the modified AAV vector is a AAV2-DJ8 vector.

[0200] In yet another aspect, the present invention provides a cell (e.g., a host cell) comprising the above-mentioned nucleic acid and/or vector. The invention further provides a recombinant expression system, vectors and

host cells, such as those described above, for the expression/production of a recombinant protein, using for example culture media, production, isolation and purification methods well known in the art.

[0201] In another aspect, the present invention provides a composition (e.g., a pharmaceutical composition) comprising the above-mentioned gRNA and/or CRISPR nuclease (e.g., Cas9 or Cpf1), or nucleic acid(s) encoding same or vector(s) comprising such nucleic acid(s). In an embodiment, the composition further comprises one or more pharmaceutically acceptable carriers, excipients, and/or diluents.

[0202] As used herein, "pharmaceutically acceptable" (or "biologically acceptable") refers to materials characterized by the absence of (or limited) toxic or adverse biological effects *in vivo*. It refers to those compounds, compositions, and/or dosage forms which are, within the scope of sound medical judgment, suitable for use in contact with the biological fluids and/or tissues and/or organs of a subject (e.g., human, animal) without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio.

[0203] The present invention further provides a kit or package comprising at least one container means having disposed therein at least one of the above-mentioned gRNAs, nucleases, vectors, cells, targeting systems, combinations or compositions, together with instructions for restoring the correct reading frame of a *DYS* gene in a cell or for treatment of DMD in a subject.

[0204] The present invention is illustrated in further details by the following non-limiting examples.

EXAMPLE 1

MATERIALS AND METHODS FOR EXAMPLES 1 TO 7

[0205] Identification of targets and gRNA cloning. The plasmid pSpCas(BB)-2A-GFP (pX458) (Addgene plasmid # 48138) (FIG. 1a) [58] containing two BbsI restriction sites necessary for insertion of a protospacer (see below) under the control of the U6 promoter was used in the experiments shown in FIGs. 1 to 8. The pSpCas(BB)-2A-GFP plasmid also contains the Cas9, of *S. pyogenes,* and eGFP genes under the control of the CBh promoter; both genes are separated by a sequence encoding the peptide T2A.

[0206] The nucleotide sequences targeted by the gRNAs in exons 50 and 54 were identified using the Leiden Muscular Dystrophy website by screening for Protospacer Adjacent Motifs (PAM) in the sense and antisense strands of each exon sequence (FIG. 1b). The PAM sequence for *S. pyogenes* Cas9 is NGG. An oligonucleotide coding for the target sequence, and its complementary sequence, were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) and cloned into BbsI sites as protospacers leading to the individual production of 10 gRNAs targeting exon 50 and 14 gRNAs targeting exon 54, according to Addgene's instructions. Briefly, the oligonucleotides were phosphorylated using T4 PNK (NEB, Ipwisch, MA) then annealed and cloned into the BbsI sites of the plasmid pSpCas(BB)-2A-GFP using the Quickligase (NEB, Ipwisch, MA). Following clone isolation and DNA amplification, samples were sequenced using the primer U6F (5'-GTCGGAACAGGAGAGCGCACGAGGGAG) (SEQ ID NO: 173) and sequencing results were analyzed using the NCBI BLAST platform (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

[0207] Cell Culture. Transfection of the expression plasmid in 293T cells and in DMD patient myoblasts.

[0208] The gRNA activities were tested individually or in pairs by transfection of the pSpCas(BB)-2A-GFP-gRNA plasmid encoding each gRNA in 293T cells and in DMD myoblasts having a deletion of exons 51 to 53. The 293T cells were grown in Dulbecco's modified Eagle medium (DMEM) medium (Invitrogen, Grand Island, NY) containing 10% fetal bovine serum (FBS) and antibiotics (penicillin 100 U/ml / streptomycin 100 µg/ml). DMD patient myoblasts were grown in MB-1 medium (Hyclone, Thermo Scientific, Logan, UT) containing 15% FBS, without antibiotics. Cells in either 24-well or 6-well plates were transfected at 70-80% confluency using respectively 1 or 5 µg of plasmid DNA and 2 or 10 µl of Lipofectamine™ 2000 (Invitrogen, Carlsbad, CA) previously diluted in Opti-Mem™ (Invitrogen, Grand Island, NY). For gRNA pair transfection, half of the DNA mixture was coming from the plasmid encoding the gRNA-50 and half from the gRNA-54. The cells were incubated at 37°C in the presence of 5% CO_2 for 48 hours. The transfection success was evaluated by the GFP expression in the transfected cells under microscopy with a Nikon TS 100 (Eclipse, Japan).

[0209] Myoblast transfection with Lipofectamine™ 2000 following the previous standard protocol was not sufficiently effective and was improved as follows. The MB-1 medium was aspirated before transfection and myoblasts were washed once with 500 µl of 1X Hanks Balanced Salt Solution (HBSS) (Invitrogen, Grand Island, NY). The complex Lipofectamine 2000 plasmid DNA (diluted in Opti-Mem™ as above) was then poured directly on cells, instead of being in media, and the cells/DNA complex was incubated at 37°C during 15 min. After this time, the antibiotic-free medium was added to the cells and the plate was returned to the incubator for 18-24 hours. After that time, the medium was aspired and replaced with the fresh medium. The plate was incubated for another 24 hours.

[0210] Myoblast differentiation in myotubes and dystrophin expression. The DMD myoblasts (transfected with gRNA2-50 and gRNA2-54) were allowed to fuse in myotubes to induce the expression of dystrophin. To permit this myoblast fusion, the MB-1 medium (Hyclone, Thermo Scientific, Logan, UT) was aspirated from the myoblast culture and replaced by the minimal DMEM medium containing 2% FBS (Invitrogen, Grand Island, NY). Myoblasts were incubated at 37°C in 5% CO_2 for 7 days. Untransfected myoblasts (negative control) of the DMD patient and immortalized wild-type myoblasts from a healthy donor (positive control) were also grown under the same conditions to induce their differentiation in myotubes.

[0211] Genomic DNA extraction and analysis. Forty-eight (48) hours after transfection with the pSpCas(BB)-2A-GFP-gRNA plasmid(s), the genomic DNA was extracted from the 293T or myoblasts using a standard phenol-chloroform method. Briefly, the cell pellet was resuspended in 100 µl of lysis buffer containing 10% sarcosyl and 0.5 M pH 8 ethylenediaminetetraacetic acid (EDTA). Twenty (20) µl of proteinase K (10 mg/ml) were added. The suspension was mixed by up down and incubated 10 min at 55°C. It was then centrifuged at 13200 rpm for 2 min. The supernatant was collected in a new microfuge tube. One volume of phenol-chloroform was added and following centrifugation, the aqueous phase was recovered in a new microfuge tube and ethanol-precipitated with 1:10 volume:volume of NaCl 5 M and two volumes of 100% ethanol. The pellet was washed with 70% ethanol, centrifuged and the DNA was resuspended in 50 µl of double-distilled water. The genomic DNA concentration was assayed with a NanoDrop™ (Thermo Scientific, Logan, UT).

[0212] To confirm the successful individual cuts or deletions, exons 50 and 54 and the hybrid exon 50-54 were then amplified by PCR. For exon 50, the sense primer targeted the end of intron 49 (called Sense 49 5'-TTCACCAAATGGATTAAGATGTTC) (SEQ ID NO: 174) and the antisense primer targeted the start of intron 50 (called Antisense 50 5'- ACTCCCCATATCCCGTTGTC) (SEQ ID NO: 175). For exon 54, the forward and reverse primers targeted respectively the end of the intron 53 (called Sense 53 5'- GTTTCAAGTGATGAGATAGCAAGT) (SEQ ID NO: 176) and the start of intron 54 (called Antisense 54 5'-TATCAGATAACAGGTAAGGCAGTG) (SEQ ID NO: 177). For the hybrid exon 50-54, the forward Sense 49 and reverse Antisense 54 were used. All PCR amplifications were performed in a thermal cycler C1000 Touch of BIO RAD (Hercules, CA) with the Phusion high fidelity polymerase (Thermo scientific, EU, Lithuania) using the following program for exon 50, exon 54 and the hybrid exon 50-54: 98°C / 10 sec, 58°C / 20 sec, 72°C / 1 min for 35 cycles.

[0213] The amplicons of individual exons 50 and 54 were used to perform the Surveyor assay. The first part of the test was the hybridization of amplicons using the slow-hybridization program (denaturation at 95°C followed by gradual cooling of the amplicons) with BIO RAD thermal cycler C1000Touch (Hercules, CA). Subsequently, the amplicons were digested with nuclease Cel (Integrated DNA Technologies, Coralville, IA) in the thermal cycler at 42°C for 25 min. The digestion products were visualized on agarose gel 1.5%

[0214] Cloning and sequencing of the hybrid exons. The amplicon of hybrid exons obtained by the amplification of genomic DNA extracted from 293T cells or myoblasts transfected with 2 different pSpCas(BB)-2A-GFP-gRNAs was purified by gel extraction (Thermo Scientific, EU, Lithuania). The bands of about 480 to 655 bp were cloned into the linearized cloning vector pMiniT (NEB, Ipwisch, MA). On day 3, the plasmid DNA was extracted with the Miniprep Kit (Thermo Scientific, EU, Lithuania) and the cloning vector was digested simultaneously with EcoRI and PstI to confirm the insertion of the amplicon. In the cloning vector pMiniT, the insert was flanked by two EcoRI restriction sites. Digestion with EcoRI generated two fragments of 2500 bp (plasmid without insert) and of 480 to 655 bp (amplicon inserted). It should be noted that there was a PstI restriction site in the remaining part of exon 54. A PstI digestion generated two fragments. The clones, which gave after double digestion with EcoRI and PstI these two fragments, were sent for sequencing using primers provided by the manufacturer (NEB, Ipwisch, MA). Sequencing results were analyzed with the NCBI BLAST platform (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the Expert Protein Analysis System (ExPASy) platform (htt://www.expasy.org). This software allowed the visualization both the nucleotide sequences of the hybrid exon 50-54 and of the corresponding amino acid sequences.

[0215] *In vivo* mouse assay. Sperm from transgenic hDMD mice expressing the full-length human dystrophin gene were inseminated [59]. The hDMD mice were crossed with mdx mice to produce the hDMD/mdx mice.

[0216] Forty (40) µg of pSpCas-2A-GFP-gRNAs (20 µg gRNA2-50 and 20 µg gRNA2-54) were suspended in 20 µl of double-distilled water and mixed with 20 µl of Tyrode's buffer (119 mM NaCl, 5 mM KCl, 25 mM HEPES buffer, 2 mM $CaCl_2$ 2 mM $MgCl_2$, 6 g/L glucose, pH was adjusted to 7.4 with NaOH, Sigma-Aldrich). The hDMD/mdx mice were electrotransferred with an Electro Square Porator (Model ECM630, BTX Harvard Apparatus, St-Laurent, Canada) following a single transcutaneous longitudinal injection in the *Tibialis anterior* (TA) of the pSpCas(BB)-2A-GFP plasmids. An electrode electrolyte cream (Teca, Pleasantville, NY) was applied on the skin to favor the

passage of the electric field between the two electrode plates. Muscles were submitted to electric field (8 pulses of 20 ms duration spaced by 1 s). The voltage was adjusted at 100 volts/cm depending the width of the mice leg. Electroporated and control mice were sacrificed 7 days later. Genomic DNA was extracted with phenol-chloroform method as above and DNA analysis performed as previously described.

[0217] Protein analysis. Myotubes were harvested and proteins were extracted with the methanol-chloroform method. Briefly, cell pellets were resuspended in lysis buffer containing 75 mM Tris-HCl pH 7.4, 1 mM dithiotreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1% sodium dodecyl sulfate (SDS). Protein extracts were dried with the speed vacuum Univapo 100 ECH (Uniequip, Martinsried, Germany) to remove all traces of methanol. Samples were then diluted in a buffer containing 0.5% mercaptoethanol and heated at 95°C for 5 min. The protein concentrations were assayed by Amido Black using Imager2200 AlphaDigiDoc (Alpha Innotech, Fisher Scientific, Suwanee, GA).

[0218] Seventy-five (75) µg of protein of each sample were separated on a 7% polyacrylamide gel and transferred onto nitrocellulose membrane at 4°C for 16 hrs. In order to detect dystrophin on the membrane, a primary mouse monoclonal antibody (cat# NCL-DYS2, Leica Biosystems, Newcastle, UK) recognizing the C-terminus of the human dystrophin was used. The antibody was diluted 1:25 in 0.1X PBS containing 5% milk and 0.05% Tween20 and incubated at 4°C for 16 hrs.

EXAMPLE 2

DYSTROPHIN EXON TARGETING IN DMD MYOBLASTS USING THE SpCAS9/CRISPR SYSTEM

[0219] Twenty-four different pSpCas(BB)-2A-GFP-gRNA plasmids (FIG. 1a) were made: 10 containing gRNAs targeting different sequences of the exon 50 of the *DYS* gene and 14 containing gRNAs targeting the exon 54 (Table 3 and FIGs. 1b-c). To test the activity of these gRNAs, these plasmids were first transfected in 293T cells. Under standard transfection conditions, 80% of cells showed expression of the GFP confirming the effectiveness of the transfection (FIG. 2a). The DNA from those cells was extracted 48 hours after transfection. The exon 50 of the *DYS* was amplified by PCR using primers Sense 49 and Antisense 50 and exon 54 was amplified with primers Sense 53 and Antisense 54 (see Example 1 for details on primer sequences). The presence of INDELs, produced by non-homologous end-joining (NHEJ) following the DSBs generated by the gRNAs and the Cas9, was detected using the Surveyor/Cel I enzymatic assay (Figs. 3a-b). An expected pattern of three bands was detected with most gRNAs; the upper band representing the uncut PCR product and the two lowest bands the Cel I products whose lengths are related to the guide used to induce the DSB.

Table 3: Exemplary gRNAs targeting exons 50 and 54 of the DYS gene

gRNA#	Exon	Strand (AS=Anti-Sense)	Target sequence	SEQ ID NOs Target/gRNA	Cut sites in *DYS* gene	Cut sites in amino acid sequence
gRNA1-50	50	Sense	TAGAAGATCTGAGCTCTGAG	82/124	7224-7225	2408 TCT (Ser): 2409 GAG (Glu)
gRNA2-50	50	Sense	AGATCTGAGCTCTGAGTGGA	83/125	7228-7229	2410 T: GG (Trp)
gRNA3-50	50	Sense	TCTGAGCTCTGAGTGGAAGG	84/126	7231-7232	2411 A: AA (Lys)
gRNA4-50	50	Sense	CCGTTTACTTCAAGAGCTGA	85/127	7258-7259	2420 C: TG (Leu)
gRNA5-50	50	Sense	AAGCAGCCTGACCTAGCTCC	86/128	7283-7284	2428 GC: T (Ala)
gRNA6-50	50	Sense	GCTCCTGGACTGACCACTAT	87/129	7298-7299	2433 AC: T (Thr)
gRNA7-50	50	AS	CCCTCAGCTCTTGAAGTAAA	88/130	7247-7248	2416 TT: A (Leu)
gRNA8-50	50	AS	GTCAGTCCAGGAGCTAGGTC	89/131	7278-7279	2426 GAC (Asp): 2427 CTA (Leu)
gRNA9-50	50	AS	TAGTGGTCAGTCCAGGAGCT	90/132	7283-7284	2428 GC:T (Ala)
gRNA10-50	50	AS	GCTCCAATAGTGGTCAGTCC	91/133	7290-7291	2430 GGA (Gly): 2431 CTG (Leu)
gRNA1-54	54	Sense	TGGCCAAAGACCTCCGCCAG	92/134	7893-7894	2631 CGC (Arg): 2632 CAG (Gln)
gRNA2-54	54	Sense	GTGGCAGACAAATGTAGATG	93/135	7912-7913	2638 G:AT (Asp)
gRNA3-54	54	Sense	TGTAGATGTGGCAAATGACT	94/136	7924-7925	2642 G:AC Asp)
gRNA4-54	54	Sense	CTTGGCCCTGAAACTTCTCC	95/137	7941-7942	2648 C:TC (leu)
gRNA5-54	54	Sense	CAGAGAATATCAATGCCTCT	96/138	8004-8005	2668 GCC (Ala):2669 TCT (Ser)
gRNA6-54	54	AS	CTGCCACTGGCGGAGGTCTT	97/139	7885-7886	2629 G:AC (Asp)

SUBSTITUTE SHEET (RULE 26)

gRNA#	Exon	Strand (AS=Anti-Sense)	Target sequence	SEQ ID NOs Target/gRNA	Cut sites in *DYS* gene	Cut sites in amino acid sequence
gRNA7-54	54	AS	CATTTGTCTGCCACTGGCGG	98/140	7892-7893	2631 CG:C (Arg)
gRNA8-54	54	AS	CTACATTTGTCTGCCACTGG	99/141	7895-7896	2632 CA:G (Gln)
gRNA9-54	54	AS	CATCTACATTTGTCTGCCAC	100/142	7898-7899	2633 TG:G (Trp)
gRNA10-54	54	AS	ATAATCCCGGAGAAGTTTCA	101/143	7936-7937	2646 A:AA (Lys)
gRNA11-54	54	AS	TATCATCTGCAGAATAATCC	102/144	7949-7950	2650 GA:T (Asp)
gRNA12-54	54	AS	TGTTATCATGTGGACTTTTC	103/145	7972-7973	2658 A:AA (Lys)
gRNA13-54	54	AS	TGATATATCATTTCTCTGTG	104/146	7982-7983	2661 AT:G (Met)
gRNA14-54	54	AS	TTTATGAATGCTTCTCCAAG	105/147	8008-8009	2670 T:GG (Trp)

[0220] The gRNAs were also subsequently tested individually in immortalized myoblasts from a DMD patient having a deletion of exons 51 through 53. Unfortunately, transfection efficiency was very low in myoblasts under the standard Lipofectamine™ 2000 transfection [14] (FIG. 2b). However, the protocol was improved and we were able to see approximately 20 to 25% of myoblasts expressing GFP (FIG. 2c). The Surveyor assay revealed the presence of INDELs in amplicons of exons 50 (FIG. 3c) and 54 (FIG. 3d) obtained from these myoblasts.

EXAMPLE 3

TESTING OF gRNA PAIRS IN THE SpCAS9/CRISPR SYSTEM

[0221] Given that the CRISPR/Cas9 induces a DSB at exactly 3 bp from the PAM in the 5' direction, it was possible to predict the consequence of cutting of the exons 50 and 54 with the various pairs of gRNAs. This analysis predicted four possibilities, as illustrated in FIG. 4a and detailed in Table 4: 1) the total number of coding nucleotides, which are deleted (i.e., the sum of the nucleotides of exons 51, 52 and 53 and the portions of exons 50 and 54, which are deleted) is a multiple of three and the junction of the remains of 50 exons and 54 does not generate a new codon, 2) the number of deleted nucleotides coding for DYS is a multiple of three but a new codon, derived from the junction of the remains of 50 exons and 54, encodes a new amino acid, 3) the number of coding nucleotides, which are deleted is not a multiple of three resulting in an incorrect reading frame of the *DYS* gene; and 4) the sum of deleted nucleotides coding for DYS is a multiple of three, but the new codon, formed by the junction of the remaining parts of exons 50 and 54, is a stop codon.

CA 03037635 2019-03-20

Table 4: Hybrid exon junctions following first and second cuts in exons 50 and 54 with various gRNA pairs

Combination	End of	Beginning of	Observation	New codon	New amino acid
	Exon 50 remain	Exon 54 remain		generated	generated
gRNA1 Ex 50/gRNA1 Ex 54	Ser 2408	Gln 2632	Junction Ser 2408-Gln 2632	None	None
gRNA1 Ex 50/gRNA5 Ex 54	Ser 2408	Ser 2669	Junction Ser 2408-Ser2669	None	None
gRNA2 Ex 50/gRNA2 Ex 54	T	AT	T+AT=TAT	TAT	Tyr
gRNA2 Ex 50/gRNA3 Ex 54	T	AC	T+AC=TAC	TAC	Tyr
gRNA2 Ex 50/gRNA6 Ex 54	T	AC	T+AC=TAC	TAC	Tyr
gRNA2 Ex 50/gRNA 14 Ex 54	T	GG	T+GG=TGG	TGG	Trp
gRNA3 Ex 50/gRNA2 Ex 54	A	AT	A+AA=AAT	AAT	Asn
gRNA3 Ex 50/gRNA3 Ex 54	A	AC	A+AC=AAC	AAC	Asn
gRNA3 Ex 50/gRNA6 Ex 54	A	AC	A+AC=AAC	AAC	Asn
gRNA3 Ex 50/gRNA10 Ex 54	A	AA	A+AA=AAA	AAA	Lys
gRNA3 Ex 50/gRNA12 Ex 54	A	AA	A+AA=AAA	AAA	Lys
gRNA3 Ex 50/gRNA14 Ex 54	A	GG	A+GG=AGG	AGG	Arg
gRNA4 Ex 50/gRNA2 Ex 54	C	AT	C+AT=CAT	CAT	His
gRNA4 Ex 50/gRNA3 Ex 54	C	AC	C+AC=CAC	CAC	His
gRNA4 Ex 50/gRNA6 Ex 54	C	AC	C+AC=CAC	CAC	His
gRNA4 Ex 50/gRNA 10 Ex 54	C	AA	C+AA=CAA	CAA	Gln
gRNA4 Ex 50/gRNA12 Ex 54	C	AT	C+AT=CAT	CAT	His
gRNA4 Ex 50/gRNA14 Ex 54	C	GG	C+GG=CGG	CGG	Arg

SUBSTITUTE SHEET (RULE 26)

Combination	End of	Beginning of	Observation	New codon	New amino acid
	Exon 50 remain	Exon 54 remain		generated	generated
gRNA5 Ex 50/gRNA7 ex 54	GC	C	GC+C=GCC	GCC	Ala
gRNA5 Ex 50/gRNA 8ex 54	GC	G	GC+G=GCG	GCG	Ala
gRNA5 EX 50/gRNA9 Ex 54	GC	G	GC+G=GCG	GCG	Ala
gRNA5 Ex 50/gRNA11 Ex 54	GC	T	GC+T=GCT	GCT	Ala
gRNA5 Ex50/gRNA13 EX 54	GC	G	GC+G=GCG	GCG	Ala
gRNA6 Ex 50/ gRNA7 Ex 54	AC	C	AC+C=ACC	ACC	Thr
gRNA6 Ex 50/ gRNA8 Ex 54	AC	G	AC+G=ACG	ACG	Thr
gRNA6 Ex 50/gRNA9 Ex 54	AC	G	AC+G=ACG	ACG	Thr
gRNA6 Ex 50/gRNA11 Ex 54	AC	T	AC+T=ACT	ACT	Thr
gRNA6 Ex 50/gRNA13 Ex 54	AC	G	AC+G=ACG	ACG	Thr
gRnA7 Ex 50/gRNA7 Ex 54	TT	C	TT+C=TTC	TTC	Phe
gRNA7 Ex 50/gRNA8 Ex 54	TT	G	TT+G=TTG	TTG	Leu
gRNA7 Ex 50/gRNA9 Ex 54	TT	G	TT+G=TTG	TTG	Leu
gRNA7 Ex 50/gRNA11 Ex 54	TT	T	TT+T=TTT	TTT	Phe
gRNA7 Ex 50/ gRNA13 Ex 54	TT	G	TT+G=TTG	TTG	Leu
gRNA8 Ex 50/gRNA1 Ex 54	Asp2426	Gln2632	Junction Asp2426-Gln2632	None	None
gRNA8 Ex 50/gRNA5 Ex 54	Asp2426	Ser 2669	Junction Asp2426-Ser2669	None	None
gRNA9 Ex 50/gRNA7 Ex 54	GC	C	GC+C=GCC	GCC	Ala
gRNA9 eEx 50/gRNA8 Ex 54	GC	G	GC+G=GCG	GCG	Ala
gRNA9 Ex 50/gRNA9 Ex 54	GC	G	GC+G=GCG	GCG	Ala
gRNA9 Ex 50/gRNA11 Ex 54	GC	T	GC+T=GCT	GCT	Ala

SUBSTITUTE SHEET (RULE 26)

Combination	End of	Beginning of	Observation	New codon	New amino acid
	Exon 50 remain	Exon 54 remain		generated	generated
gRNA9 Ex 50/gRNA13 Ex 54	GC	G	GC+G=GCG	GCG	Ala
gRNA10 Ex 50/gRNA1 Ex 54	Gly2430	Gln2632	Junction Gly 2430-Gln2632	None	None
gRNA10 Ex 50/gRNA5 Ex 54	Gly2430	Ser 2669	Junction Gly 2430-Ser2669	None	None

[0222] The deletion of part of the *DYS* gene was investigated by transfecting 293T cells and human myoblasts with different pairs of plasmids encoding gRNAs: one targeting exon 50 and the other the exon 54 (FIGs. 4b and 4c). To detect successful deletions, genomic DNA was extracted from these transfected and non-transfected cells 48 hours later and amplified by PCR using primers Sense 49 and Antisense 54 (see Example 1 for details regarding primer sequences). No amplification was obtained from DNA extracted from untransfected cells (FIG. 4c, lanes 1 and 6) because of the expected amplicon size (about 160 Kbp) of the wild-type *DYS* gene (i.e., exon 50 to exon 54) is too big. However, amplicons, named hybrid exons, of the expected sizes were obtained when a pair of gRNAs was used (FIG. 4b, lanes 2-5 and lanes 7-10), confirming the excision of the 160 Kbp sequence in 293T cells.

[0223] As shown in FIG. 4b, several different gRNA pairs (targeting exons 50 and 54) were tested and all produced exactly the expected modification of the *DYS* gene according to the four possibilities explained above.

EXAMPLE 4

CHARACTERIZATION OF HYBRID EXON 50-54 IN 293T CELLS (SpCAS9/CRISPR SYSTEM)

[0224] The amplicons obtained following transfection of the gRNA pairs were gel purified and cloned into the pMiniT plasmid, transformed in bacteria and clones were screened for successful insertions. Positive clones, according to the digestion pattern, were sent for sequencing to demonstrate the presence of a hybrid exon formed by the fusion of a part of exon 50 with a portion of exon 54. For example, in 100% (7/7) of sequences obtained for the gRNA5-50 and gRNA1-54 pair, the DYS gene was cut in both exons at exactly 3 nucleotides in the 5' direction from the PAM (data not shown). This exercise was repeated with different pairs of gRNAs and for each functional gRNA pair, the CinDel technique removed successfully a portion of about 160 100 bp in the DYS gene of 293T cells.

EXAMPLE 5

CHARACTERIZATION OF HYBRID EXON 50-54 IN MYOBLASTS (SpCAS9/CRISPR SYSTEM)

[0225] We also wanted to confirm the accuracy of cuts produced by the Cas9 from our expression plasmids in the myoblasts of a DMD patient already having a deletion of exons 51 to 53. We thus transfected the gRNA 2-50 and gRNA 2-54 pair previously characterized to produce a deletion in the *DYS* gene restoring the reading frame. As control, we also used another gRNA pair (i.e., gRNA5-50 and gRNA1-54) that should not restore the reading frame. As in 293T, genomic DNA of these myoblasts was extracted 48 hours later and amplified with primers Sense 49 and Antisense 54 and amplicons were cloned into the plasmid pMiniT. The plasmids were extracted from bacterial clones, screened according to their digestion pattern and positive clones were sequenced. The sequences of 45 clones were analyzed for the gRNA2-50 and gRNA2-54 pair and the most abundant product (25/45, i.e. 56%) contained exactly the expected junction between the remaining parts exons 50 and 54 to produce a 141 bp hybrid exon (FIGs. 5a and 5b). For 60% (27/45), a new codon (Y) was created (FIGs. 5a and 5b). A percentage of 62% (28/35) was detected as in-frame hybrid exons (FIG. 5b) and 38% (17/45) as out-of-frame hybrid exons (FIG. 5b).

[0226] For the second gRNA pair (gRNA5-50 and gRNA1-54), the plasmids were extracted from eight bacterial clones and sequenced. The sequence of these clones also demonstrated that 75% (6 out of 8) of these hybrid exons

50-54 (amplicon 655 bp) contained the expected reading frame shift. One of the two remaining clones showed a 1 bp insertion in addition of the expected deletion, this restored the *DYS* reading frame. Another clone showed an additional deletion of 11 bp that did not restore the reading frame.

EXAMPLE 6

IN VIVO CORRECTION IN THE HDMD/*MDX* MOUSE (SpCAS9/CRISPR SYSTEM)

[0227] As the CinDel method was effective in 293T cells and in DMD myoblasts in culture, plasmids coding for a pair of gRNAs were electroporated in the *Tibialis anterior* (TA) of a hDMD/*mdx* mouse to confirm CinDel effects *in vivo*. Genomic DNA was extracted 7 days later from the gRNA2-50/2-54 electroporated TA and from a non-electroporated TA. Exons 50 and 54 of the human dystrophin gene were PCR amplified. We were able to detect additional bands following digestion of the amplicon of these exons by the Cell enzyme of the Surveyor assay (FIG. 6a, CinDel lanes). These results confirmed that both gRNAs were able to induce mutations of their targeted exon *in vivo*. Moreover, the hybrid exon 50-54 was also PCR amplified (FIG. 6b, lane 3) demonstrating that both gRNAs were able to cut simultaneously *in vivo* leading to a deletion of more than 160 kb. The amplicons of the hybrid exon 50-54 were cloned in bacteria and 11 clones were sequenced. The sequences of 7 of these clones were the same as those of the obtained for *in vitro* experiments with the same gRNA pair (FIG. 5b), thus 64% (7 out of 11) of the sequences showed a correct restoration of the reading frame ***in vivo*.**

EXAMPLE 7

DYS EXPRESSION IN MYOTUBES FORMED BY GENETICALLY CORRECTED MYOBLASTS (SpCAS9/CRISPR SYSTEM)

[0228] In order to verify whether the CinDel gene therapy method was efficient in restoring the expression of the DYS protein, DMD myoblasts transfected with gRNA2-50 and gRNA2-54 were differentiated into myotubes *in vitro*. The proteins from the resulting myotubes (FIG. 7a) were extracted after 7 days in the fusion medium. A western blot confirmed the presence of a truncated (Trunc.) DYS protein with a molecular weight of about 400 kDa (FIG. 7b, lane 3). The size of this protein corresponds to the weight expected in the absence of exons 51-53 and of portions of exons 50 and 54, while the molecular weight of the full-length (FL) DYS protein is 427 kDa in normal myotubes (FIG. 7b, lane 2). No DYS protein was detected in proteins extracted from the DMD myotubes that had not been genetically corrected (FIG. 7b, lane 1). This result indicates that myotubes formed *in vitro* by myoblasts of a DMD patient in which the reading frame has been restored by the CinDel are able to express an internally truncated DYS protein.

EXAMPLE 8

MATERIALS AND METHODS FOR EXAMPLES 9 TO 23

[0229] Identification of targets and gRNA cloning. The plasmid pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA (Addgene plasmid # 61591; SEQ ID NO: 167) containing two BsaI restriction sites necessary for insertion of a protospacer (see below) under the control of the U6 promoter was used in our study.

The pX601 plasmid also contains the Cas9 of *S. aureus*.

[0230] The nucleotide sequences targeted by the gRNAs along exons 46 and 58 were identified using the benchling software website by screening for Protospacer Adjacent Motifs (PAM) in the sense and antisense strands of each exon sequence. The PAM sequence for *S. aureus* Cas9 is NNGRRT. An oligonucleotide coding for the target sequence, and its complementary sequence, were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) and cloned into BsaI sites as protospacers leading to the individual production of 2 gRNAs targeting exon 46, 3 gRNAs targeting exon 47, 1 gRNA targeting exon 49, 2 gRNAs targeting exon 51, 2 gRNAs targeting exon 52, 5 gRNAs targeting exon 53 and 3 gRNAs targeting exon 58 (see Table 6 below for sequences), according to Addgene's instructions. Briefly, the oligonucleotides were phosphorylated using T4 PNK (NEB, Ipwisch, MA) then annealed and cloned into the BsaI sites of the plasmid pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA using the Quickligase (NEB, Ipwisch, MA). Following clone isolation and DNA amplification, samples were sequenced using the primer U6F2 (5' GAGGGCCTATTTCCCATGATT 3') (SEQ ID NO: 178) and sequencing results were analyzed using the CLC Sequence Viewer software (CLC Bio).

Cell Culture. Transfection of the expression plasmid in 293T cells and in DMD patient myoblasts.

[0231] The gRNA activities were tested individually or in pairs by transfection of the pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA plasmid encoding each gRNA in 293T cells and in DMD myoblasts having a deletion of exons 49 to 50 or a deletion of exons 51 to 53, or a deletion of exons 51 to 56. The 293T cells were grown in Dulbecco's modified Eagle medium (DMEM) medium (Invitrogen, Grand Island, NY) containing 10% fetal bovine serum (FBS) and antibiotics (penicillin 100 U/ml / streptomycin 100 µg/ml). DMD patient myoblasts were grown in MB-1 medium (Hyclone, Thermo Scientific, Logan, UT) containing 15% FBS, without antibiotics.

[0232] 293T in 24-well were transfected at 70-80% confluency using respectively 1 µg of plasmid DNA and 3 µl of Lipofectamine™ 2000 (Invitrogen, Carlsbad, CA) previously diluted in Opti-Mem™ (Invitrogen, Grand Island, NY). For gRNA pair transfection, half of the DNA mixture was coming from the plasmid encoding a gRNA with a target sequence upstream of exon 50 and half from a gRNA with a target sequence downstream of exon50. The cells were incubated at 37°C in the presence of 5% CO_2 for 48 hours.

[0233] Myoblast were transfected at 60-70% confluency in 6-well plates using 5 µg of plasmid DNA and 2 µL of TransfeX™ transfection reagent (ATCC® ACS-4005™) previously diluted in Opti-MEM™. The MB-1 medium was replaced by fresh medium before transfection. The complex TransfeX plasmid DNA (diluted in Opti-Mem™ as above) was then poured on cells, and the cells/DNA complex was incubated at 37°C overnight followed by replacement of culture medium with the fresh MB-1. Cells sere incubated at 37°C in the presence of 5% CO_2 for 48 hours.

[0234] Genomic DNA extraction and analysis. Forty-eight (48) hours after transfection with the pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA plasmid(s), the genomic DNA was extracted from the 293T or myoblasts using a standard phenol-chloroform method. Briefly, the cell pellet was resuspended in 100 µl of lysis buffer containing 10% sarcosyl and 0.5 M pH 8 ethylene diamine tetra acetic acid (EDTA). Twenty (20) µl of

proteinase K (10 mg/ml) were added. The suspension was mixed by up down and incubated 10-15 min at 55°C. Suspension was then centrifuged at 13200 rpm for 5 min. The supernatant was collected in a new microfuge tube. One volume of phenol-chloroform was added and following centrifugation, the aqueous phase was recovered in a new microfuge tube. Then DNA was precipitated using 1/10 volume of NaCl 5 M and two volumes of 100% ethanol followed by 5 min centrifugation ate 13200rpm. The pellet was washed with 70% ethanol, centrifuged and the DNA was resuspended in double-distilled water. The genomic DNA concentration was assayed with a NanoDrop™ spectrophotometer (Thermo Scientific, Logan, UT).

[0235] To confirm the successful individual cuts or deletions, exons 46, 47, 49, 51, 52, 53, 58 and the hybrid exon 46-51, 46-53, 49-52, 49-53, 47-58 were then amplified by PCR. For exon 46, the sense primer targeted the end of intron 45 (called Sense 46 5'- CCTCCCTAAGCGCTAGGGTTACAGG) (SEQ ID NO: 179) and the antisense primer targeted the start of intron 46 (called Antisense 46 5'- ACTCCCCATATCCCGTTGTC) (SEQ ID NO: 180). For exon 47, the forward and reverse primers targeted respectively the end of the intron 46 (called Sense 47 5'-GTATTTGAGGTACCACTGGGCCCTC) (SEQ ID NO: 181) and the start of intron 47 (called Antisense 47 5'-GCCACTGAGCTGGACACACGAAATG) (SEQ ID NO: 182). For exon 49, the forward and reverse primers targeted respectively the end of the intron 48 (called Sense 49 5'-GTCATGCTTCAGCCTTCTCCAGAC) (SEQ ID NO: 183) and the start of intron 49 (called Antisense 49 5'-GTTTATCCCAGGCCAGCTTTTTGC) (SEQ ID NO: 184). For exon 51, the forward and reverse primers targeted respectively the end of the intron 50 (called Sense 51 5'-GGCTTTGATTTCCCTAGGGTCCAGC) (SEQ ID NO: 185) and the start of intron 51 (called Antisense 51 5'-GGAGAAGGCAAATTGGCACAGACAA) (SEQ ID NO: 186). For exon 52, the forward and reverse primers targeted respectively the end of the intron 51 (called Sense 52 5'- GTAATCCGAGGTACTCCGGAATGTC) (SEQ ID NO: 187) and the start of intron 52 (called Antisense 52 5'- GTTTCCCCTACTCCTTCGTCTGTC) (SEQ ID NO: 188). For exon 53, the forward and reverse primers targeted respectively the end of the intron 52 (called Sense 53 5'-CACTGGGAAATCAGGCTGATGGGTG) (SEQ ID NO: 189 and the start of intron 53 (called Antisense 53 5'-GCCAAGGAAGGAGAATTGCTTGAGG) (SEQ ID NO: 190). For exon 58, the forward and reverse primers targeted respectively the end of the intron 57 (called Sense 58 5'- GGCTCACGGTATACCTCACGATCC) (SEQ ID NO: 191) and the start of intron 58 (called Antisense 58 5'- CCTCCTCACAGATAACTCCCTTTG) (SEQ ID NO: 192) For the hybrid exons 46-51, the forward Sense 46 and reverse Antisense 51 were used. For the hybrid exons 46-53, the forward Sense 46' (5-'CACTGCGCCTGGCCAGGAATTTTTGC) (SEQ ID NO: 193) and reverse Antisense 51 were used. For the hybrid exon 47-52, the forward Sense 47 and reverse Antisense 52 were used. For the hybrid exon 49-52, the forward Sense 49 and reverse Antisense 52 were used. For the hybrid exon 49-53, the forward Sense 49 and reverse Antisense 53 were used. From 293T cells, for the hybrid exons 47-58 the primer forward Sense 47(SEQ ID NO: 181) and the primer reverse Antisense 58 (SEQ ID NO: 192) were used. From myoblasts cells, for the hybrid exons 47-58 the forward Sense 47' (5'-CAATAGAAGCAAAGACAAGGTAGTTG) (SEQ ID NO: 194) and the reverse Antisense 58' (5'- GCACAAACTGATTTATGCATGGTAG) (SEQ ID NO: 195) were used. From genomic DNA of mice injected with AAVs, for an optimal detection of the formation of the hybrid exons 47 – 58, we performed a nested-PCR. The first PCR was done using the primer forward Sense 47 (SEQ ID NO: 181) and the primer reverse Antisense 58 (SEQ ID NO: 192). The second PCR was done using the primer forward Sense 47' (SEQ ID NO: 194)

and the primer reverse Antisense 58' (SEQ ID NO: 195). All PCR amplifications were performed in a thermal cycler C1000 Touch of BIO RAD (Hercules, CA) with the Phusion™ high fidelity polymerase (Thermo scientific, EU, Lithuania). Exon 46 was amplified using the following program: 98°C / 10 sec, 64.5°C / 30 sec, 72°C/ 40 sec for 35 cycles. Exons 47, 49, 51 and 53 were amplified using the following program: 98°C / 10 sec, 61.2°C / 30 sec, 72°C / 45 sec for 35 cycles. Exons 52 and 58 were amplified using the following program: 98°C / 10 sec, 63°C / 30 sec, 72°C / 40 sec for 35 cycles. The hybrid exons 46-51 were amplified using the following program: 98°C / 10 sec, 66°C / 30 sec, 72°C/ 30 sec for 35 cycles. The hybrid exons 46-53 were amplified using the following program: 98°C / 10 sec, 65.5°C / 30 sec, 72°C/ 40 sec for 35 cycles. The hybrid exon 47-52 was amplified using the following program: 98°C / 10 sec, 61.2°C / 30 sec, 72°C/ 30 sec for 35 cycles. The hybrid exon 49-52 was amplified using the following program: 98°C / 10 sec, 66°C / 30 sec, 72°C/ 45 sec for 35 cycles. The hybrid exon 49-53 was amplified using the following program: 98°C / 10 sec, 63°C / 30 sec, 72°C/ 45 sec for 35 cycles. From 293T cells, the hybrid exons 47-58 were amplified using the following program: 98°C / 10 sec, 61.2°C / 30 sec, 72°C/ 30 sec for 35 cycles. From myoblasts cells, the hybrid exons 47-58 were amplified using the following program: 98°C / 10 sec, 63°C / 30 sec, 72°C/ 30 sec for 35 cycles. The amplicons of individual exons 46, 47, 49, 51, 52, 53 and 58 were used to perform the Surveyor assay. There was first a hybridization step of the amplicons using a slow-hybridization program (denaturation at 95°C for 5 min followed by gradual cooling of the amplicons) with BIO RAD thermal cycler C1000Touch™ (Hercules, CA). Subsequently, the amplicons were digested with nuclease Cel (Integrated DNA Technologies, Coralville, IA) in the thermal cycler at 42°C for 1 hour. The digestion products were visualized on agarose gel 2%.

[0236] Cloning and sequencing of the hybrid exons. The amplicons of hybrid exons obtained by the amplification of genomic DNA extracted from 293T cells or myoblasts transfected with 2 different pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA plasmid was purified using the GeneJET™ PCR Purification Kit (Thermo Scientific, EU, Lithuania). The purified PCR products were cloned into the linearized cloning vector pMiniT™ (NEB, Ipwisch, MA). Then, plasmid DNA was extracted with the Miniprep Kit (Thermo Scientific, EU, Lithuania). The clones were sent for sequencing using primers provided by the manufacturer (NEB, Ipwisch, MA). Sequencing results were analyzed with the CLC Sequence Viewer software (CLCBio).

[0237] sgRNA *in vitro* Transcription. SgRNAs 1-50 and 5-54 were amplified by PCR from plasmids presented in Iyombe et al. (2016) using a C1000 Touch thermocycler (Bio Rad Inc., Hercules, CA, USA). The forward primer permitted to add in 5' the T7 promoter sequence (primers Fw-IVT-1-50 and Fw-IVT-5-54, Table 1). Reverse primer corresponds to the end of the tracrRNA (Rv-IVT-tracr, Table 5). Amplicon sequences coding for sgRNAs were purified on column (PCR Purification Kit GeneJET, Thermo Scientific Inc., Waltham, MA, USA) and dosed with a NanoDrop™ (Thermo Scientific Inc.). Transcription was made with the kit HiScribe Quick™ T7 High Yield RNA Synthesis Kit (NEB Inc. Ipswich, MA, USA) using 500 ng of amplicons with an incubation of 16 h at 37°C. Thereafter, a treatment with DNase I was made followed by an extraction with phenol/chloroform and a precipitation with cold ethanol. After centrifugation, pellets were resuspended in 0.1 mM EDTA solution and the RNA concentration was estimated using a NanoDrop™ (Thermo Scientific Inc.). The sgRNAs obtained were stored at -80°C.

[0238] **Table 5**: Primers used for the amplification of *in vitro* transcription templates.

Primer name	Sequence
Fw-IVT-1-50 (SEQ ID NO: 201)	5'-TAATACGACTCACTATAAGAAGATCTGAGCTCTGAGGTTTT-3'
Fw-IVT-5-54 (SEQ ID NO: 202)	5'- TAATACGACTCACTATAAGAGAATATCAATGCCTCTGTTTTAG -3'
Rv-IVT-tracr (SEQ ID NO: 203)	5'-AAAAAAGCACCGACTCGGTGCCA-3'

[0239] *In vitro* Activity Analysis of crRNA:tracrRNA:SpCas9 Ribonucleic Complex. In order to analyze the activity of the crRNA:tracrRNA:SpCas9 complex, exon 54 of the dystrophin gene was amplified by PCR using a C1000 Touch thermocycler. CrRNA and tracrRNA were obtained commercially from Dharmacon Inc. and Cas9 protein was obtained from Feldan Therapeutics Inc. The crRNA:tracrRNA: SpCas9 was then mixed with the amplicon of exon 54 and incubated at 37°C for 30 min. The reaction was stopped with RNase (1 mg/mL) and heating at 56°C for 5 min. Cleavage analysis was done by migration on a 1.5% agarose gel in 1X TBE buffer.

[0240] *In vitro* Protein Delivery. For experiments, HeLa cells were plated in 96 well plates at a confluence of 10000 cells per well and incubated overnight in DMEM medium (Dulbecco's Modified Eagle Medium, Invitrogen Inc. Carlsbad, CA, USA) supplemented with 10% fetal bovine serum and 5% penicillin and streptomycin. The cells were incubated at 37°C under an atmosphere of 5% CO2. The analysis of the specificity and efficiency of sgRNA was then done with these cells using Feldan Shuttle technology (Feldan Therapeutics Inc.). Ribonucleic complexes were made using 2.5 µM of SpCas9 protein and 2 µM of each sgRNA separately and incubated for 5 min at room temperature (RT). Complexes were incubated with Feldan Shuttle A at a concentration of 50 µM and volume was completed using PBS 1X. Mixture was added directly to the cells for exactly 1 min incubation at RT. Cells were then washed with medium once followed by a washing step with PBS 1X. Fresh medium was added and cells were incubated for 48 h before genomic DNA extraction.

[0241] *In vivo* Protein Delivery

[0242] Electroporation. The Tibialis anterior (TA) muscles of hDMD mice were injected longitudinally with 40 µL of a solution containing the protein SpCas9 complexed with 2 sgRNAs and a protein delivery agent. Complexes formed with 2.5 µM of SpCas9 protein and 2 µM of sgRNA (total amount of both sgRNA used) were incubated for 5 min and electroporated (electroporator Electro Square Porator, Model ECM630, BTX Harvard Apparatus Inc. Holliston, MA, USA). An electrolyte cream (Teca Inc. Louisville, Quebec, Canada) was applied to the mouse leg to promote the passage of electrical pulses. Eight pulses of 20 ms separated by 1 s were applied to the mouse muscle. The voltage was adjusted at 100 V/cm depending on the width of the treated mouse leg. The mice were sacrificed 72 hours later. The mice were then anesthetised with isoflurane and euthanized by CO2 inhalation. The muscles were collected and separated into 4 pieces lengthwise.

[0243] Lipofectamine™ RNAimax. 1 µM of SpCas9 protein and 0.7 µM of sgRNA (total amount of both sgRNA used) was delivered using 2 µL of Lipofectamine RNAimax (Thermo Fisher Scientific Inc.). The ribonucleic complex

was initially incubated for 5 min and then mixed with Lipofectamine RNAimax, followed by another incubation of 20 min before the longitudinally injection in TA muscles of hDMD mice with a final volume of 40 µL. The mice were sacrificed 72 hours later. The muscles were collected with the same method as for the electroporation.

[0244] Feldan Shuttles. 2.5 µM of SpCas9 protein were complexed with 2 µM of sgRNAs (total amount of both sgRNA used) and then incubated for 5 min at RT. Resulting complexes were mixed with 2 different Feldan Shuttles, Feldan Shuttle A (FSA, SEQ ID NO: 196) at 50 µM or Feldan Shuttle B (FSB, SEQ ID NO: 197) at 35 µM. In both cases, the final volume of the reaction was maintained at 40 µL. This solution was injected longitudinally in the TA muscles. The mice were sacrificed 72 hours later. The muscles were collected with the same method as for the electroporation.

[0245] Genomic DNA Extraction from Muscles. Treated TA muscles were collected and separated into 4 pieces lengthwise. The fractions were subsequently treated with a lysing buffer containing proteinase K (2 mg/mL). This was followed by incubation for 16 hours at 55°C. Lysis was followed by extraction with phenol/chloroform to obtain purified genomic DNA. DNA concentration was estimated using a NanoDrop™ (Thermo Scientific Inc.).

[0246] Genomic DNA extraction from tissue cuts. First, OCT (optimum cutting temperature) compound was removed by washing once the tissue cuts with PBS 1X. Then, 400 µL of lysis buffer was applied onto the sample. The mixture was placed into a 1.5 mL eppendorf tube and supplemented with 10 µL of Proteinase K (20mg/mL) followed by incubation at 56°C for 1h. Suspension was brought to 500 µL using distilled water. Then one volume of phenol-chloroform was added and following centrifugation, the aqueous phase was recovered in a new microfuge tube. Then DNA was precipitated using 1/10 volume of NaCl 5 M and two volumes of 100% ethanol followed by 5 min centrifugation ate 13200rpm. The pellet was washed with 70% ethanol, centrifuged and the DNA was resuspended in double-distilled water. The genomic DNA concentration was assayed with a Nanodrop™ spectrophotometer (Thermo Scientific, Logan, UT).

[0247] Immunohistochemistry for the identification of the dystrophin expression in muscle fibers. First, to withdraw the OCT compound we washed the slides three times for 5 min using PBS 1X to remove traces of OCT compound. Then muscle sections were blocked for 1h using PBS1X and 10% of FBS. Immediately, blocking solution was replaced by a solution containing the primary antibody mouse anti-dystrophin (NCL-Dys2, Novocastra) diluted 1:50 into PBS 1X and 10 % FBS. The primary antibody was incubated for 1h. Then slides were wash three times for 10 min using PBS1X. The secondary antibody goat anti-mouse Alexa Fluor 546. To prevent the sample from drying, slides were covered with 50% glycerol in PBS then covered with a cover slide. Fluorescence were observed using. Pictures were taken using a Nikon ISO 3200, with an exposure time of for ¼ sec.

[0248] PCR analysis of hybrid exons formed in the hDMD/mdx mouse model. Genomic DNA, extracted from hDMD muscles, was used to amplify the hybrid exon using a C1000 Touch thermocycler and primers Fw-Int50 (5'-TGCCTGGAGAAAGGGTTTTTGT-3', SEQ ID NO: 222) and Rv-Ex54 (5'-TATCAGATAACAGGTAAGGCAGTG-3', SEQ ID NO: 177). Analysis was then made by migration on 1.5 % agarose gel in 1X TBE buffer containing Red Safe 1X dye (Chembio Inc.).

[0249] Cloning and Sequencing for the Hybrid Exon. The hybrid exon amplicons were purified using a gel extraction kit (Gel Extraction Kit GeneJET, Thermo Scientific Inc.). PCR purified fragments obtained were then dosed using a NanoDrop™ (Thermo Scientific Inc.). 15 ng of the PCR product was cloned into the plasmid vector pMiniT of the PCR Cloning Kit (NEB Inc. Ipswich, MA, USA). Ligations were then transformed into competent bacteria provided by the manufacturer, and then seeded into LB Agar plates containing ampicillin (50 µg/mL). After 16 h incubation at 37°C, clones were taken and inoculated into 5 mL of LB medium (lysogeny broth) with ampicillin (50 µg/mL). The samples were incubated 16 h at 37°C. Bacterial cultures were used for plasmid DNA extraction using a Miniprep kit (GeneJET Plasmid Miniprep Kit, Thermo Scientific Inc.). The plasmids were dosed with a NanoDrop™ (Thermo Scientific Inc.) and then sent to the sequencing platform of the CHUL Research Centre (CHUQ). The sequences obtained were analyzed using the BLAST homology platform (NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi).

EXAMPLE 9

ANALYSIS OF *IN VITRO* CLEAVAGE BY THE SpCAS9 / RNA COMPLEX

[0250] SgRNAs 1-50 and 5-50, (see Table 3 above), were produced by *in vitro* transcription. Commercially produced crRNA and tracrRNA were also used in some experiments. The sgRNAs and crRNA targeted exon 50 preceding the deletion and exon 54 that follows the deletion causing the disease. The gRNA-targeted sequences were chosen to not only correct the reading frame but also the dystrophin protein structure. The cutting efficacy of the gRNAs and of the crRNA in complex with the Cas9 protein was initially tested on a PCR amplicon of the targeted region in exon 54. Results showed high cutting efficiency *in vitro*, since complete digestions were obtained in only 30 minutes of incubation at 37°C (FIG. 8A).

EXAMPLE 10

SURVEYOR TEST ON EXON 54 OF THE DMD GENE, FOLLOWING TRANSDUCTION OF THE CRISPR / CAS9 SYSTEM WITH A FELDAN SHUTTLE IN HELA CELLS.

[0251] The SpCas9:crRNA:tracrRNA or the SpCas9:sgRNA complexes were then transduced in Hela cells using the Feldan Shuttle technology (FIG. 8B). Feldan Shuttle A (FSA) and Feldan Shuttle B (FSB) are polypeptide-based shuttle agents, which were kindly provided by Feldan Therapeutics Inc. Our experiments showed efficient cutting of the dystrophin gene using the Surveyor assay on the genomic extracted DNA only when the SpCas9 protein was delivered with the Feldan Shuttle in complex with the sgRNAs or with the crRNA:tracrRNA in cells (FIG. 8B). There was no difference in the cutting efficiency obtained using the sgRNA or the crRNA:tracrRNA. This shows that the use of sgRNA transcribed *in vitro* does not interfere with the activity of the Cas9 protein.

EXAMPLE 11

DELIVERY OF THE SpCAS9: SGRNA COMPLEX IN THE hDMD/mdx MOUSE MODEL

[0252] Following positive *in vitro* results, this gene editing approach was tested *in vivo* in the hDMD/mdx mouse model. This mouse model contains the complete human DMD gene integrated in its genome ('t Hoen et al. , 2008).

This allows us to verify whether the sgRNAs targeting the human DMD exons 50 and 54 were also able to induce specific deletions *in vivo* following delivery with different methods, i.e., a single injection of the SpCas9 protein in complex with the 2 sgRNAs was made into the *Tibialis anterior* (TA) of the mouse. This complex was transduced in the muscle fibers of the TA either by electroporation, lipofection (Lipofectamine™ RNAimax) or with two different Feldan Shuttles. The mice were sacrificed 72 hours later and the treated muscle was extracted and divided in 4 equivalent parts to analyze the distribution of the gene editing. Genomic DNA was extracted for PCR amplification and sequencing. The region between exons 50 and 54 is about 160 kb, and thus cannot be amplified by conventional PCR. However, if the hybrid exon is formed following DSBs in both targeted exons, a PCR fragment of about 600 bp can be observed (FIG. 8C). The amplicons obtained following transduction of the sgRNA pair in complex with the SpCas9 protein were gel purified and cloned into the pMiniT plasmid, transformed in bacteria and clones were screened for successful insertions. Positive clones, according to the digestion pattern (data not shown), were sent for sequencing to demonstrate the presence of a hybrid exon formed by the fusion of a part of exon 50 with a portion of exon 54 (FIG. 9A(i)). The expected hybrid exon 50-54 was present (FIG. 9A(ii), SEQ ID NO: 200), as indicated by the perfect alignment between the observed sequence and the expected sequence (FIG. 9b). This result was obtained for about 60% of the clone sequences. The other 40% of the sequences contained INDELs at the junction site. These INDELs are made in the correction of the gene by NHEJ, which is also observed by our group when the same hybrid exon was formed using plasmids encoding the SpCas9 gene and the same sgRNAs used in our experiment.

EXAMPLE 12

STRATEGY FOR THE IDENFICATION OF gRNAs OF INTEREST TARGETING THE HUMAN DYSTROPHIN GENE

[0253] Our primary goal was to establish a strategy based on the creation of a hybrid exon allowing the correction of the dystrophin gene reading frame, in the case of DMD patients affected by the deletion of exon 50. Thus, we screened all possible gRNAs target sites surrounding exon 50. We identified all sites from exon 46 to exon 58. We identified nearly 50 gRNAs that can be used with the Cas9 protein from *Staphylococcus aureus* (*S. aureus*). As SaCas9 nuclease induces a double strand break (DSB) precisely 3 nucleotides upstream of the PAM (NNGRRT), we were able to select gRNAs that can be combined to create a new hybrid exon that restores a normal reading frame in the dystrophin gene. This hybrid junction could permit the production of an internally-truncated dystrophin. We focused on combinations of gRNAs where the hybrid junction maintains the configuration of a normal spectrin-like repeat, where hydrophobic amino acids are localized in position “a” and “d” of the heptad motif (spectrin-like repeats are composed of α-helixes each containing 7 amino-acids (a to f) where hydrophobic amino acid are in the location “a” and “d”; see for example FIG. 16). In order to assess the correct localization of those amino acids, we referred to the eDystrophin database which provides information about the structural domains of the dystrophin protein (http://edystrophin.genouest.org/). As a result of our preliminary analysis, we focused our work on 18 gRNAs (see Table 6) that generate 12 hybrid exons not only correcting the reading frame of the dystrophin gene but also maintaining the structure of spectrin-like repeats.

EXAMPLE 13

TESTING OF INDIVIDUAL gRNAS FOR THE SaCAS9/CRISPR SYSTEM

[0254] We designed 18 gRNAs that we cloned into the PX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA plasmid (Addgene # 61591) (Table 6). In order to assess the activity of these gRNAs, we transfected 293T cells. 48 hours post transfection we extracted genomic DNA. Targeted exons 46, 47, 49, 51, 52, 53, 58 were amplified by PCR then submitted to a "Surveyor" enzyme assay for the detection of INDELs (FIG. 12). Among the 18 gRNAs, gRNA 2, gRNA 4, gRNA 9, gRNA 11, gRNA 15 show few or no activity whereas all remaining gRNAs exhibited good cleavage efficiency and specificity as they generated cleave bands at the expected size.

EXAMPLE 14

TESTING OF gRNA PAIRS USING THE SaCas9/CRISPR SYSTEM IN 293T CELLS

[0255] 293T cells were co-transfected with combinations (see Table 7) of gRNAs that might allow to create large genomic deletion and precisely connect exons, surrounding a deletion of the exon 50, to the right nucleotides. Thus, we tested the 12 combinations of gRNAs we identified. Hybrid exons were identified by PCR using the forward primer for one of the targeted exon and a reverse primer for the other, previously used when gRNAs were individually tested (FIG. 13). Thus, PCR amplification is only possible if a deletion occurred between the two targeted exons as there are too spaced by the introns.

EXAMPLE 15

CHARACTERIZATION OF HYBRID EXONS IN 293T CELLS (SaCas9/CRISPR SYSTEM)

[0256] PCR products from amplification of hybrid exons were purified then cloned into a pMiniT vector for the sequencing of the hybrid exons resulting from large genomic deletion by the SaCas9 and 2 gRNAs surrounding the exon 50 deletion (FIG. 18a). Finally, combinations of gRNAs 3 and 16 and gRNAs 5 and 18 seemed the most promising as they allow to create the largest genomic deletion thus permitting to cover up to 40% of mutations identified in DMD patients. In addition, the combination of gRNAs 3 and 16 creating a 450kbp deletion is the most precise combination as the junction is the expected one in 75% of the tested clones.

EXAMPLE 16

DYSTROPHIN EXON TARGETING IN DMD MYOBLASTS USING THE SaCas9/CRISPR SYSTEM

[0257] Myoblast cells from DMD patients were co-transfected with two combinations of gRNAs, combination of gRNAs 3 and 16 and the combination of gRNAs 5 and 18, that might allow to create large genomic deletion and precisely connect exons 47 and 58. These combinations should be able to create a hybrid exon with a correct reading frame leading to the production of a dystrophin protein with a hybrid spectrin like repeat. Hybrid exons 47 – 58 were identified by PCR using the primers forward Sense 47' (SEQ ID NO: 194) and the reverse Antisense 58' (SEQ ID NO: 195) (FIG. 15). Thus, PCR amplification is only possible if a deletion occurred between the two targeted exons as there are too spaced by the introns.

EXAMPLE 17

CHARACTERIZATION OF THE HYBRID EXON 47-58 IN MYOBLASTS (SaCas9 CRISPR SYSTEM)

[0258] PCR products from amplification of the hybrid exons 47-58 were purified then cloned into a pMiniT vector for the sequencing of the hybrid exons resulting from large deletion by SaCas9 and 2 gRNAs surrounding exon 50 deletion (FIG. 18b).

EXAMPLE 18

IN VIVO CORRECTION IN THE HDMD/*MDX* MOUSE (SaCas9 CRISPR SYSTEM)

[0259] Following the identification and characterization of pairs of gRNAs in 293T then in myoblasts from different DMD patients, we tested the viral AAV-mediated delivery of the CRISPR system from *S. aureus* and the pair of gRNAs 3 and 16 and the pair of gRNAs 5 and 18 using the AAV serotype 9. One AAV9 permitted the expression of the SaCas9 nuclease under the control of the CMV promotor while a second AAV9 coded for a pair of gRNAs (gRNAs 3 and 16 or gRNAs 5 and 18). We compared the administration of the viral particles through intraperitoneal and intravenous injection path into 6 week old hDMD/mdx mice. Mice were sacrificed 6 weeks after the injection. The tissues – heart, diaphragm, *Tibilalis anterior* (TA), brain/*cerebellum,* liver- were collected and incubated overnight in 30% sucrose. After harvesting in OCT medium and liquid nitrogen freezing followed by processing of the tissue onto glass slides, genomic DNA was extracted and purified. For an optimal detection of the successful formation of the hybrid exon 47-58 *in vivo* we performed a nested-PCR. The first PCR was done using the forward primer sense 47 and the reverse primer antisense 58. The second PCR the forward Sense 47' (SEQ ID NO: 194) and the reverse Antisense 58' (SEQ ID NO: 195). FIG. 21 shows the detection of the formation of the hybrid exon 47-58 following nested-PCR with genomic DNA from the heart (a), the diaphragm (b), the TA (c). We observed that in the heart, as in the diaphragm, the viral delivery of the SaCas9 and of a pair of gRNAs, gRNAs 3 and 16 as well as gRNAs 5 and 18, permit the formation of the hybrid exon 47-58, in comparison to the non-treated mouse (NT). Besides, in the TA, the amplification of the hybrid exon 47-58 was detected when the SaCas9 and the pair of gRNAs 3 and 16 or the pair of gRNAs 5 and 18 were only delivered through intravenous injection.

EXAMPLE 19

METHODS FOR DELIVERY OF THE SpCAS9: sgRNA COMPLEX IN THE RAG/MDX DYSTROPHIC MOUSE MODEL (Examples 20-23)

[0260] gRNAs cloning into the plasmid pSpCas9(BB)-2A-Puro. The oligonucleotides coding for the target sequences and their complementary sequences were ordered from IDT and cloned into the plasmid pSpCas9(BB)-2A-Puro (pX459) (Addgene plasmid # 48139) containing two BbsI restriction sites necessary for insertion of a protospacer under the control of the U6 promoter. Cloning was performed as previously described in Example 8. The pSpCas(BB)-2A-Puro plasmid contains the Cas9 of *S. pyogenes,* and a puromcyin resistance gene under the control of the CBh promoter; both genes are separated by a sequence encoding the self-cleavable peptide T2A. The puromycin resistance gene allows to select and enriched the transfected C2C12 cells (ATCC® CRL1772), a hard to

transfect cell line.

[0261] The gRNA activities were tested individually or in pairs by transfection of the pSpCas9(BB)-2A-puromycin plasmid encoding each gRNA in C2C12 cells (murine myoblast). The C2C12 cells were grown in Dulbecco's modified Eagle medium (DMEM) medium (Invitrogen, Grand Island, NY) containing 10% fetal bovine serum (FBS) and antibiotics (penicillin 100 U/ml / streptomycin 100 µg/ml). 24 hours after the transfection, cells were submitted to 2 µg/mL of puromycin for 48 hours to select transfected C2C12 cells. Following puromycin selection, cells were grown for 48 hours in the absence of puromycin.

[0262] Genomic DNA was extracted as described in Example 8.

[0263] PCR amplification of modified DYS gene. To assess the correct cuts and deletions the targeted region was amplified by PCR using the forward primer Fw2-i22 (5'- CTGTGATGTGAGGACATATAAAGAC 3') (SEQ ID NO: 204) located in intron 22 and the reverse primer Rev1-i23 (5'- TCAATGTAGGGAAGGAAATATGGCA 3') (SEQ ID NO: 205) located in intron 23. PCR amplifications was performed in a thermal cycler C1000 Touch of BIO RAD (Hercules, CA) with the Phusion™ high fidelity polymerase (Thermo scientific, EU, Lithuania). Exon 23 was amplified using the following program: 98°C / 10 sec, 61.2°C / 30 sec, 72°C/ 45 sec for 35 cycles. For the analysis of individual gRNA activities, PCR products were submitted to Surveyor enzyme assay. For the analysis of the pairs of gRNAs, PCR products were loaded onto an agarose gel. If a pair of gRNAs permits the deletion of exon 23, PCR amplification will allow the detection of a wild type band along with a shorten band corresponding to remaining parts of intron 22 and intron 23 following the deletion.

[0264] Amplicons generated for exon 23 were used to perform the Surveyor assay. There was first a hybridization step of the amplicons using a slow-hybridization program (denaturation at 95°C for 5 min followed by gradual cooling of the amplicons) with BIO RAD thermal cycler C1000Touch™ (Hercules, CA). Subsequently, the amplicons were digested with nuclease Cel (Integrated DNA Technologies, Coralville, IA) in the thermal cycler at 42°C for 1 hour. The digestion products were visualized on agarose gel 2%.

[0265] For *in vivo* experiments, we used the sgRNA 20, sgRNA 22 and sgRNA 23 that were ordered from Synthego (Redwood City, CA). We received 1 nmole of each sgRNA that were resuspended into nuclease and RNase free water to a stock concentration of 50µM, harvested at -80°C. As a working solution, sgRNAs were diluted to 10 µM and stored at -20°C.

[0266] *In vivo* protein delivery. 2.5 µM of SpCas9 protein were complexed with 2 µM of sgRNAs and then incubated for 5 to 10 min at RT. The sgRNAs mixture is composed of 1 µM of one sgRNA targeting a region upstream of exon 23 and 1 µM of a sgRNA ttargeting a region downstream of exon 23. Resulting complexes were mixed with 3 different Feldan Shuttles, Feldan Shuttle A (FSA) at 50 µM or Feldan Shuttle B (FSB) at 35 µM or Feldan Shuttle C (FSC) at 50 µM. In each condition, the final volume of the reaction was maintained at 40 µL. For the negative control, SpCas9 complexed with sgRNAs were injected without Feldan Shuttle. The tested solutions were injected longitudinally in the *Tibialis anterior* muscles. The mice were sacrificed 3 weeks later and the muscles were collected. The muscles were incubated overnight at 4°C in 30% of sucrose. Then muscles were embedded into

OCT medium and flash frozen in liquid nitrogen. The tissues were then processed using a cryostat (Leica system) allowing to make 16 µm transversal cross sections onto glycined glass slides.

[0267] Strategy for the PCR amplification of hybrid introns. For the detection of the deletion of exon 23 we performed an optimized nested-PCR approach. In the first PCR reaction, we used the external forward primer F1-i22 (5'- GAACATGTCTTATCAGTCAAGAGATC) (SEQ ID NO: 223) and the external reverse primer Rev1-i23 (5'- TCAATGTAGGGAAGGAAATATGGCA) (SEQ ID NO: 224) along with a forward "poison" primer Fw-P (5'- AACTATCTGAGTGACACTGTGAAGG) (SEQ ID NO: 225) targeting exon 23. This strategy allows to mainly amplify the wild-type region generated from the forward "poison" primer and the external reverse primer. However, if exon 23 is deleted, the forward poison primer Fw-P is not able to generate an amplification product. Thus, from genomic DNA, the two external primers will amplify and enriched genomic DNA where exon 23 is deleted, resulting in amplification of a hybrid intron formed by the remaining part of intron 22 and a part of intron 23. Following the first PCR, nested-PCR is performed using the internal forward primer F2-i22 (5'- CTGTGATGTGAGGACATATAAAGAC) (SEQ ID NO: 226) and the internal reverse primer Rev2-i23 (5'- CAGACAATCCAAGAAGGTATGACAC) (SEQ ID NO: 227). These two internal primers will amplify from the previous amplicons generated with the external primers only; i.e., amplicons lacking exon 23. This combination of internal primers only permits amplification from the amplicons obtained with the two external primers F1-i22 and Rev1-i23.

[0268] Hybrid intron analysis. The amplicons of hybrid introns obtained by the amplification of genomic DNA extracted from TA muscle were cloned into the linearized cloning vector pMiniT™ (NEB, Ipwisch, MA). Then, plasmid DNA was extracted with the Miniprep™ Kit (Thermo Scientific, EU, Lithuania). The clones were sent for sequencing using primers provided by the manufacturer (NEB, Ipwisch, MA). Sequencing results were analyzed with the CLC Sequence Viewer software (CLCBio).

[0269] Dystrophin positive fibers count. We counted the number of dystrophin positive fibers in three successive muscle cuts exhibiting the highest number of positive and established means and standard deviation.

EXAMPLE 20

gRNAs targeting intron 22 and intron 23 to remove mutated exon 23 in the mouse model Rag/mdx

[0270] In the mouse model Rag/mdx the dystrophin gene is mutated in exon 23. This nonsense point mutation C572T, thus creates a premature STOP codon and abrogates the synthesis of the 427 kDa wild type dystrophin protein. As the number of nucleotides in exon 23 is a multiple of three, its deletion should allow correcting the dystrophin gene reading frame. Consequently, we aimed at deleting exon 23 by designing a pair of gRNAs that targets its surroundings introns. Thus, we identified 4 gRNAs target sites upstream of exon 23, in intron 22 and 4 gRNAs target sites downstream of exon 23, in intron 23 (see Table 8). Two target sites span the exon 23/intron 23 junction but the cut which is introduced into the gene in located within the intronic sequence (see gRNA # 20 and 26 in Table 8).

[0271] We thus prepared 8 gRNAs expression construct targeting intron 22 or intron 23 that we cloned into PX459. In order to assess the activity of these 8 gRNAs, C2C12 cells were transfected. 24 hours post transfection we

selected transfected cells using 1 µg/mL of puromycin for 48 hours. Following another 48 hours of culture in selection free medium, we extracted genomic DNA. We amplified by PCR exon 23 and its surrounding introns and submitted the samples to a "Surveyor" detection assay to assess the creation of INDELs (FIG. 22a). Among the 8 gRNAs we tested, gRNA 19, gRNA 20, gRNA 21, gRNA 22, gRNA 23 and gRNA 24 demonstrated good cleavage efficiency whereas no activity was detected for gRNA 25 and gRNA 26 under the conditions tested.

EXAMPLE 21

Testing of gRNAs pairs for the deletion of exon 23 in C2C12 cells

[0272] Based on the gRNAs identified in Example 20, we tested combinations of gRNAs that could delete exon 23. Thus, we tested the combination of gRNAs 22 and 20, gRNAs 22 and 21, gRNAs 22 and 24, gRNAs 23 and 20, gRNAs 23 and 21 and gRNAs 23 and 24. To assay the deletion of exon 23, PCR products were amplified from purified genomic DNA using the primers Fw2-i22 and Rev1-i23 and loaded onto an agarose gel (FIG. 22b). In each condition, PCR reaction allowed the detection of the wild type band formed by the complete intron 22, exon 23 and intron 23. Besides for each gRNAs combination that were tested, the amplification permits to detect a second PCR product which size is smaller than the wild type product. These truncated products correspond to the sizes resulting from the deletion of the exon 23. For further *in vivo* experiments, we choose to focus on the combination of gRNAs 20 and 22 and the combination of gRNA 20 and 23.

EXAMPLE 22

Testing gRNAs pairs for the deletion of mutated exon 23 in the mouse model Rag/mdx

[0273] For the in vivo experiments, we used a non-viral delivery approach based on a peptide-mediated delivery of the Cas9 protein along with a pair of gRNAs that target DNA sequences upstream and downstream of mutated exon 23 in the dystrophic mouse model Rag/mdx. This non-viral delivery tool, known as the Feldan Shuttle (FS), permits to bind the ribonucleoprotein complex Cas9/gRNA through non-covalent binding. Upon cell entry, pH variation leads to the disruption of the interaction between the FS and the ribonucleoprotein complex. Three shuttles from Feldan Therapeutics were tested: FSA, FSB and FSC.

[0274] For the first *in vivo* experiment, 4 month old mice were injected in their *Tibialis anterior* through a single trans longitudinal injection from the lower to the upper part of the muscle. 3 weeks after the injection, mice were sacrificed for the collection of injected muscles. Even if previous results demonstrated that 3 days post injection we were able to detect the desired genomic deletion, we waited for 3 weeks in order to allow accumulation of the restored dystrophin expression into myotubes. Following muscle collection, harvesting and processing into 16 µm thick transversal cuts, we performed an immunohistochemistry against dystrophin using the NCL-Dys2 antibody (FIG. 23) and analysed dystrophin positive fibers.

[0275] We observed that non-treated muscles contain nearly 37 dystrophin positive fibers. These fibers, that naturally occur, are named revertant fibers and might results from another mutation into exon 23 or from a splicing of exon 23, thus correcting the reading frame of the dystrophin gene. In comparison to non-treated muscles, muscles

treated with Cas9 and gRNAs 22 and 20 complexed with FSB, or the delivery of Cas9 with gRNAs 23 and 20 complexed with FSA generated an increase in the number of dystrophin positive fibers, respectively showing 49 and 39 dystrophin positive fibers. However, the combination of gRNAs 22 and 20 with FSA and the combination of gRNAs 23 and 20 with the FSB allowed to obtain 83 and 99 dystrophin positive fibers respectively. Finally, the FSC complexed with the ribonucleoprotein delivering the gRNAs 22 and 20 or the gRNAs 23 and 20, exhibited the highest number of dystrophin positive fibers with 186 and 199 fibers, respectively. Consequently, in further experiments we focused on the use of the FSC which seems to be the most suitable non-viral delivery system for the Cas9/gRNAs delivery in the muscle and thus for the correction of the reading frame of the dystrophin gene.

[0276] To confirm that the dystrophin positive fibers observed come from the effect of the CRISPR/Cas9 system programmed for the deletion of exon 23, we performed a PCR amplification of the targeted genomic region. As preliminary results of PCR amplification, using the primers Fw2-i22 and Rev1-i23, does not allow the detection of the deletion of the exon 23, we used an optimize nested-PCR approach that allow the detection of a small deletion, as we described in Example 19. Here, the first PCR was made using the primers Fw1-i22 and Rev1-i23 and the poison primer Fw-P (FIG. 25a) only permit the visualization of a product corresponding to a wild type sequence. Then, nested-PCR was performed using internal primers Fw2-i22 and Rev2-i23 (FIG. 25 b). The combination of gRNAs 23 and 20 with the SpCas9 complexed with the FSA does not allow the detection of the deletion of exon 23. However, the other combinations involving the combination of gRNAs with FSB or FSC permitted the detection of a PCR product corresponding to the deletion of exon 23.

[0277] For the second round of *in vivo* experiments, 6 weeks old mice were injected in their *Tibialis anterior* through a single trans longitudinal injection. 3 weeks after the injection, mice were sacrificed for the collection of injected muscles. Following muscle collection, harvesting and processing into 16 µm thick transversal cuts, we performed an immunohistochemistry against dystrophin using the NCL-Dys2 antibody (FIG. 26) and analysed dystrophin positive fibers. In these experiment, we injected SpCas9/gRNAs alone or complexed with the Feldan Shuttle FSC that demonstrated the most efficient dystrophin protein recovery in the first experiment.

[0278] In these second round experiments, we aimed to compare the delivery of the SpCas9 along with a pair of gRNAs, gRNAs 20 and 22 or gRNAs 20 and 23, complexed or not with FSC. We focused on shuttle FSC as it demonstrated the best delivery efficiency as shown by the elevated number of recovered dystrophin positive fibers detected using this shuttle. Intramuscular injections of SpCas9 complexed with the pair of gRNAs 20 and 22 and the pair of gRNAs 20 and 23 generated 101 and 120 dystrophin positive fibers, respectively. In addition, the use of the FSC for the delivery of the SpCas9 along with a pair of gRNAs permits to increase the number of dystrophin positive fibers. Indeed, the use of FSC for the delivery of the SpCas9 and gRNAs 20 and 22 generated 151 dystrophin positive fibers and the SpCas9 and the gRNAs 20 and 23 generated 166 dystrophin positive fibers.

[0279] To confirm that the dystrophin positive fibers we observed come from the effect of the CRISPR/Cas9 system and a specific pair of gRNAs for the deletion of exon 23, we performed PCR amplification of the targeted genomic region. The first PCR made using primers Fw1-i22, Rev1-i23 and poison primer Fw-P (FIG. 28a) only allowed amplification of a product corresponding to the wild type DYS sequence. Then, the nested-PCR was

performed using internal primers Fw2-i22 and Rev2-i23 (FIG. 28b). The combination of gRNAs 22 and 20 with the SpCas9 alone or complexed with the FSC allowed the detection of the deletion of exon 23. The same observation was made with the combination of gRNAs 23 and 20. However, one injection of the gRNAs 23 and 20 with the SpCas9 complexed with the FSC did not permit the detection of a PCR product corresponding to the deletion of the exon 23 which might result from a failed injection.

SUBSTITUTE SHEET (RULE 26)

Table 6: Exemplary gRNAs in exons 46-58. Nucleotides position are provided with reference to the DMD gene sequence ENS00000198947 (Chromosome X reverse strand 31,097,677-33,339,441) analysed using Benchling web tool (https://benchling.com/)

gRNA #	Exon cutting site #	Strand	gRNA target sequences* (excluding PAM)	SEQ ID NOs. Target/gRNA	gRNA target sequence position	PAM nts Position (cs: coding sequence, in: intron)	Cut sites in *DYS* gene**	Cut sites in amino acid sequence	gRNA involved in the formation of hybrid exon(s)
1	46	Sense	**TTCTCCAG**GCTAGAAGAACAA	106/148	1407207-1407227	1407228-1407233	6624 - 6225	2208 GAA (Glu) : 2209 CAA (Gln)	46-51; 46-53
2	46	Antisense	CTGCTCTTTTCCAGGTTCAAG	107/149	1407312-1407332	1407306-1407311	6714 - 6715	2238 CTT (Leu) : 2239 GAA (Glu)	46-53
3	47	Sense	**GTCTGTTTCAG**TTACTGGTGG	108/150	1409686-1409706	1409707-1409712	6769 - 6770	2257 G:TG (Val)	47-58
4	47	Antisense	TCCAGTTTCATTTAATTGTTT	109/151	1409736-1409756	1409730-1409735	6824 - 6825	2268 AAA (Lys) : 2267 CAA (Gln)	47-58
5	47	Antisense	CTTATGGGAGCACTTACAAGC	110/152	1409765-1409785	1409759-1409764	6833 - 6834	2278 CT:T (Leu)	47-58
6	49	Antisense	**TTGCTTCATTAC**CTTCACTGG	111/153	1502716-1502736	1502710-1502715	7194 - 7195	2398 CCA (Pro) : 2399 GTG (Val)	49-52 ; 49-53
7	51	Antisense	TTGTGTCACCAGAGTAACAGT	112/154	1565282-1565302	1565276-1564281	7323 - 7324	2441 ACT (Thr) : 2442 GTT (Val)	46-51
8	51	Antisense	AGTAACCACAGGTTGTGTCAC	113/155	1565294-1565314	1565288-1565293	7335 - 7336	2445 GTG (Val) : 2446 ACA (Thr)	46-51
9	52	Antisense	TTCAAATTTTGGGCAGCGGTA	114/156	1609765-1609785	1609759-1609764	7595 - 7596	2532 AC:C (Thr)	47-52

gRNA #	Exon cutting site #	Strand	gRNA target sequences* (excluding PAM)	SEQ ID NOs. Target/gRNA	gRNA target sequence position	PAM nts Position (cs: coding sequence, in: intron)	Cut sites in *DYS* gene**	Cut sites in amino acid sequence	gRNA involved in the formation of hybrid exon(s)
10	52	Sense	CAAGAGGCTAGAACAATCATT	115/157	1609802-1609822	1609823-1609828	7647 -7648	2549 ATC (Ile) : 2550 ATT (Ile)	49-52
11	53	Antisense	TTGTACTTCATCCCACTGATT	116/158	1659891-1659911	1659885-1659890	7677 - 7678	2559 AAT (Asn) : 2560 CAG (Gln)	49-53
12	53	Sense	CTTCAGAACCGGAGGCAACAG	117/159	1659918-1659938	1659939-1659944	7719 - 7720	2573 CAA (Gln) : 2574 CAG (Gln)	46-53
13	53	Sense	CAACAGTTGAATGAAATGTTA	118/160	1659933-1659953	1659954-1659959	7734 - 7735	2578 ATG (Met) : 2579 TTA(Leu)	46-53
14	53	Sense	GCCAAGCTTGAGTCATGGAAG	119/161	1660017-1660037	1660038-1660043	7818 - 7819	2606 TGG (Trp) : 2607 AAG (Lys)	46-53
15	53	Antisense	CTTGGTTTCTGTGATTTTCTT	120/162	1660068-1660088	1660062-1660067	7854 - 7855	2618 AAG (Lys) : 2619 AAA (Lys)	46-53
16	58	Sense	**TCATTTCACAG**GCCTTCAAGA	121/163	1860349-1860369	1860370-1860375	8554 - 8555	2852 A:AG (Lys)	47-58
17	58	Antisense	CAGAAATATTCGTACAGTCTC	122/164	1860411-1860431	1860405-1860410	8601 - 8602	2867 GAG (Gln) : 2868 ACT (Thr)	47-58
18	58	Antisense	**CAATTAC**CTCTGGGCTCCTGG	123/165	1860467-1860487	1860461-1860466	8657 - 8658	2886 CA:G (Gln)	47-58

*sequences shown in bold are intronic sequences (i.e., portions adjacent to the indicated exon)

SUBSTITUTE SHEET (RULE 26)

** Numbering with reference to the DYS cDNA sequence found at http://www.dmd.nl/seqs/murefDMD.html (Leiden Muscular Dystrophy pages, DMD cDNA Reference sequence). The reference sequence is based on GenBank file NM_004006.1 (with one difference 12505G>A), containing the Dp427m isoform (muscle) of dystrophin.

Table 7: Hybrid exon junctions following introduction of first and second cuts with various gRNA pairs

Combinations	Hybrid exons	Part of the upstream exon	Part of the downstream exon	Observation	New codon	New amino acid
gRNA 1 / gRNA 7	46 - 51	Glu	Val	Junction Glu - Val	None	None
gRNA 6 / gRNA 8	46 - 51	Pro	Thr	Junction Pro - Thr	None	None
gRNA 1 / gRNA 12	46 - 53	Glu	Gln	Junction Glu - Gln	None	None
gRNA 1 / gRNA 13	46 - 53	Glu	Gln	Junction Glu - Gln	None	None
gRNA 2 / gRNA 14	46 - 53	Leu	Lys	Junction Leu - Lys	None	None
gRNA 2 / gRNA 15	46 - 53	Leu	Lys	Junction Leu - Lys	None	None
gRNA 5 / gRNA 9	47 - 52	CT	C	CT + C = CTC	CTC	Leu
gRNA 6 /gRNA 10	49 - 52	Pro	Glu	Junction Pro - Glu	None	None
gRNA 6 / gRNA 11	49 - 53	Pro	Glu	Junction Pro - Glu	None	None
gRNA 3 / gRNA 16	47 - 58	G	AG	G + AG = GAG	GAG	Glu
gRNA 4 / gRNA 17	47 - 58	Gln	Thr	Junction Gln - Thr	None	None
gRNA 5 /gRNA 18	47 - 58	CT	G	CT + G = CTG	CTG	Leu

Table 8: Exemplary gRNAs targeting introns 22 and 23. Nucleotides position are provided with reference to the mouse DMD gene sequence ENSMUSG00000045103 (analysed using Benchling web tool (https://benchling.com/)

SUBSTITUTE SHEET (RULE 26)

#gRNA	intron cutting site	Strand	gRNA target sequences (excluding PAM)	SEQ ID NOs. Target/gRNA	gRNA target position	PAM nt position	Cut site localization
19	22	Sense	GAACTTCTATTTAATTTTG	206/214	83803285 - 83803303	83803304 - 83803306	83803300-83803301
20	23	Sense	**ATTTCAG**GTAAGCCGAGGTT	207/215	83803512 - 83803531	83803532 - 83803534	83803528-83803529
21	22	Sense	TCTTAATAATGTTTCACTGT	208/216	83803014 - 83803033	83803034 - 83803036	83803030-83803031
22	22	Antisense	ATAATTTCTATTATATTACA	209/217	83803137 - 83803156	83803134 - 83803136	83803139-83803140
23	22	Antisense	TTTCATTCATATCAAGAAGA	210/218	83803237 - 83803256	83803234 - 83803236	83803239-83803240
24	23	Antisense	ATAGTTTAAAGGCCAAACCT	211/219	83803527 - 83803546	83803524 - 83803526	83803529-83803530
25	23	Antisense	TGTGAAAAAATATAGTTTAA	212/220	83803538 - 83803557	83803535 - 83803537	83803540-83803541
26	23	Sense	**CGAAAATTTCAG**GTAAGCCG	213/221	83803507 - 83803526	83803527 - 83803529	83803523-83803524

*sequences shown in bold are exonic sequences (i.e., portions adjacent to the indicated intron)

Table 9: Sequences described herein

SEQ ID NO(s)	Description
1	Dystrophin DMD-001 cDNA Ensembl (ENSG00000198947) (from Start (ATG) to Stop (TAG) codon
2	Dystrophin protein sequence DMD-001 (Translation of SEQ ID NO: 1)
3	25 nts of 5' UTR + cDNA sequence of exon 1+ 25 nts of adjacent 3' intron sequence of Dystrophin transcript (DMD-001)
4-80	cDNA exon sequences (exons 2 to 78) of Dystrophin transcript (DMD-001) with flanking 25 nts of intron sequences on each side (5' and 3') of each exon
81	cDNA of exon 79 sequence flanked by 25 nts of adjacent intron sequence in 5' and 25 nts of 3'UTR sequence in 3'
82-105	gRNA target sequences on the Dystrophin gene listed in Table 3 (Example 2)
106-123	gRNA target sequences on the Dystrophin gene listed in Table 6. SEQ ID NO: 107 (target sequence of "gRNA3"); SEQ ID NO: 109 (target sequence of "gRNA5"); SEQ ID NO: 120 (Target sequence of "gRNA16"); and SEQ ID NO: 122 (target sequence of "gRNA18")
124-147	gRNA RNA sequences corresponding to the target sequences of SEQ ID NOs: 82-104 listed in Table 3 (Example 2)
148-165	gRNA RNA sequences of the target sequences of SEQ ID NOs: 105-122 listed in Table 6 (Example 19). SEQ ID NO: 149 ("gRNA3"); SEQ ID NO: 151 ("gRNA5"); SEQ ID NO: 162 ("gRNA16"); and SEQ ID NO: 164 ("gRNA18")
166	*S. pyogenes* Cas9 RNA recognition sequence (TracrRNA/crRNA)
167	Sequence of plasmid pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA (Addgene Plasmid # 61591)
168	Cpf1 recognition sequence (TracrRNA)
169	Protein sequence of humanized Cas 9 from *S. pyogenes* (without NLS and without TAG)
170	Protein sequence of humanized Cas9 from *S. pyogenes* (with NLS and without TAG)
171	Protein sequence of humanized Cas 9 from *S.* aureus (without NLS and without TAG)
172	Protein sequence of humanized Cas 9 from *S. aureus* (with NLS and without TAG)
173-177	Primer sequences listed in Example 1
178-195	Primer sequences listed in Example 8
196-198	Feldan Shuttles A, B and C, respectively
199-200	Expected and obtained sequence of hybrid exon 50-54 (FIG. 9)
201-203	Primers for amplification of *in vitro* transcription templates (Table 5, Example 8)
204-205	Primers used for PCR amplification Fw2-i22 (SEQ ID NO: 204); Rev1-i23 (SEQ ID NO: 205)
206-213	Target sequence of gRNAs listed in Table 8
214-221	RNA sequences of gRNA listed in Table 8
222	Primer used for HDMD Hybrid Exon Analysis by PCR (Fw-Int50))
223-227	Primers for PCR amplification of the hybrid introns (see Example 19)

228	Partial intron 22 and exon 23 sequence shown in FIG. 11i
229	Partial exon 23 and intron 23 sequence shown in FIG. 11j

[0280] Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

REFERENCES

1. Engel, A, and Banker, BQ (1986). Myology : basic and clinical, McGraw-Hill: New York.

2. Rybakova, IN, Patel, JR, and Ervasti, JM (2000). The dystrophin complex forms a mechanically strong link between the sarcolemma and costameric actin. J Cell Biol 150: 1209-1214.

3. Hoffman, EP, Brown, RH, Jr., and Kunkel, LM (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 51: 919-928.

4. Hoffman, EP, Brown, RH, and Kunkel, LM (1992). Dystrophin: the protein product of the Duchene muscular dystrophy locus. 1987 [classical article]. Biotechnology 24: 457-466.

5. Bladen, CL, Salgado, D, Monges, S, Foncuberta, ME, Kekou, K, Kosma, K, et al. (2015). The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. Hum Mutat 36: 395-402.

6. Koenig, M, Beggs, AH, Moyer, M, Scherpf, S, Heindrich, K, Bettecken, T, et al. (1989). The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. Am J Hum Genet 45: 498-506.

7. Hoffman, EP (1993). Genotype/phenotype correlations in Duchenne/Becker dystrophy. Mol Cell Biol Hum Dis Ser 3: 12-36.

8. Emery, AE (2002). The muscular dystrophies. Lancet 359: 687-695.

9. Duan, D (2011). Duchenne muscular dystrophy gene therapy: Lost in translation? Research and reports in biology 2011: 31-42.

10. Goyenvalle, A, Seto, JT, Davies, KE, and Chamberlain, J (2011). Therapeutic approaches to muscular dystrophy. Hum Mol Genet 20: R69-78.

11. Konieczny, P, Swiderski, K, and Chamberlain, JS (2013). Gene and cell-mediated therapies for muscular dystrophy. Muscle Nerve 47: 649-663.

12. Mendell, JR, et al. (2012). Gene therapy for muscular dystrophy: Lessons learned and path forward. Neurosci Lett 527: 90-99.

13. Verhaart, IE, and Aartsma-Rus, A (2012). Gene therapy for Duchenne muscular dystrophy. Curr Opin Neurol 25: 588-596.

14. Monaco, AP, Neve, RL, Colletti-Feener, C, Bertelson, CJ, Kurnit, DM, and Kunkel, LM (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 323: 646-650.

15. Kunkel, LM, Hejtmancik, JF, Caskey, CT, Speer, A, Monaco, AP, Middlesworth, W, et al. (1986). Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. Nature 322: 73-77.

16. Gregorevic, P, Blankinship, MJ, Allen, JM, Crawford, RW, Meuse, L, Miller, DG, et al. (2004). Systemic delivery of genes to striated muscles using adeno-associated viral vectors. Nat Med 10: 828-834.

17. Gregorevic, P, et al. (2006). rAAV6-microdystrophin preserves muscle function and extends lifespan in severely dystrophic mice. Nat Med 12: 787-789.

18. Wang, Z, Kuhr, CS, Allen, JM, Blankinship, M, Gregorevic, P, Chamberlain, JS, et al. (2007). Sustained AAV-mediated dystrophin expression in a canine model of Duchenne muscular dystrophy with a brief course of

immunosuppression. Mol Ther 15: 1160-1166.

19. Qiao, C, Koo, T, Li, J, Xiao, X, and Dickson, JG (2011). Gene therapy in skeletal muscle mediated by adeno-associated virus vectors. Methods Mol Biol 807: 119-140.

20. Aartsma-Rus, A (2012). Overview on DMD exon skipping. Methods Mol Biol 867: 97-116.

21. Aartsma-Rus, A, and van Ommen, GJ (2007). Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. RNA 13: 1609-1624.

22. Aartsma-Rus, A, and van Ommen, GJ (2009). Less is more: therapeutic exon skipping for Duchenne muscular dystrophy. Lancet Neurol 8: 873-875.

23. Dunckley, MG, et al. (1998). Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides. Hum Mol Genet 7: 1083-1090.

24. Lu, QL, Mann, CJ, Lou, F, Bou-Gharios, G, Morris, GE, Xue, SA, et al. (2003). Functional amounts of dystrophin produced by skipping the mutated exon in the mdx dystrophic mouse. Nat Med 9: 1009-1014.

25. Mann, CJ, Honeyman, K, McClorey, G, Fletcher, S, and Wilton, SD (2002). Improved antisense oligonucleotide induced exon skipping in the mdx mouse model of muscular dystrophy. J Gene Med 4: 644-654.

26. Takeshima, Y, Yagi, M, Wada, H, Ishibashi, K, Nishiyama, A, Kakumoto, M, et al. (2006). Intravenous infusion of an antisense oligonucleotide results in exon skipping in muscle dystrophin mRNA of Duchenne muscular dystrophy. Pediatr Res 59: 690-694.

27. van Deutekom, JC, Janson, AA, Ginjaar, IB, Frankhuizen, WS, Aartsma-Rus, A, Bremmer-Bout, M, et al. (2007). Local dystrophin restoration with antisense oligonucleotide PRO051. The New England journal of medicine 357: 2677-2686.

28. Kinali, M, Arechavala-Gomeza, V, Feng, L, Cirak, S, Hunt, D, Adkin, C, et al. (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. Lancet Neurol 8: 918-928.

29. Aartsma-Rus, A (2010). Antisense-mediated modulation of splicing: therapeutic implications for Duchenne muscular dystrophy. RNA biology 7: 453-461.

30. Ousterout, DG, Kabadi, AM, Thakore, PI, Perez-Pinera, P, Brown, MT, Majoros, WH, et al. (2015). Correction of dystrophin expression in cells from duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. Mol Ther 23: 523-532.

31. Rousseau, J, Chapdelaine, P, Boisvert, S, Almeida, LP, Corbeil, J, Montpetit, A, et al. (2011). Endonucleases: tools to correct the dystrophin gene. J Gene Med 13: 522-537.

32. Li, HL, Fujimoto, N, Sasakawa, N, Shirai, S, Ohkame, T, Sakuma, T, et al. (2015). Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. Stem cell reports 4: 143-154.

33. Ousterout, DG, Perez-Pinera, P, Thakore, PI, Kabadi, AM, Brown, MT, Qin, X, et al. (2013). Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. Mol Ther 21: 1718-1726.

34. Long, C, McAnally, JR, Shelton, JM, Mireault, AA, Bassel-Duby, R, and Olson, EN (2014). Prevention of

muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. Science 345: 1184-1188.

35. Nakamura, K, Fujii, W, Tsuboi, M, Tanihata, J, Teramoto, N, Takeuchi, S, et al. (2014). Generation of muscular dystrophy model rats with a CRISPR/Cas system. Scientific reports 4: 5635.

36. Ousterout, DG, Kabadi, AM, Thakore, PI, Majoros, WH, Reddy, TE, and Gersbach, CA (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. Nature communications 6: 6244.

37. Ran, FA, Cong, L, Yan, WX, Scott, DA, Gootenberg, JS, Kriz, AJ, et al. (2015). *In vivo* genome editing using Staphylococcus aureus Cas9. Nature 520: 186-191.

38. Cong, L, Ran, FA, Cox, D, Lin, S, Barretto, R, Habib, N, et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819-823.

39. Jinek, M, East, A, Cheng, A, Lin, S, Ma, E, and Doudna, J (2013). RNA-programmed genome editing in human cells. eLife 2: e00471.

40. Sander, JD, and Joung, JK (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32: 347-355.

41. Mali, P, Yang, L, Esvelt, KM, Aach, J, Guell, M, DiCarlo, JE, et al. (2013). RNA-guided human genome engineering via Cas9. Science 339: 823-826.

42. Cho, SW, Kim, S, Kim, JM, and Kim, JS (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol.

43. Doudna, JA, and Charpentier, E (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346: 1258096.

44. Zheng, Q, Cai, X, Tan, MH, Schaffert, S, Arnold, CP, Gong, X, et al. (2014). Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. Biotechniques 57: 115-124.

45. Deltcheva, E, Chylinski, K, Sharma, CM, Gonzales, K, Chao, Y, Pirzada, ZA, et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471: 602-607.

46. Marraffini, LA, and Sontheimer, EJ (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11: 181-190.

47. Jinek, M, Chylinski, K, Fonfara, I, Hauer, M, Doudna, JA, and Charpentier, E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816-821.

48. Canver, MC, Bauer, DE, Dass, A, Yien, YY, Chung, J, Masuda, T, et al. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J Biol Chem 289: 21312-21324.

49. Aartsma-Rus, A, Kaman, WE, Weij, R, den Dunnen, JT, van Ommen, GJ, and van Deutekom, JC (2006). Exploring the frontiers of therapeutic exon skipping for Duchenne muscular dystrophy by double targeting within one or multiple exons. Mol Ther 14: 401-407.

50. Beroud, C, Tuffery-Giraud, S, Matsuo, M, Hamroun, D, Humbertclaude, V, Monnier, N, et al. (2007). Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63% of patients with Duchenne muscular dystrophy. Hum Mutat 28: 196-202.

51. Skuk, D, and Tremblay, JP (2014). Clarifying misconceptions about myoblast transplantation in myology. Mol Ther 22: 897-898.

52. Skuk, D, and Tremblay, JP (2011). Intramuscular cell transplantation as a potential treatment of myopathies: clinical and preclinical relevant data. Expert Opin Biol Ther 11: 359-374.

53. Bruusgaard, JC, Liestol, K, Ekmark, M, Kollstad, K, and Gundersen, K (2003). Number and spatial distribution of nuclei in the muscle fibres of normal mice studied *in vivo*. J Physiol 551: 467-478.

54. Kinoshita, I, Vilquin, JT, Asselin, I, Chamberlain, J, and Tremblay, JP (1998). Transplantation of myoblasts from a transgenic mouse overexpressing dystrophin prduced only a relatively small increase of dystrophin- positive membrane. Muscle Nerve 21: 91-103.

55. Pavlath, GK, Rich, K, Webster, SG, and Blau, HM (1989). Localization of muscle gene products in nuclear domains. Nature 337: 570-573.

56. Nicolas, A, Raguenes-Nicol, C, Ben Yaou, R, Ameziane-Le Hir, S, Cheron, A, Vie, V, et al. (2015). Becker muscular dystrophy severity is linked to the structure of dystrophin. Hum Mol Genet 24: 1267-1279.

57. Kaspar, RW, Allen, HD, Ray, WC, Alvarez, CE, Kissel, JT, Pestronk, A, et al. (2009). Analysis of dystrophin deletion mutations predicts age of cardiomyopathy onset in becker muscular dystrophy. Circ Cardiovasc Genet 2: 544-551.

58. Ran, FA, Hsu, PD, Wright, J, Agarwala, V, Scott, DA, and Zhang, F (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8: 2281-2308.

59. t Hoen, PA, de Meijer, EJ, Boer, JM, Vossen, RH, Turk, R, Maatman, RG, et al. (2008). Generation and characterization of transgenic mice with the full-length human DMD gene. J Biol Chem 283: 5899-5907.

60. Mohanraju, P. et al. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 353(6299:aad5147.

61. Shmakov, S et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell 60(3):385-97.

62. Zetsche, B. et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163(3):759-71.

63. Kleinstiver, BP. et al. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol 33(12):1293-1298.

64. Tsai SQ. et al. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature Biotechnology, 32, 569–576.

65. Koo T. et al. (2015). Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. Mol Cells 38(6):475-481.

WHAT IS CLAIMED IS:

1. A method of modifying a dystrophin gene and restoring the correct reading frame for dystrophin expression within a cell having an endogenous frameshift or nonsense mutation within the dystrophin (DYS) gene, the method comprising:

a) introducing a first cut within an exon or an intron of the DYS gene creating a first exon end or intron end, wherein said first cut is located upstream of the endogenous frameshift or nonsense mutation;

b) introducing a second cut within an exon or an intron of the DYS gene creating a second exon end or second intron end, wherein said second cut is located downstream of the frameshift or nonsense mutation;

wherein upon ligation of said first and second exon ends or said first and second intron ends, a modified dystrophin gene comprising a hybrid exon or intron is created and dystrophin expression is restored.

2. The method of claim 1, wherein said first and second cuts are introduced by providing a cell with i) a CRISPR nuclease; and ii) a pair of gRNAs consisting of a) a first gRNA which binds to an exon or intron sequence of the DYS gene located upstream of the endogenous frameshift or nonsense mutation for introducing the first cut; b) a second gRNA which binds to an exon or intron sequence of the DYS gene located downstream of the endogenous frameshift or nonsense mutation for introducing the second cut.

3. The method of claim 1 or 2, wherein the endogenous frameshift or nonsense mutation is located within a region of the dystrophin gene encompassing exons 45-58 of the dystrophin gene.

4. The method of any one of claims 1 to 3, wherein (i) the first cut is within exon 46, 47, 48, 49 or 50 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene, or (ii) the first cut is within intron 22 and the second cut is within intron 23, of the dystrophin gene.

5. The method of any one of claims 1 to 4, wherein the first cut is within exon 47 and the second cut is within exon 58.

6. The method of any one of claims 1 to 5, wherein said modified dystrophin gene encodes a modified dystrophin protein comprising a hybrid spectrin-like repeat (SLR) comprising a portion of a first SLR and a portion of second SLR, thereby comprising a hybrid SLR junction, and wherein the wild-type configuration of said hybrid SLR is maintained where hydrophobic amino acids are localized in position “a” and “d” of the heptad motif.

7. The method of any one of claims 1 to 6, wherein said modified dystrophin gene encodes a modified dystrophin protein comprising the following hybrid spectrin-like repeat (SLR): (i) a hybrid SLR comprising a portion of SLR 17 and a portion of SLR 19 (SLR 17-19); (ii) a hybrid SLR comprising a portion of SLR 17 and a portion of SLR 20 (SLR 17-20); (iii) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 20 (SLR 18-20); (iv) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 21 (SLR 18-21); (v) a hybrid SLR comprising a portion of SLR 19 and a portion of SLR 20 (SLR 19-20); or (vi) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 23 (SLR 18-23).

8. The method of claim 7, wherein said hybrid SLR has a hybrid SLR junction selected from the hybrid SLR junctions set forth in FIGs. 5a, 5b and 16b.

9. The method of any one of claims 1 to 8, wherein a hybrid exon is generated and said hybrid exon has a nucleic acid sequence as set forth in FIG. 19(a)(ii) or FIG. 17.

10. The method of any one of claims 1 to 9, wherein each of said first gRNA and said second gRNA has a target sequence adjoining a PAM sequence as set forth in Table 6, Table 8, FIG. 1 or FIG. 11.

11. The method of any one of claims 1 to 9, wherein the pair of gRNA is selected from a gRNA pair set forth in FIG. 4, 13, 15, 17, 21 or 22 or Table 4 or 7.

12. The method of any one of claims 1 to 9, wherein said first gRNA and said second gRNA are selected from the gRNAs listed in Table 3, 6 or 8.

13. The method of any one of claims 1 to 12, wherein said first gRNA has a target sequence which is GTCTGTTTCAGTTACTGGTGG (SEQ ID NO: 108) and said second gRNA has a target sequence which is TCATTTCACAGGCCTTCAAGA (SEQ ID NO: 121).

14. The method of any one of claims 1 to 12, wherein said first gRNA has a target sequence which is CTTATGGGAGCACTTACAAGC (SEQ ID NO: 110) and said second gRNA has a target sequence which is CAATTACCTCTGGGCTCCTGG (SEQ ID NO: 123).

15. The method of any one of claims 1 to 3, wherein the gRNA pair consists of a first gRNA having a first target sequence which is TCTTAATAATGTTTCACTGT (SEQ ID NO: 208) and a second gRNA having a second target sequence which is TTTCATTCATATCAAGAAGA (SEQ ID NO: 210).

16. The method of any one of claims 1 to 3, wherein the gRNA pair consists of a first gRNA having a first target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO: 209) and a second gRNA having a second target sequence which is ATAGTTTAAAGGCCAAACCT (SEQ ID NO: 211).

17. The method of any one of claims 1 to 3, wherein the gRNA pair consists of a first gRNA having a first target sequence which is TTTCATTCATATCAAGAAGA (SEQ ID NO: 210) and a second gRNA having a second target sequence which is ATAGTTTAAAGGCCAAACCT (SEQ ID NO: 211).

18. The method of any one of claims 1 to 3, wherein the gRNA pair consists of a first gRNA having a first target sequence which is TCTTAATAATGTTTCACTGT (SEQ ID NO: 208) and a second gRNA having a second target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO: 209).

19. The method of any one of claims 2 to 18, wherein said first gRNA, said second gRNA and/or said CRISPR nuclease are delivered into the cell using one or more adeno-associated virus (AAV) vectors.

20. The method of any one of claims 2 to 18, wherein said first gRNA, said second gRNA and/or said CRISPR nuclease are delivered into said cell using one or more polypeptide-based shuttles (e.g., Feldan Shuttle).

21. The method of claim 18, wherein the one or more polypeptide-based shuttles comprise:

(i) a polypeptide-based shuttle having amino acid sequence HHHHHHKWKLFKKIGAVLKVLTTGYARAAARQARA (FSA);

(ii) a polypeptide-based shuttle having amino acid sequence HHHHHHKWKLFKKIGAVLKVLTTGYARAAARQARAHHHHHH (FSB);

(iii) a polypeptide-based shuttle having amino acid sequence HHHHHHLLKLWSRLLKLWTQGRRLKAKRAKAHHHHHH (FSC); or

(iv) any combination of at least two of (i), (ii) and (iii).

22. A gRNA pair for restoring dystrophin expression in a cell comprising an endogenous frameshift or nonsense mutation within the dystrophin (DYS) gene, wherein said pair consists of a first gRNA and a second gRNA, wherein said first gRNA comprises a first target sequence upstream of the endogenous frameshift or nonsense mutation and can direct a nuclease-mediated first cut in an exon or intron sequence of the DYS gene located upstream of the endogenous frameshift or nonsense mutation and wherein said second gRNA comprises a second target sequence downstream of the endogenous frameshift or nonsense mutation and can direct a nuclease-mediated second cut in an exon or intron sequence of the DYS gene located downstream of the endogenous frameshift or nonsense mutation.

23. The gRNA pair of claim 22, wherein the first cut is (i) within exon 46, 47, 48, 49 or 50 and the second cut is within exon 51, 52, 53, 54, 55, 56, 57 or 58, of the dystrophin gene, or (ii) the first cut is (i) within intron 22 and the second cut is within intron 23, of the dystrophin gene.

24. The gRNA pair of claim 22 or 23, wherein said gRNA pair allows to generate a modified dystrophin gene encoding a modified dystrophin protein comprising a hybrid spectrin-like repeat (SLR) comprising a portion of a first SLR and a portion of second SLR and a hybrid SLR junction and wherein the hybrid SLR has a wild-type configuration where hydrophobic amino acids are localized in position “a” and “d” of the heptad motif.

25. The gRNA pair of claim 24, wherein said modified dystrophin protein comprises the following hybrid spectrin-like repeat (SLR): (i) a hybrid SLR comprising a portion of SLR 17 and a portion of SLR 19 (SLR 17-19); (ii) a hybrid SLR comprising a portion of SLR 17 and a portion of SLR 20 (SLR 17-20); (iii) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 20 (SLR 18-20); (iv) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 21 (SLR 18-21); (v) a hybrid SLR comprising a portion of SLR 19 and a portion of SLR 20 (SLR 19-20); or (vi) a hybrid SLR comprising a portion of SLR 18 and a portion of SLR 23 (SLR 18-23).

26. The gRNA pair of claim 24 or 25, wherein said hybrid SLR has a hybrid SLR junction selected from the hybrid SLR junctions set forth in FIGs. 5a, 5b and 16b.

27.The gRNA pair of any one of claims 22 to 26, wherein said gRNA pair allows to generate a hybrid exon from (i) exon 46 and exon 51 (hybrid exon 46-51); (ii) exon 46 and exon 53 (hybrid exon 46-53); (iii) exon 47 and exon 52 (hybrid exon 47-52); (iv) exon 49 and exon 52 (hybrid exon 49-52); (v) exon 49 and exon 53 (hybrid exon 49-53); or (vi) exon 47 and exon 58 (hybrid exon 47-58).

28. The gRNA pair of claim 27, wherein said gRNA pair allows to generate a hybrid exon in said DYS gene having a nucleic acid sequence as set forth in FIG. 19(a)(ii) or FIG. 17.

29. The gRNA pair of any one of claims 22 to 27, wherein each of said first target sequence and said second target sequence of said first and second gRNAs is adjoining a PAM sequence in said DYS gene set forth in Table 6, Table 8, FIG. 1 or FIG. 11.

30. The gRNA pair of any one of claims 22 to 27, wherein said pair is selected from a gRNA pair set forth in FIG. 4, 13, 15, 17, 21 or 22 or Table 4 or 7.

31. The gRNA pair of any one of claims 22 to 27, wherein said first gRNA and said second gRNA are selected from the gRNAs listed in Table 3, 6 or 8 or FIG. 11.

32. The gRNA pair of claim 31, wherein said first target sequence of said first gRNA is GTCTGTTTCAGTTACTGGTGG (SEQ ID NO: 108) and said second target sequence of said second gRNA is TCATTTCACAGGCCTTCAAGA (SEQ ID NO: 121).

33. The gRNA pair of claim 31, wherein said first target sequence of said first gRNA is CTTATGGGAGCACTTACAAGC (SEQ ID NO: 110) and said second target sequence of said second gRNA is CAATTACCTCTGGGCTCCTGG (SEQ ID NO: 123).

34 The gRNA pair of claim 31, wherein the gRNA pair consists of a first gRNA having a first target sequence which is ATTTCAGGTAAGCCGAGGTT (SEQ ID NO: 207) and a second gRNA having a second target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO:209).

35. The gRNA pair of claim 31, wherein the gRNA pair consists of a first gRNA having a first target sequence which is ATTTCAGGTAAGCCGAGGTT (SEQ ID NO: 207) and a second gRNA having a second target sequence which is TTTCATTCATATCAAGAAGA (SEQ ID NO: 210).

36. The gRNA pair of claim 31, wherein the gRNA pair consists of a first gRNA having a first target sequence which is, TTTCATTCATATCAAGAAGA (SEQ ID NO: 210) and a second gRNA having a second target sequence which is ATAGTTTAAAGGCCAAACCT (SEQ ID NO: 211).

37. The gRNA pair of claim 31, wherein the gRNA pair consists of a first gRNA having a first target sequence which is TCTTAATAATGTTTCACTGT (SEQ ID NO: 208) and a second gRNA having a second target sequence which is ATAATTTCTATTATATTACA (SEQ ID NO: 209).

38. A nucleic acid comprising one or more sequences encoding one or both members of the gRNA pair of any one of claims 22 to 37.

39. The nucleic acid of claim 38, further comprising a sequence encoding a CRISPR nuclease.

40. A nucleic acid comprising a modified dystrophin gene comprising ligated first and second exon ends or first and second intron ends as defined in any one of claims 1 to 21.

41. The nucleic acid of claim 40, wherein the modified dystrophin gene comprises ligated first and second exon ends defined by the cut sites shown in Table 3 or 6 or FIG. 11b, c, d, e, f, g, or h.

42. The nucleic acid of claim 41, wherein the first cut site is between nucleotides 6769 and 6770 of the DYS gene and the second cut site is between nucleotides 8554 and 8555 of the DYS gene.

43. The nucleic acid of claim 41, wherein the first cut site is between nucleotides 6833 and 6834 of the DYS gene and the second cut site is between nucleotides 8657 and 8658 of the DYS gene.

44. The nucleic acid of claim 40, wherein the modified dystrophin gene comprises ligated first and second intron ends defined by the cut sites shown in Table 8 or FIG. 11i or j.

45. A nucleic acid comprising a modified dystrophin gene comprising a hybrid exon sequence set forth in FIG. 5a, 9a(i) or 17.

46. A nucleic acid comprising a modified dystrophin gene comprising a polynucleotide sequence encoding a hybrid spectrin-like repeat (SLR) having a polypeptide sequence set for in FIG. 16b.

47. A modified dystrophin polypeptide encoded by the nucleic acid of any one of claims 40 to 46.

48. A vector comprising the nucleic acid of any one of claims 38 to 46.

49. The vector of claim 48, wherein said vector is a viral vector.

50. The vector of claim 49, wherein said vector is an AAV or a Sendai virus derived vector.

51. A cell comprising one or both members of the gRNA pair of any one of claims 22 to 37, the nucleic acid of any one of claims 38 to 46, the polypeptide of claim 47 and/or the vector of any one of claims 48 to 50.

52. A composition comprising one or both members of the gRNA pair of any one of claims 22 to 37, the nucleic acid of any one of claims 38 to 46, the polypeptide of claim 47, the vector of any one of claims 48 to 50, and/or the cell of claim 47.

53. The composition of claim 52, further comprising a CRISPR nuclease or a nucleic acid encoding a CRISPR nuclease.

54. The composition of claim 52 or 53, further comprising a pharmaceutically acceptable carrier.

55. A kit comprising one or both members of the gRNA pair of any one of claims 22 to 37, the nucleic acid of any one of claims 38 to 46, the polypeptide of claim 47, the vector of any one of claims 48 to 50, the cell of claim 51, the composition of any one of claims 52 to 54, a CRISPR nuclease and/or a nucleic acid encoding a CRISPR nuclease.

56. The kit of claim 55, further comprising instructions for performing the method of any one of claims 1 to 21.

57. The kit of claim 55 or 56, for use in treating muscular dystrophy in a subject in need thereof.

58. The kit of claim 57, wherein the muscular dystrophy is Duchenne muscular dystrophy.

59. A method for treating muscular dystrophy in a subject, comprising modifying a dystrophin gene and restoring the correct reading frame for dystrophin expression within a cell of said subject according to the method of any one of claims 1 to 21.

60. A method for treating muscular dystrophy in a subject, comprising contacting a cell of the subject with (i)(a) the gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) the composition of claim 53.

61. The method of claim 59 or 60, wherein the muscular dystrophy is Duchenne muscular dystrophy.

62. Use of (i)(a) the gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) the composition of claim 53, for treating muscular dystrophy in a subject.

63. Use of (i)(a) the gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) the composition of claim 53, for the preparation of a medicament for treating muscular dystrophy in a subject.

64. The use of claim 62 or 63, wherein the muscular dystrophy is Duchenne muscular dystrophy.

65. The (i)(a) gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) the composition of claim 53, for use in treating muscular dystrophy in a subject.

66. The (i)(a) gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) the composition of claim 53, for use in the preparation of a medicament for treating muscular dystrophy in a subject.

67. The (i)(a) gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide or (ii) the composition of claim 53, for use of claim 65 or 66, wherein the muscular dystrophy is Duchenne muscular dystrophy.

68. A reaction mixture comprising (a) the gRNA pair of any one of claims 22 to 37 or one or more nucleic acids encoding said gRNA pair and (b) a CRISPR nuclease polypeptide or a nucleic acid encoding a CRISPR nuclease polypeptide.

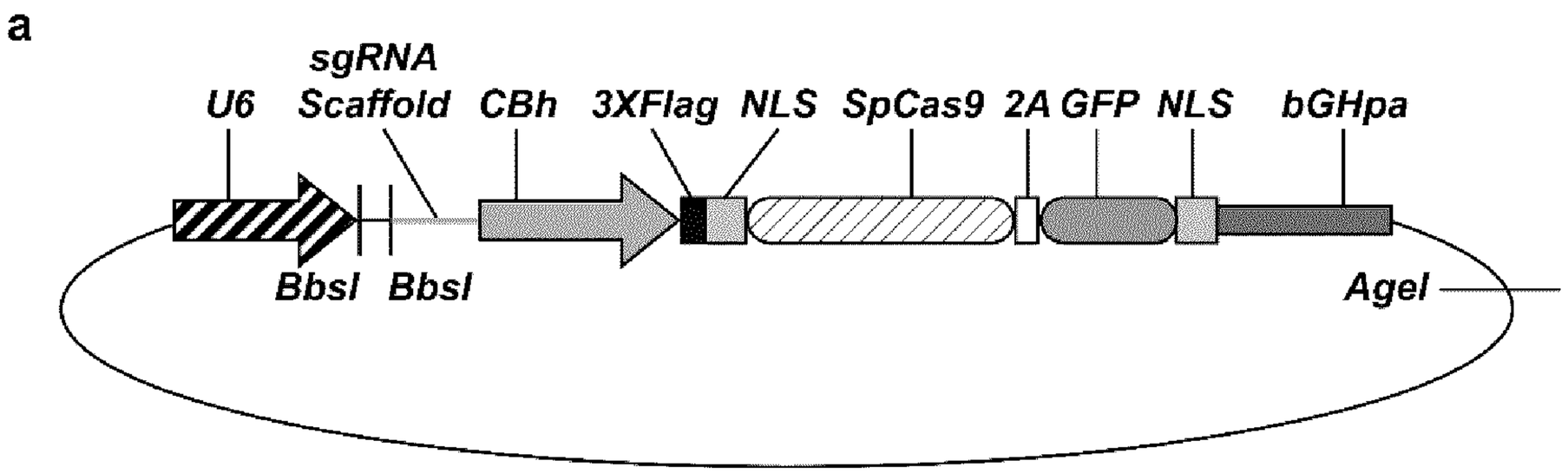

a
sgRNA
U6
Scaffold
CBh
3XFlag
NLS
SpCas9
2A
GFP
NLS
bGHpa
BbsI
BbsI
AgeI
b
Exon 50
1 2 3
5′AGGAAGTTAGAAGATCTGAGCTCTGAGTGGAAGGCGGTAAACCGTTTACTTCAAGAGCTG 3′
3′TCCTTCAATCTTCTAGACTCGAGACTCACCTTCCGCCATTTGGCAAATGAAGTTCTCGAC 5′
7
4
5
6
5′AGGGCAAAGCAGCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCT 3′
3′TCCCGTTTCGTCGGACTGGATCGAGGACCTGACTGGTGATAACCTCGGA 5′
8 9 10
C
Exon 54
1
2
5′ CAGTTGGCCAAAGACCTCCGCCAGTGCCAGACAAATGTAGATGTGCCAAATG 3′
3′ GTCAACCGGTTTCTGGAGGCGGTCACCGTCTGTTTACATCTACACCGTTTAC 5′
6 7 8 9
3
4
5′ ACTTCGCCCTGAAACTTCTCCGGCATTATTCTGCAGATGATACCAGAAAAGTCCAC 3′
3′ TGAACCGGGACTTTGAAGAGGCCCTAATAAGACGTCTACTATGGTCTTTTCAGGTG 5′
10 11 12 13
5
5′ATGATAACAGAGAATATCAATGCCTCTTGGAGAAGCATTCATAAAAG 3′
3′TGTCTCTTTACTATATAGTTACGGAGAACCTCTTCGTAAGTATTTTC 5′
14

FIG. 1

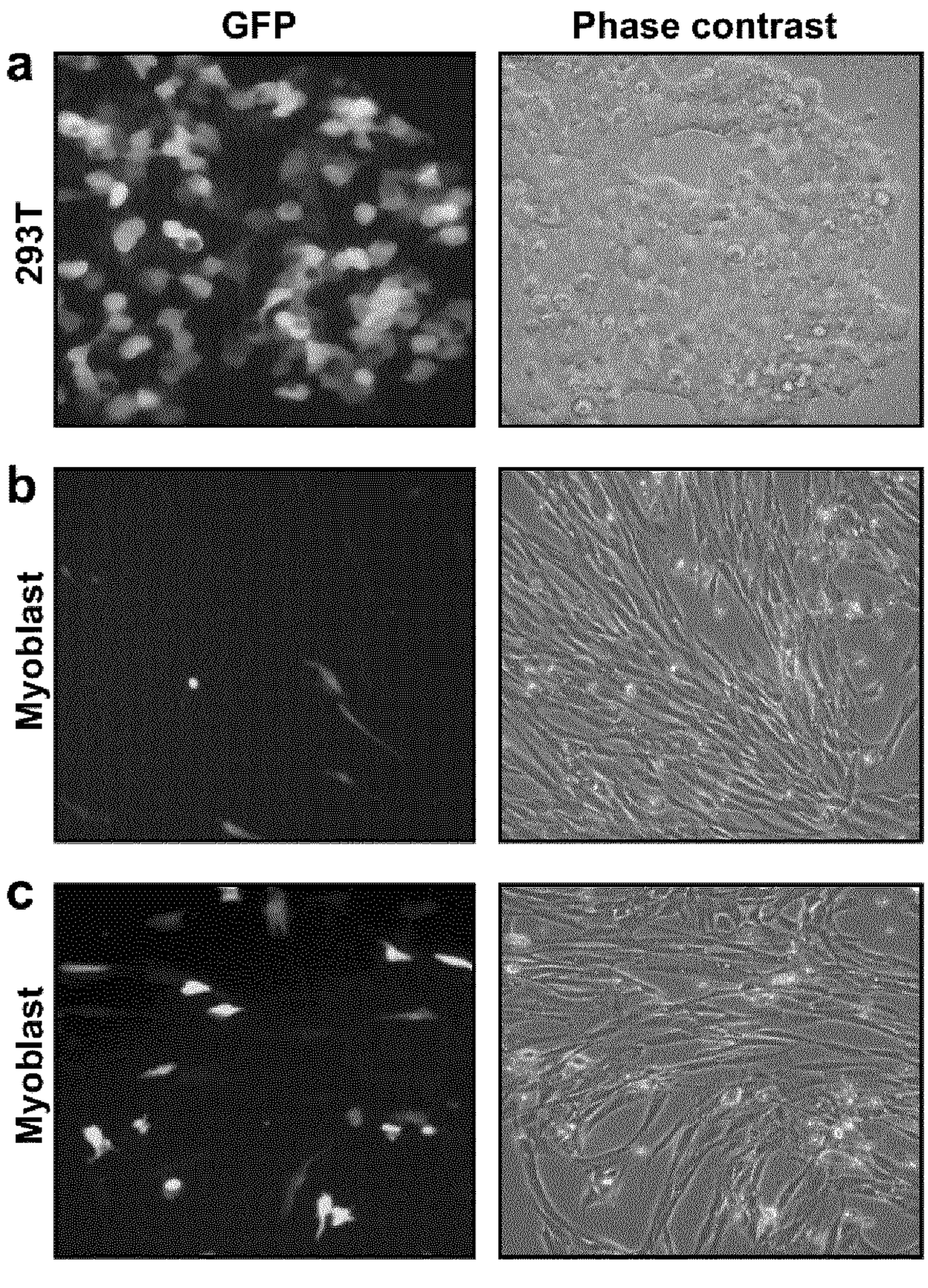


FIG. 2

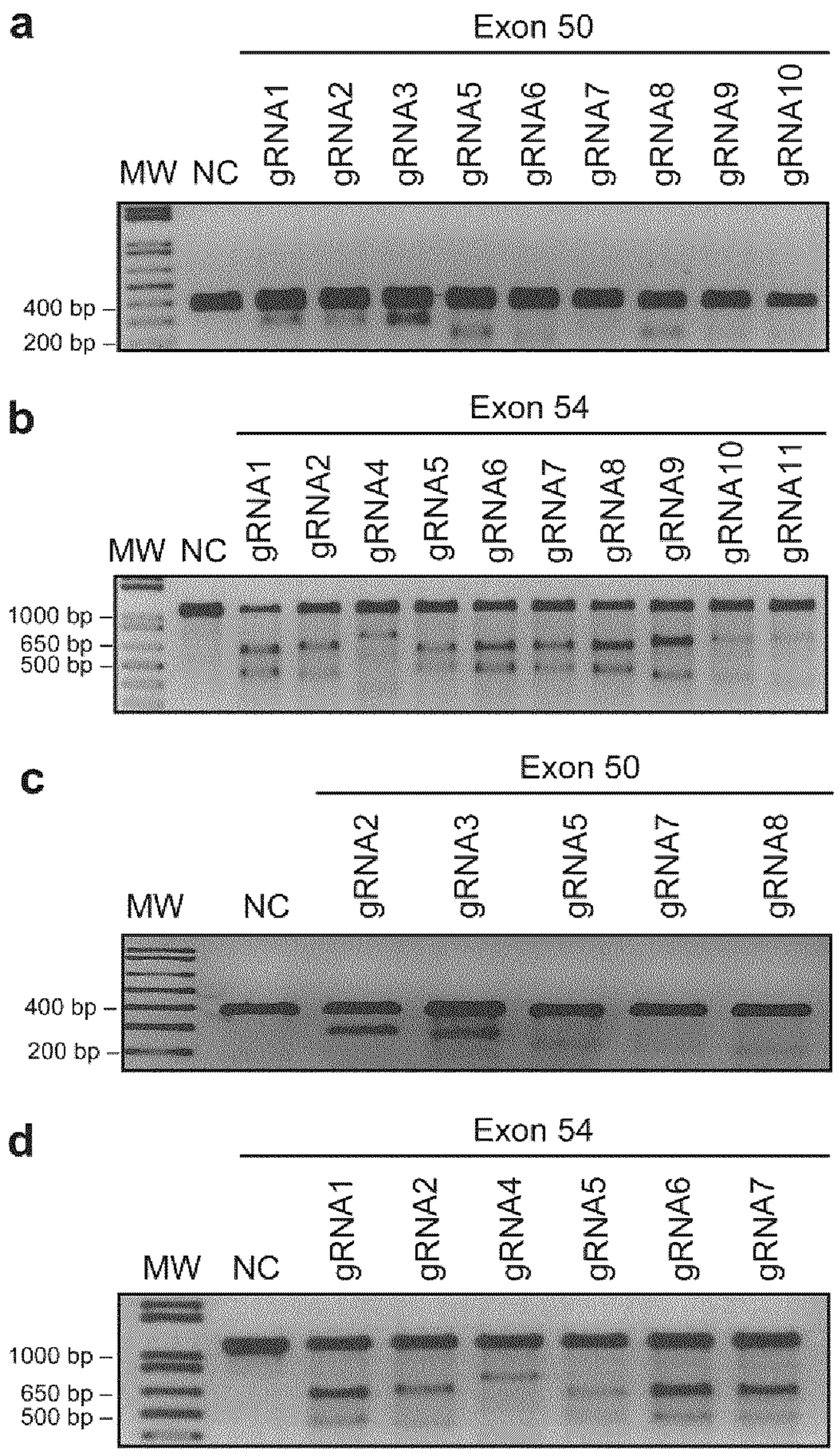

a
Exon 50
MW
NC
gRNA1
gRNA2
gRNA3
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
400 bp
200 bp
b
Exon 54
MW
NC
gRNA1
gRNA2
gRNA4
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
gRNA11
1000 bp
650 bp
500 bp
c
Exon 50
MW
NC
gRNA2
gRNA3
gRNA5
gRNA7
gRNA8
400 bp
200 bp
d
Exon 54
MW
NC
gRNA1
gRNA2
gRNA4
gRNA5
gRNA6
gRNA7
1000 bp
650 bp
500 bp

FIG. 3

a

EX 54 / EX 50	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	⊠				⊠									
2		⊠	▒	▒		▒				■		■		▒
3			▒	▒		▒				▒		▒		▒
4		▒	▒	▒		▒				▒		▒		▒
5	⊠						▒	▒	▒		▒		▒	
6							▒	▒	▒		▒		▒	
7							▒	▒	▒		▒		▒	
8	▒				▒									
9							▒	▒	▒		▒		▒	
10	▒				▒									

▒ Junction of amino acids — Out-of-frame

▒ Generation of a new codon — ■ Generation of a new stop codon

⊠ Pairs tested and sequenced

b

	gRNA pairs	Size of rem. Ex 50 amp. (Nb of nt en ex 50)	Size of rem. Ex 54 amp. (Nb of nt in ex 54)	Size of hybrid ex 50-54 amp. (Nb of nt in hybrid ex 50-54)	Restore or not the reading frame
▒	1-50/5-54	105 bp (24 nucleotides)	380 bp (23 nucleotides)	485 bp (47 nucleotides)	yes
▒	2-50/2-54	109 bp (28 nucleotides)	472 bp (115 nucleotides)	581 bp (143 nucleotides)	yes
	5-50/1-54	164 bp (83 nucleotides)	491 bp (134 nucleotides)	655 bp (217 nucleotides)	no
■	2-50/10-54	109bp (28 nucleotides)	503 bp (119 nucleotides)	612 bp (147 nucleotides)	no

c

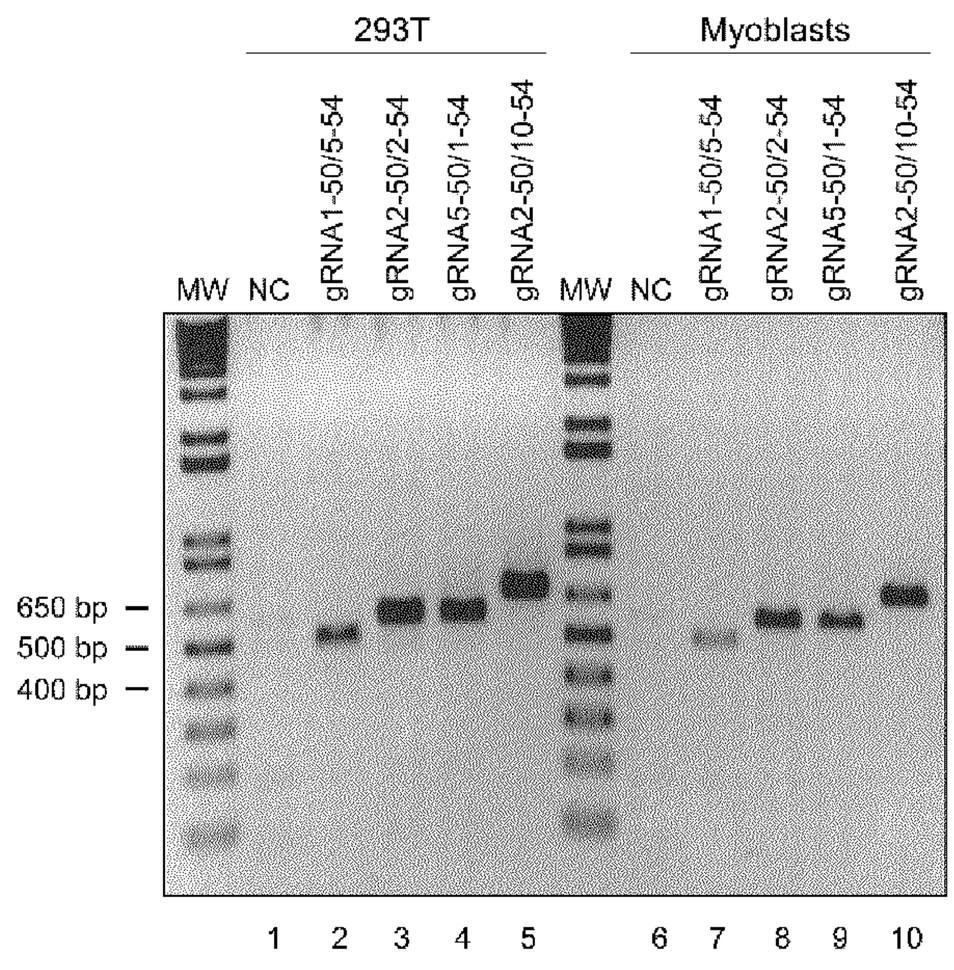
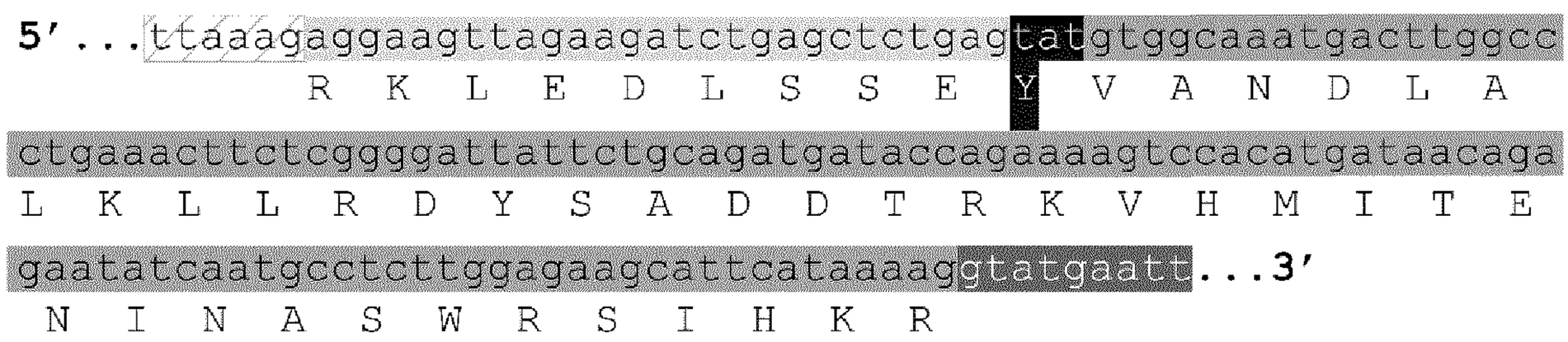


FIG. 4

a



b

Sequences obtained	Nb clones (%)
In frame hybrid 50/54	**28 (62%)**
RKLEDLSSEYVANDLALKLLRDYSADDTRKVHMITENINASWRSIHK	25 (56%)
RKLEDLSEYVANDLALKLLRDYSADDTRKVHMITENINASWRSIHK	1 (2%)
RKLEDLSYVANDLALKLLRDYSADDTRKVHMITENINASWRSIHK	1 (2%)
RKLEDLSSWKLLRDYSADDTRKVHMITENINASWRSIHK	1 (2%)
Out-of-frame hybrid 50/54	**17 (38%)**
Total	**45 (100%)**

Intron 49 Remaining exon 54

Remaining exon 50 Intron 54

FIG. 5

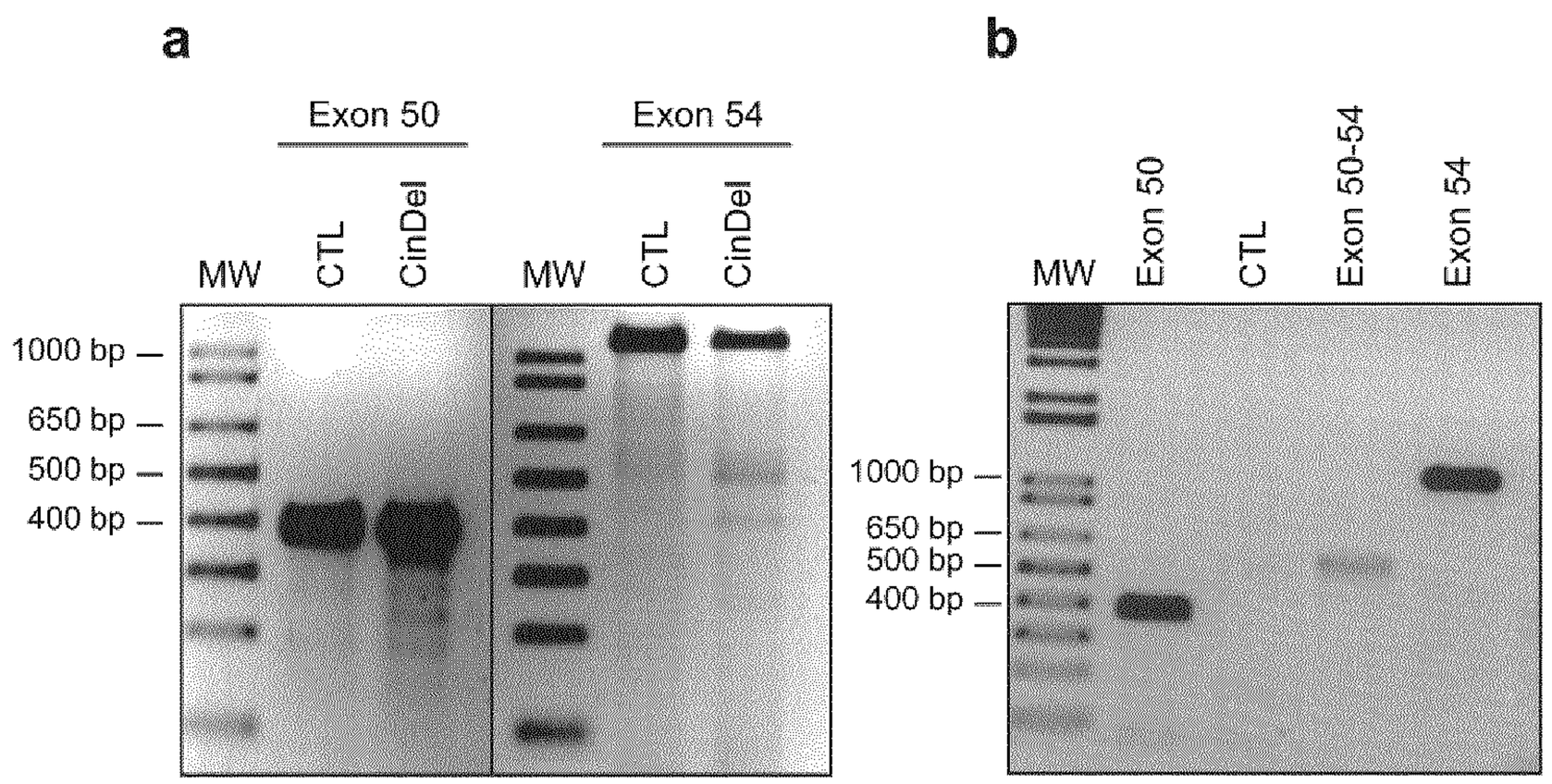

a
Exon 50
Exon 54
MW
CTL
CinDel
MW
CTL
CinDel
1000 bp
650 bp
500 bp
400 bp
b
MW
Exon 50
CTL
Exon 50-54
Exon 54
1000 bp
650 bp
500 bp
400 bp

FIG. 6

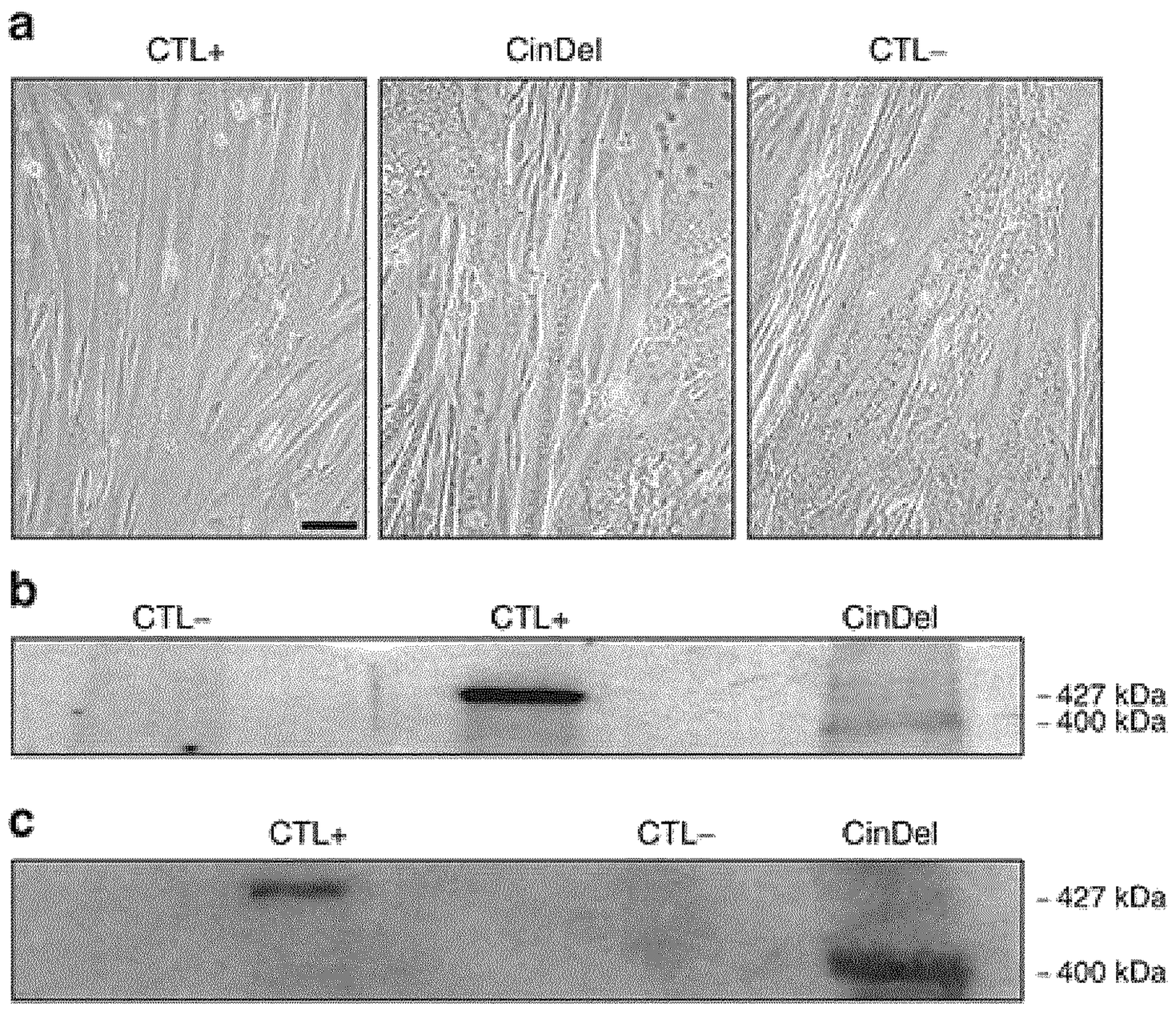

a
CTL+
CinDel
CTL−
b
CTL−
CTL+
CinDel
427 kDa
400 kDa
c
CTL+
CTL−
CinDel
427 kDa
400 kDa

FIG. 7

CA 3037635 2019-03-20

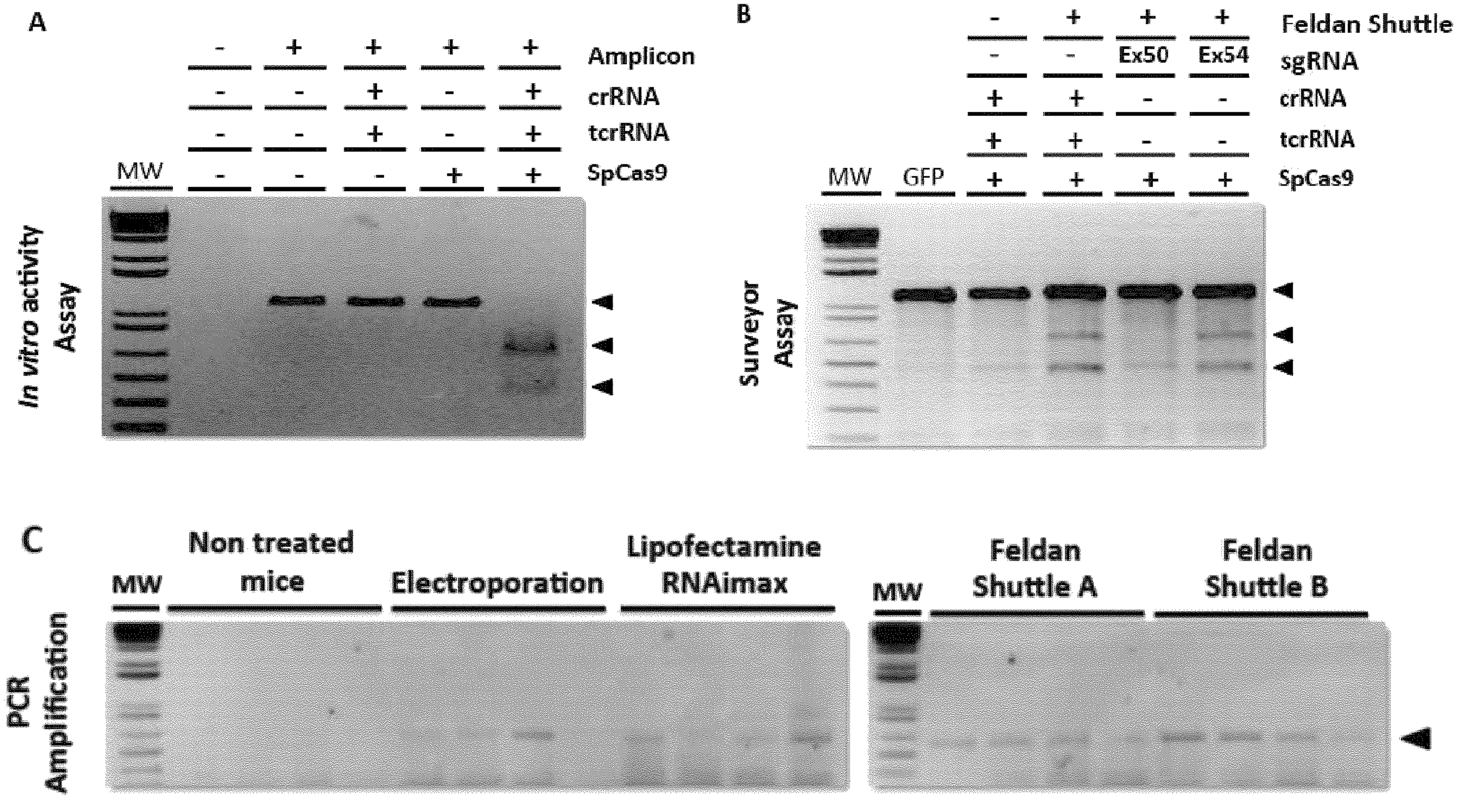


FIG. 8

A

(i)

ttcaccaaatggattaagatgttcatgaattatcttcaaagtgttaatcgaataagtaatgtgtatgcttttctgttaaag***AGGAAG***
TTAGAAGATCTGAGCTCTTCT**TGGAGAAGCATTCATAAAAG**gtatgaattacattatttctaaaactactgttggctgtaataatgg
ggtggtgaaactggatggaccatgaggatttgttttccaatccagctaaactggagcttgggagggttcaagacgataaataccaa
ctaaactcacggacttggctcagacttctattttaaaaacgaggaacataagatctcatttgcccgctgtcacaaaagtagtgacat
aaccaagagattaaacaaaaagcaaaatactgatttatagctagaagagccatttatcagtctactttgataactctatccaaagga
atatctttctatctcatcatggcgcacactgccttacctgttatctgata

(ii)

tatgcgcgaccgagttgctcttgccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaa
aacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatctt
cagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacgga
aatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaat
gtatttagaaaaataaacaaatagggqttccgcgcacatttccccgactcgagaattctgacgtcattattatcagttcaccaaatg
gattaagatgttcatgaattatcttcaaagtgttaatcgaataagtaatgtgtatgcttttctgttaaag***AGGAAGTTAGAAGATCT***
GAGCTCTTCT**TGGAGAAGCATTCATAAAAG**gtatgaattacattatttctaaaactactgttggctgtaataatggggtggtgaaac
tggatggaccatgaggatttgttttccaatccagctaaactggagcttgggagggttcaagacgataaataccaactaaactcacg
gacttggctcagacttctattttaaaaacgaggaacataagatctcatttgcccgctgtcacaaaagtagtgacataaccaagagat
taaacaaaaagcaaaatactgatttatagctagaagagccatttatcagtctactttgataactctatccaaaggaatatctttcta
tctcatcatggcgcacactgccttacctgttatctgataatcatggtttttacctcctgaattcggatccctcgagcgatacagcga
gtcgctctttgtggaggtgacgattacctaacaaaatcggacn

B

Range 1: 425 to 909 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
896 bits(485)	0.0	485/485(100%)	0/485(0%)	Plus/Plus

```
Query  1    TTCACCAAATGGATTAAGATGTTCATGAATTATCTTCAAAGTGTTAATCGAATAAGTAAT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  425  TTCACCAAATGGATTAAGATGTTCATGAATTATCTTCAAAGTGTTAATCGAATAAGTAAT  484

Query  61   GTGTATGCTTTTCTGTTAAAGAGGAAGTTAGAAGATCTGAGCTCTTCTTGGAGAAGCATT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  485  GTGTATGCTTTTCTGTTAAAGAGGAAGTTAGAAGATCTGAGCTCTTCTTGGAGAAGCATT  544

Query  121  CATAAAAGGTATGAATTACATTATTTCTAAAACTACTGTTGGCTGTAATAATGGGGTGGT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  545  CATAAAAGGTATGAATTACATTATTTCTAAAACTACTGTTGGCTGTAATAATGGGGTGGT  604

Query  181  GAAACTGGATGGACCATGAGGATTTGTTTTTCCAATCCAGCTAAACTGGAGCTTGGGAGG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  605  GAAACTGGATGGACCATGAGGATTTGTTTTTCCAATCCAGCTAAACTGGAGCTTGGGAGG  664

Query  241  GTTCAAGACGATAAATACCAACTAAACTCACGGACTTGGCTCAGACTTCTATTTTAAAAA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  665  GTTCAAGACGATAAATACCAACTAAACTCACGGACTTGGCTCAGACTTCTATTTTAAAAA  724

Query  301  CGAGGAACATAAGATCTCATTTGCCCGCTGTCACAAAAGTAGTGACATAACCAAGAGATT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  725  CGAGGAACATAAGATCTCATTTGCCCGCTGTCACAAAAGTAGTGACATAACCAAGAGATT  784

Query  361  AAACAAAAAGCAAAATACTGATTTATAGCTAGAAGAGCCATTTATCAGTCTACTTTGATA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  785  AAACAAAAAGCAAAATACTGATTTATAGCTAGAAGAGCCATTTATCAGTCTACTTTGATA  844

Query  421  ACTCTATCCAAAGGAATATCTTTCTATCTCATCATGGCGCACACTGCCTTACCTGTTATC  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  845  ACTCTATCCAAAGGAATATCTTTCTATCTCATCATGGCGCACACTGCCTTACCTGTTATC  904

Query  481  TGATA  485
            |||||
Sbjct  905  TGATA  909
```

FIG. 9

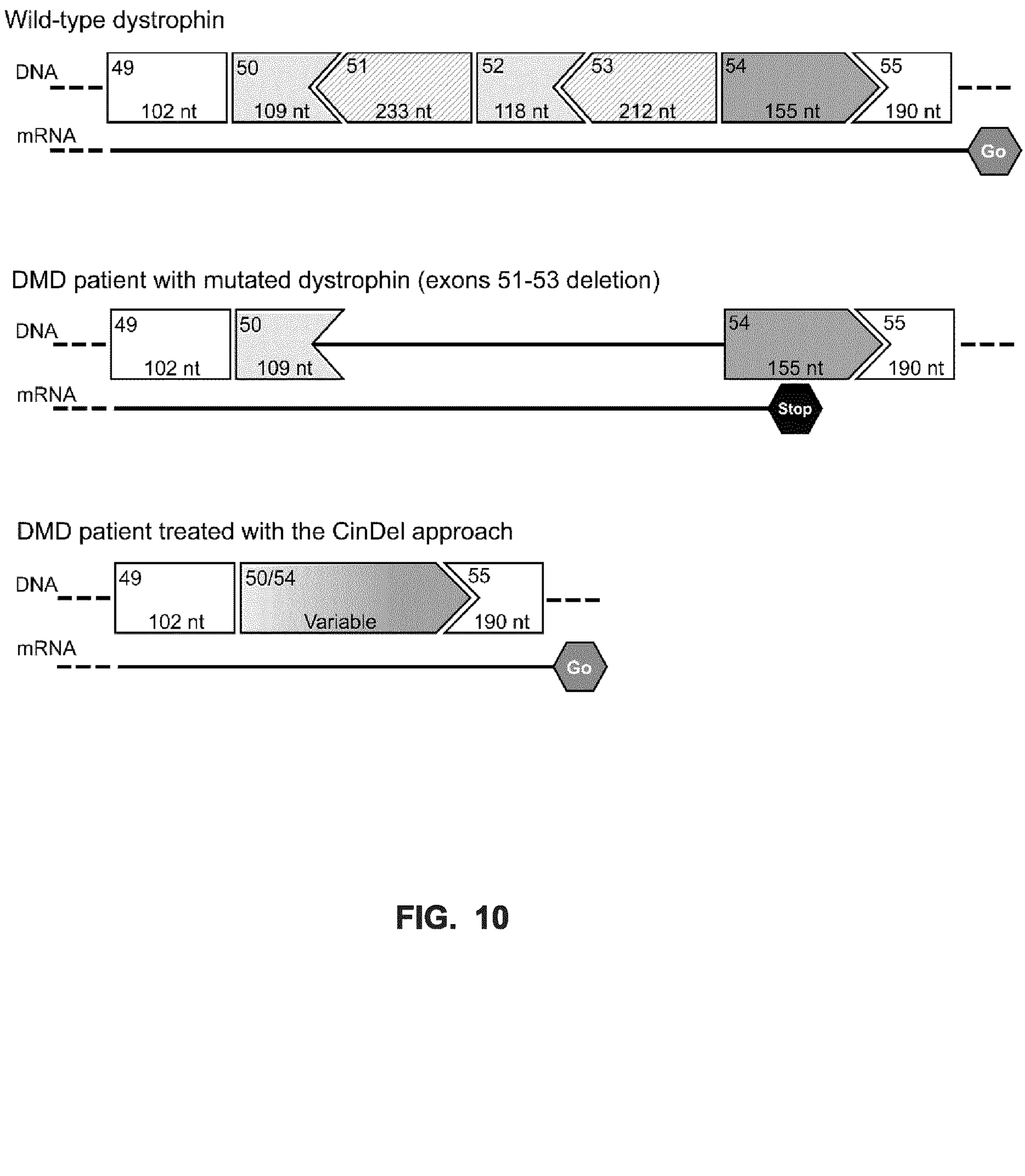
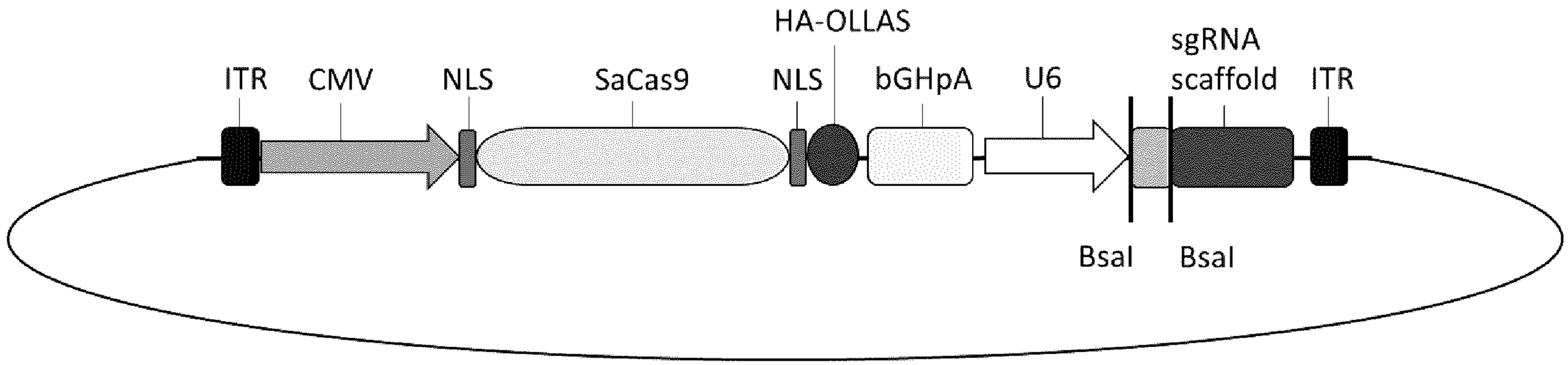

Wild-type dystrophin
DNA
49
102 nt
50
109 nt
51
233 nt
52
118 nt
53
212 nt
54
155 nt
55
190 nt
mRNA
Go
DMD patient with mutated dystrophin (exons 51-53 deletion)
DNA
49
102 nt
50
109 nt
54
155 nt
55
190 nt
mRNA
Stop
DMD patient treated with the CinDel approach
DNA
49
102 nt
50/54
Variable
55
190 nt
mRNA
Go

FIG. 10

a



b

Exon 46

1

5′ **GCTAGAAGAACAA**AAGAATATCTTGTCAGAATTTCAAAGAGATTTAAATGAATTTGTTTTATG 3′
3′ CGATCTTCTTGTTTTCTTATAGAACAGTCTTAAAGTTTCTCTAAATTTACTTAAACAAAATAC 3′

5′ GTTGGAGGAAGCAGATAACATTGCTAGTATCCCACTTGAACCTGGAAAAGAGCAGCAACTAAA 3′
3′ CAACCTCCTTCGTCTATTGTAACGATCATAGGGT**GAACTTGGACCTTTTCTCGTC**GTTGATTT 5′

2

5′ AGAAAAGCTTGAGCAAGTCAAG 3′
3′ TCTTTTCGAACTCGTTCAGTTC 5′

FIG. 11

CA 03037635 2019-03-20

C

Exon 47

3

5′ **TTACTGGTGG**AAGAGTTGCCCCTGCGCCAGGGAATTCTCAAACAATTAAATGAAACTGGAGGA 3′

3′ AATGACCACCTTCTCAACGGGGACGCGGTCCCT**TAAGAG****TTTGTTAATTTACTTTGACCT**CCT 5′

4

5′ CCCGTGCTTGTAAGTGCTCCCATAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAGCTCAAG 3′

3′ GGGCA**CGAACATTCACGAGGGTATTC**GGGTCTTCTCGTTCTATTTGAACTTTTATTCGAGTTC 5′

5

5′ CAGACAAATCTCCAGTGGATAAAG 3′

3′ GTCTGTTTAGAGGTCACCTATTTC 5′

d

Exon 49

5′ GAAACTGAAATAGCAGTTCAAGCTAAACAACCGGATGTGGAAGAGATTTTGTCTAAAGGGCAG 3′

3′ CTTTGACTTTATCGTCAAGTTCGATTTGTTGGCCTACACCTTCTCTAAAACAGATTTCCCGTC 5′

5′ CATTTGTACAAGGAAAAACCAGCCACTCAGCCAGTGAAG 3′

3′ GTAAACATGTTCCTTTTTGGTCGG**TGAGTC****GGTCACTTC** 5′

6

CONTINUATION OF FIG. 11

SUBSTITUTE SHEET (RULE 26)

e **Exon 51**

5′ CTCCTACTCAGACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCA 3′
3′ GAGGATGAGTC**TGACAATGAGACCACTGTGTTGGACACCAATGA**TTCCTTTGACGGTAGAGGT 5′
7 **8**

5′ AACTAGAAATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCTGGCAGATTTCAACCGGGCTT 3′
3′ TTGATCTTTACGGTAGAAGGAACTACAACCTCCATGGACGAGACCGTCTAAAGTTGGCCCGAA 5′

5′ GGACAGAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATCACAGAGGGTGATGG 3′
3′ CCTGTCTTGAATGGCTGACCGAAAGAGACGAACTAGTTCAATATTTTAGTGTCTCCCACTACC 5′

5′ TGGGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAG 3′
3′ ACCCACTGGAACTCCTATAGTTGCTCTACTAGTAGTTCGTCTTC 5′

f **Exon 52**

5′ GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAA 3′
3′ CGTTGTTACGTCCTAAACCTTGTCTCCGCAGGGGTCAACCTTCTTGAGTA**ATGGCGACGGGTT** 5′
9

10
5′ AATTTGAAAAACAAGACCAGCAAT**CAAGAGGCTAGAACAATCATT**ACGGATCGAA 3′
3′ TTAAACTTTTTGTTCTGGTCGTTAGTTCTCCGATCTTGTTAGTAATGCCTAGCTT 5′

CONTINUATION OF FIG. 11

SUBSTITUTE SHEET (RULE 26)

CA 3037635 2019-03-20

SUBSTITUTE SHEET (RULE 26)

g

Exon 53

5′ TTGAAAGAATTCAGAATCAGTGGGATGAAGTACAAGAACAC**CTTCAGAACCGGAGGCAACAGT** 3′

3′ AACTTTCTTAAGTC**TTAGTCACCCTACTTCATGTT**CTTGTGGAAGTCTTGGCCTCCGTTGTCA 5′

11

12 **13**

5′ **TGAATGAAATGTTA**AAGGATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCT 3′

3′ ACTTACTTTACAATTTCCTAAGTTGTGTTACCGACCTTCGATTCCTTCTTCGACTCGTCCAGA 5′

14

5′ TAGGACAGGCCAGA**GCCAAGCTTGAGTCATGGAAG**GAGGGTCCCTATACAGTAGATGCAATCC 3′

3′ ATCCTGTCCGGTCTCGGTTCGAACTCAGTACCTTCCTCCCAGGGATATGTCATCTACGTTAGG 5′

15

5′ AAAAGAAAATCACAGAAACCAAG 3′

3′ TT**TTCTTTTAGTGTCTTTGGTTC** 5′

h

Exon 58

16

5′ **GCCTTCAAG**AGGGAATTGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGA 3′

3′ CGGAAGTTCTCCCTTAACTTTTGATTTCTTGGACATTAGTACTCATGAGAA**CTCTGACATGCT** 5′

17

5′ ATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAGAG 3′

3′ **TATAAAGAC**TGTCTCGTCGGAAACCTTCCTGATCTCTTTGAGAT**GGTCCTCGGGTCTC** 5′

18

CONTINUATION OF FIG. 11

Part of intron 22

i

```
5′ ...CTGTGATGTGAGGACATATAAAGACTAATTTTTTTGTTGATTCTAAAAATCCCATGTTGTATACTTATTCTT 3′
3′ ...GACACTACACTCCTGTATATTTCTGATTAAAAAAACAACTAAGATTTTTAGGGTACAACATATGAATAAGAA 5′

                                                                21
5′ TTTAAATCTGAAAATATATTAATCATATATTGCCTAAATGTCTTAATAATGTTTCACTGTAGGTAAGTTAAA 3′
3′ AAATTTAGACTTTTATATAATTAGTATATAACGGATTTACAGAATTATTACAAAGTGACATCCATTCAATTT 5′


5′ ATGTATCACATATATAATAAACATAGTTATTAATGCATAGATATTCAGTAAAATTATGACTTCTAAATTTCT 3′
3′ TACATAGTGTATATATTATTTGTATCAATAATTACGTATCTATAAGTCATTTTAATACTGAAGATTTAAAGA 5′


5′ GTCTAAATATAATATGCCCTGTAATATAATAGAAATTATTCATAAGAATACATATATATTGCTTTATCAGAT 3′
3′ CAGATTTATATTATACGGGACATTATATTATCTTTAATAAGTATTCTTATGTATATATAACGAAATAGTCTA 5′
                    22

5′ ATTCTACTTTGTTTAGATCTCTAAATTACATAAACTTTTATTTACCTTCTTCTTGATATGAATGAAACTCAT 3′
3′ TAAGATGAAACAAATCTAGAGATTTAATGTATTTGAAAATAAATGGAAGAAGAACTATACTTACTTTGAGTA 5′
                                                  23
                                               19
5′ CAAATATGCGTGTTAGTGTAAATGAACTTCTATTTAATTTTGAGGCTCTGCAAAGTTCTTTGAAAGA 3′
3′ GTTTATACGCACAATCACATTTACTTGAAGATAAATTAAAACTCCGAGACGTTTCAAGAAACTTTCT 5′
```

Part of intron 23

j

```
                                        26   20
5′ AATAAACTTCGAAAATTTCAGGTAAGCCGAGGTTTGGCCTTTAAACTATATTTTTTCACATAGCAATTAATTGGA 3′
3′ TTATTTGAAGCTTTTAAAGTCCATTCGGCTCCAAACCGGAAATTTGATATAAAAAAGTGTATCGTTAATTAACCT 5′
                                   24          25
```

CONTINUATION OF FIG. 11

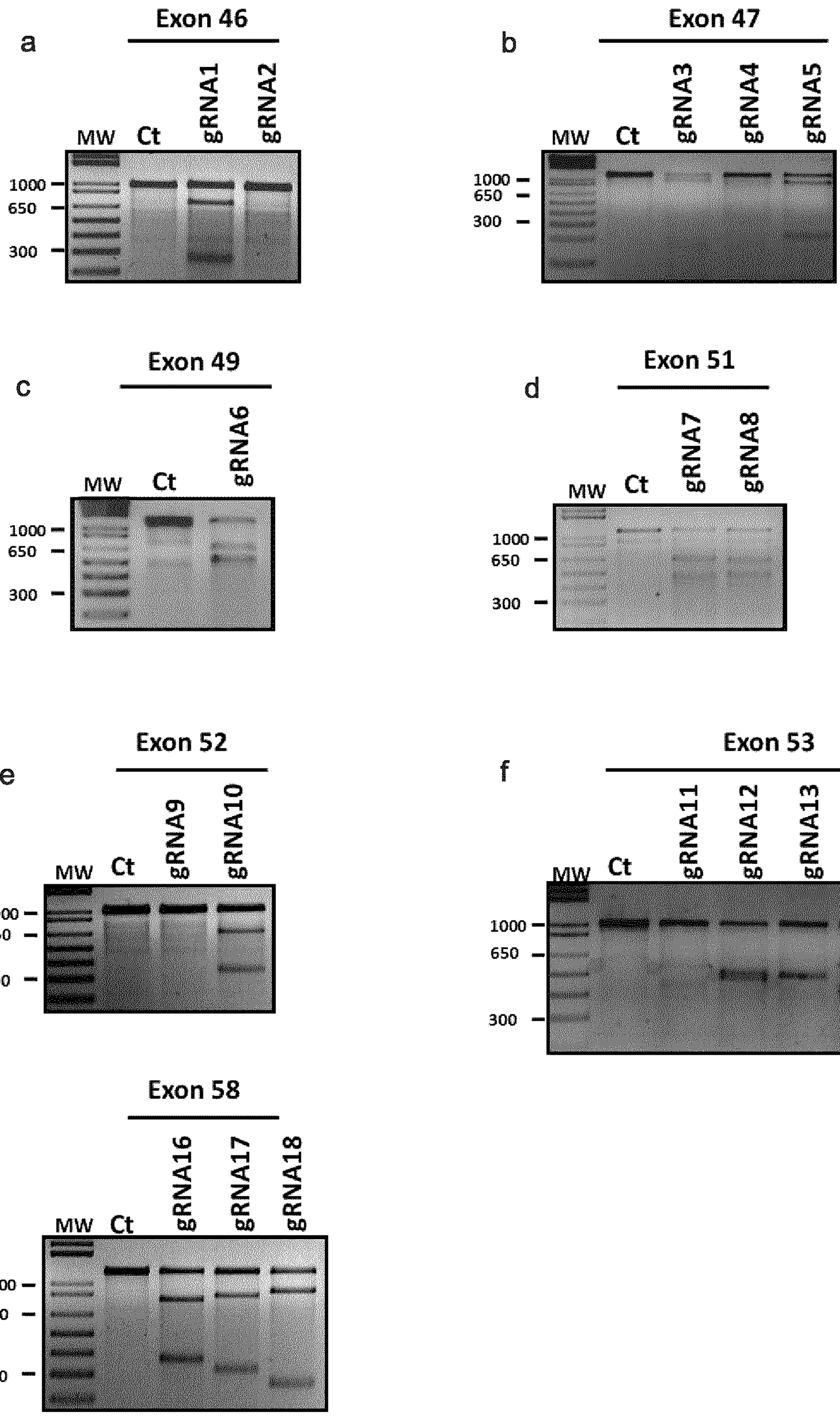

a
Exon 46
MW
Ct
gRNA1
gRNA2
1000
650
300
b
Exon 47
MW
Ct
gRNA3
gRNA4
gRNA5
1000
650
300
c
Exon 49
MW
Ct
gRNA6
1000
650
300
d
Exon 51
MW
Ct
gRNA7
gRNA8
1000
650
300
e
Exon 52
MW
Ct
gRNA9
gRNA10
1000
650
300
f
Exon 53
MW
Ct
gRNA11
gRNA12
gRNA13
gRNA14
gRNA11
1000
650
300
g
Exon 58
MW
Ct
A16
A17
A18
1000
650
300

FIG. 12

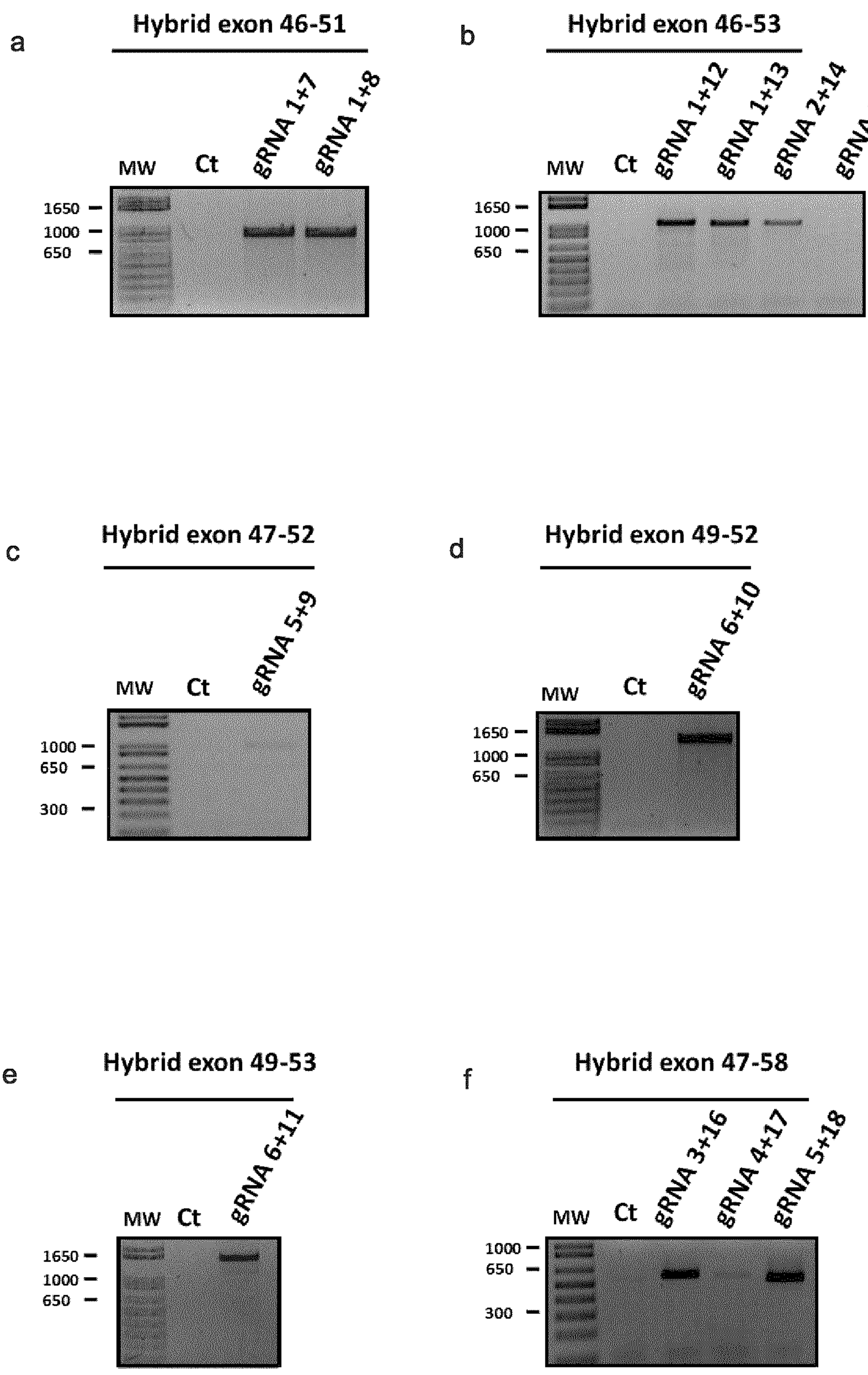

a
Hybrid exon 46-51
MW
Ct
gRNA 1+7
gRNA 1+8
1650
1000
650
b
Hybrid exon 46-53
MW
Ct
gRNA 1+12
gRNA 1+13
gRNA 2+14
gRNA 2+15
1650
1000
650
c
Hybrid exon 47-52
MW
Ct
gRNA 5+9
1000
650
300
d
Hybrid exon 49-52
MW
Ct
gRNA 6+10
1650
1000
650
e
Hybrid exon 49-53
MW
Ct
gRNA 6+11
1650
1000
650
f
Hybrid exon 47-58
MW
Ct
gRNA 3+16
gRNA 4+17
gRNA 5+18
1000
650
300

FIG. 13

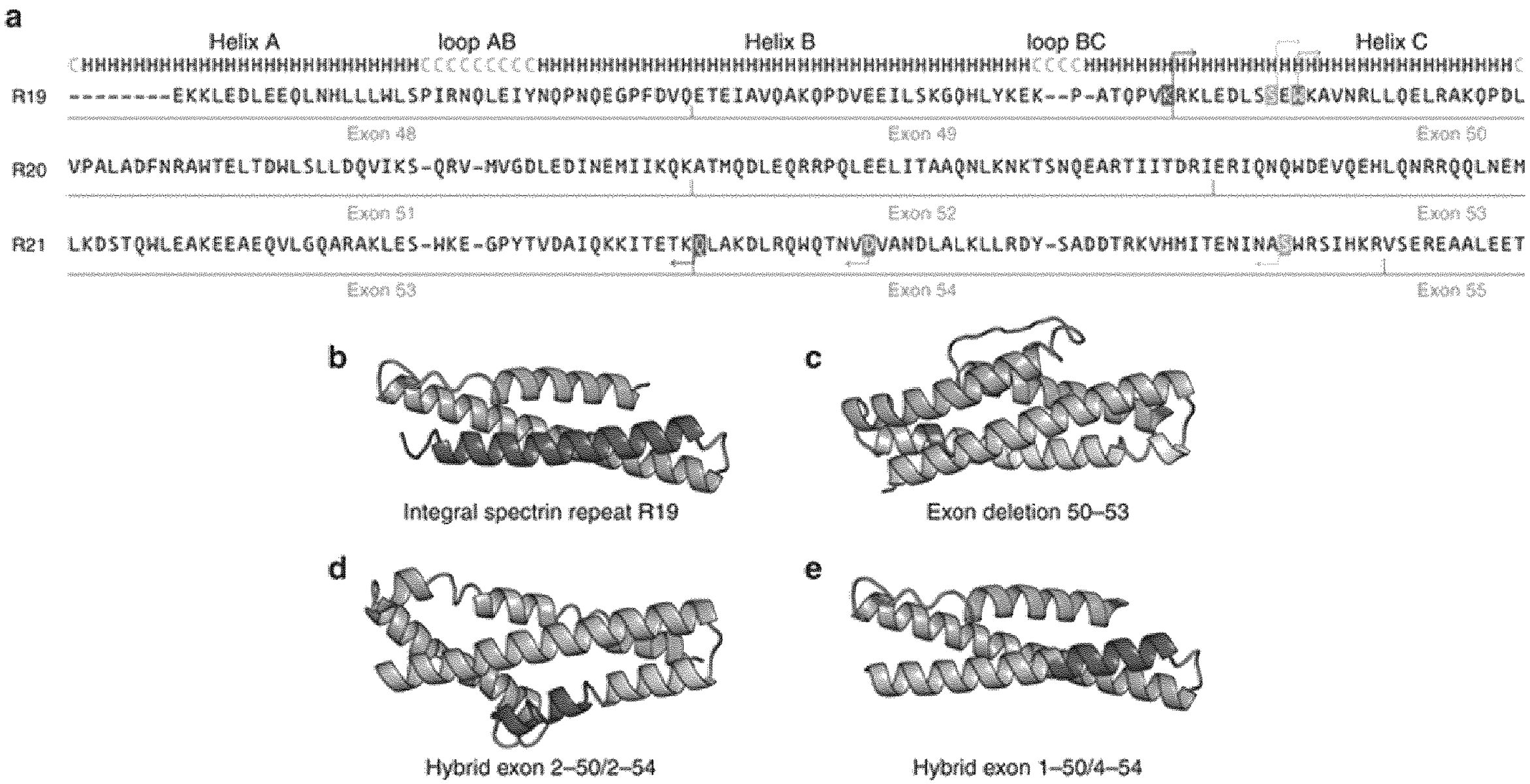

a
Helix A
loop AB
Helix B
loop BC
Helix C
R19
R20
R21
Exon 48
Exon 49
Exon 50
Exon 51
Exon 52
Exon 53
Exon 53
Exon 54
Exon 55
b
Integral spectrin repeat R19
c
Exon deletion 50–53
d
Hybrid exon 2–50/2–54
e
Hybrid exon 1–50/4–54

FIG. 14

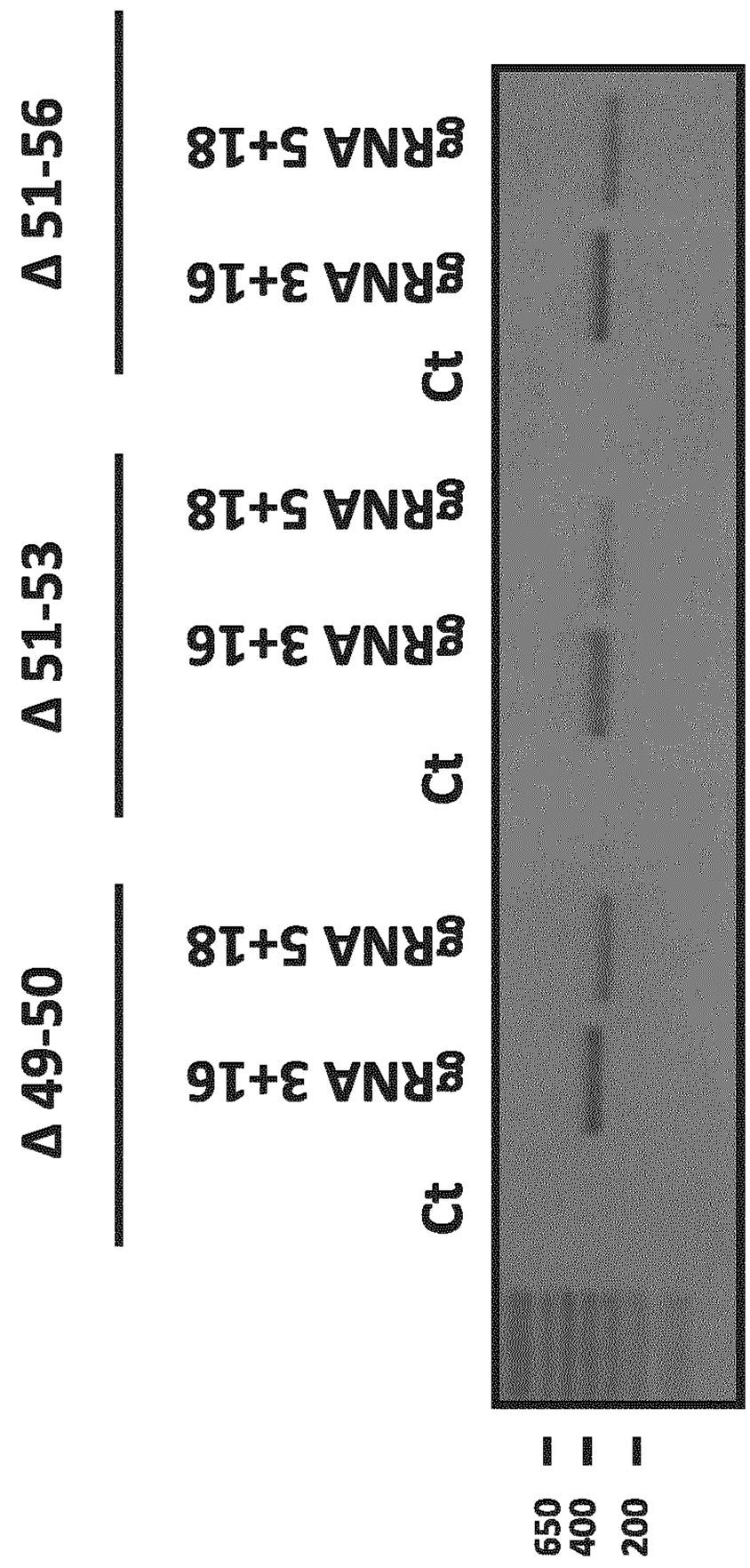

Δ 51-56
gRNA 5+18
gRNA 3+16
Ct
Δ 51-53
gRNA 5+18
gRNA 3+16
Ct
Δ 49-50
gRNA 5+18
gRNA 3+16
Ct
650
400
200

FIG. 15

A) Sa-gRNAs cutting site localization in spectrin-like repeats

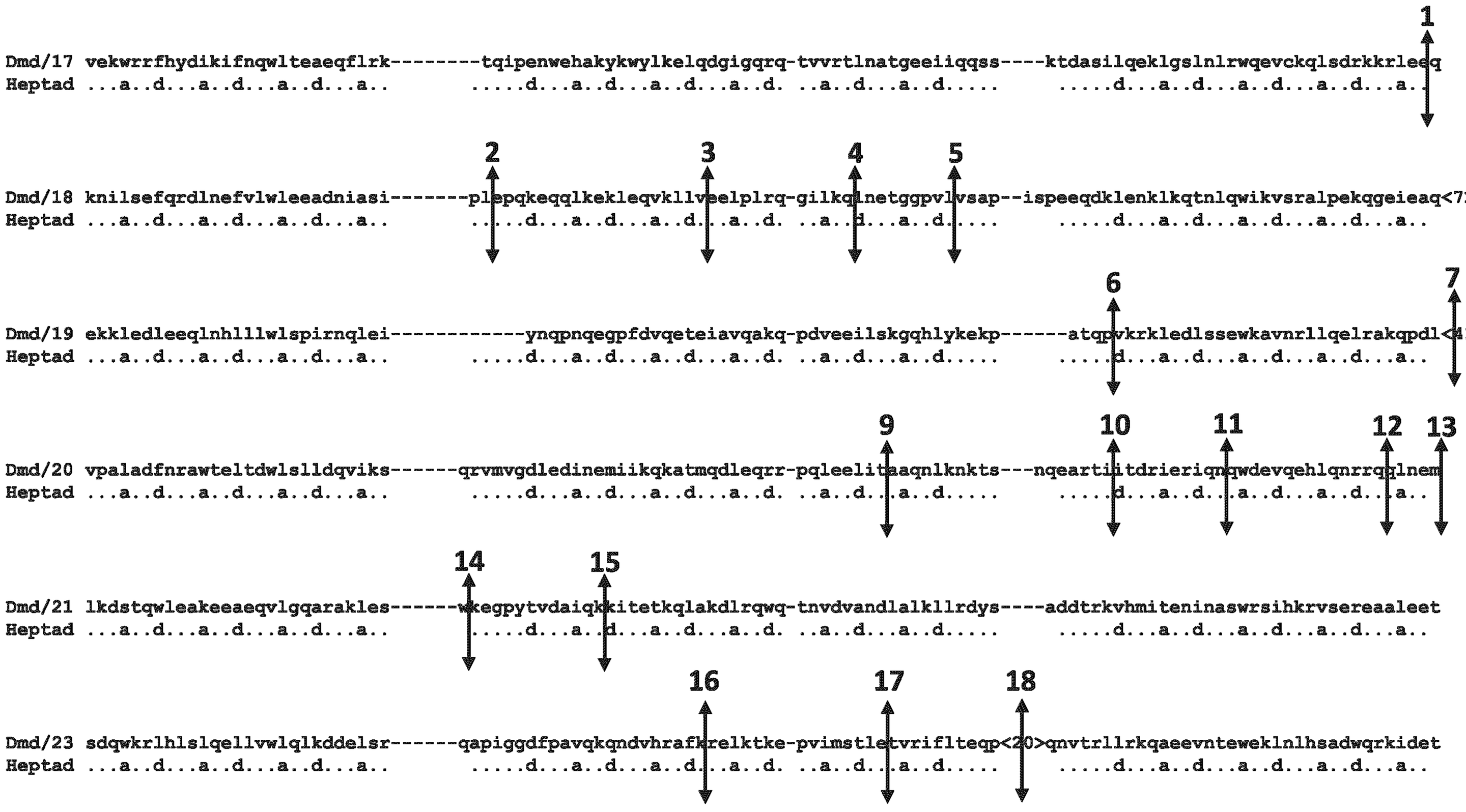
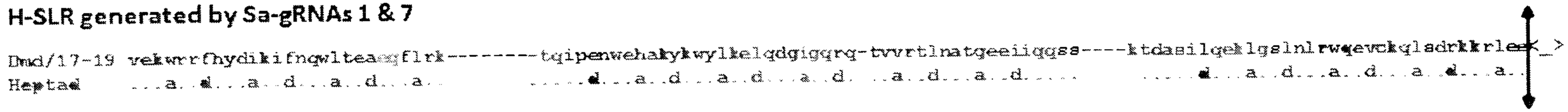


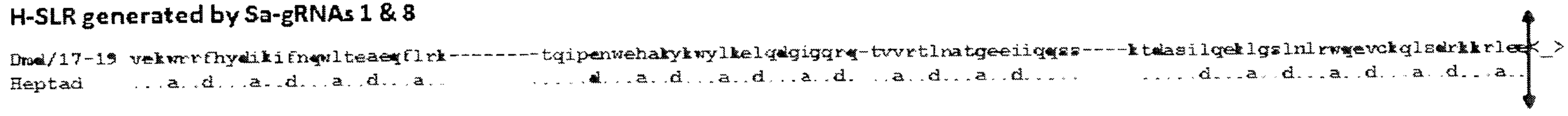
FIG. 16

B) Hybrid spectrin-like repeat generated from combination of Sa-gRNAs

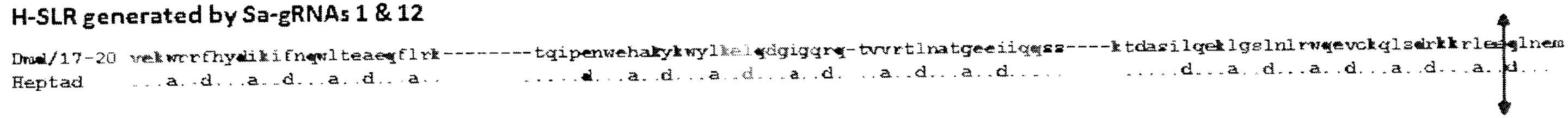
H-SLR generated by Sa-gRNAs 1 & 7



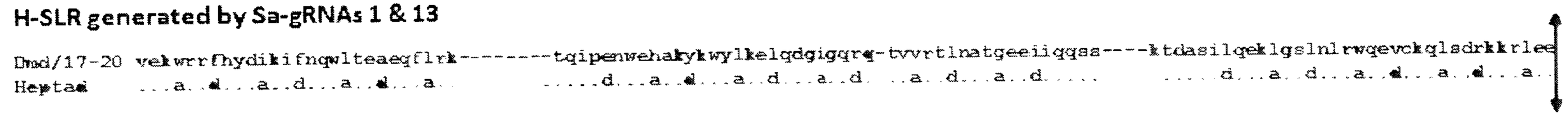
H-SLR generated by Sa-gRNAs 1 & 8



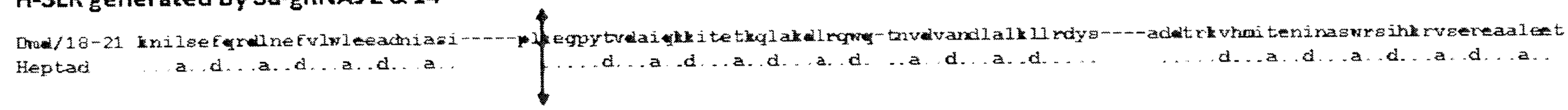
H-SLR generated by Sa-gRNAs 1 & 12



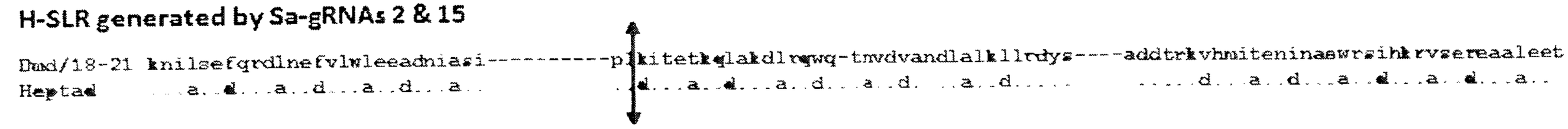
H-SLR generated by Sa-gRNAs 1 & 13



H-SLR generated by Sa-gRNAs 2 & 14



H-SLR generated by Sa-gRNAs 2 & 15

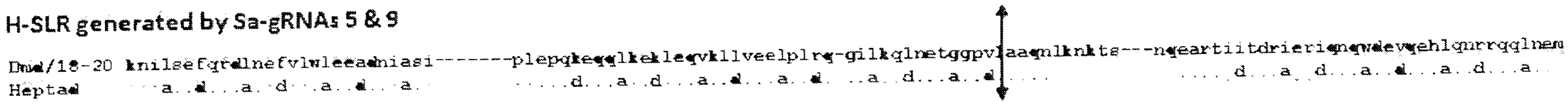


CONTINUATION OF FIG. 16

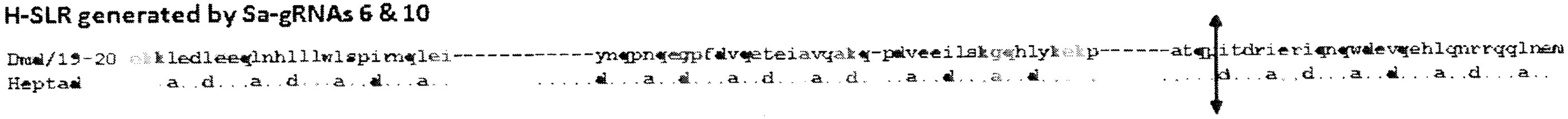
CA 03037635 2019-03-20

SUBSTITUTE SHEET (RULE 26)

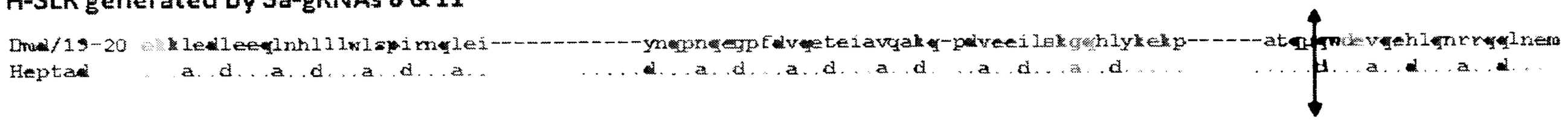
C) H-SLR generated by Sa-gRNAs 5 & 9



H-SLR generated by Sa-gRNAs 6 & 10



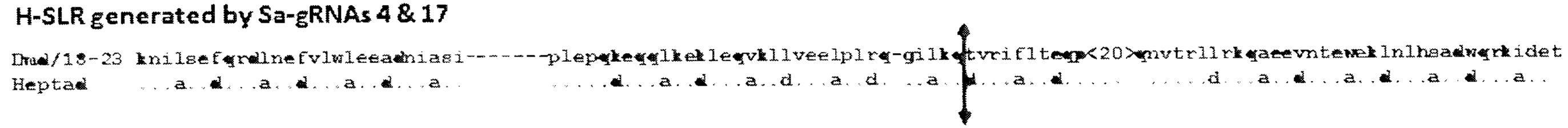
H-SLR generated by Sa-gRNAs 6 & 11



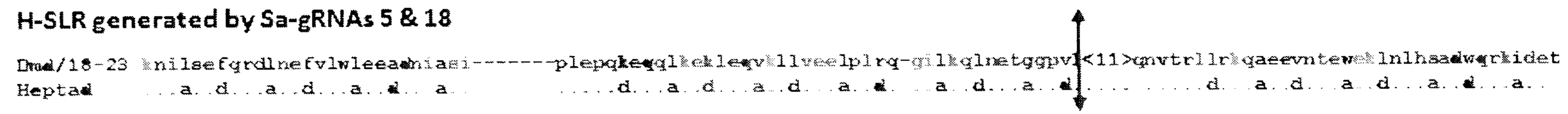
H-SLR generated by Sa-gRNAs 3 & 16



H-SLR generated by Sa-gRNAs 4 & 17



H-SLR generated by Sa-gRNAs 5 & 18



CONTINUATION OF FIG. 16

Hybrid exons 46-51

Sa-gRNAs 1+7

GCTAGAAGAA/GTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTCC
TTGATGTTGGAGGTACCTGCTCTGGCAGATTTCAACCGGGCTTGGACAGAACTTACCGACTGGCTTTCTCTGCTTGATC
AAGTTATAAAATCACAGAGGGTGATGGTGGGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAG

Sa-gRNAs 1+8

GCTAGAAGAA/ACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTCCTTGATGTTGGAG
GTACCTGCTCTGGCAGATTTCAACCGGGCTTGGACAGAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATC
ACAGAGGGTGATGGTGGGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAG

Hybrid exons 46-53

Sa-gRNAs 1+12

GCTAGAAGAA/CAGTTGAATGAAATGTTAAAGGATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCTTA
GGACAGGCCAGAGCCAAGCTTGAGTCATGGAAGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATCACAGAA
ACCAAG

Sa-gRNAs 1+13

GCTAGAAGAA/TTAAAGGATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCTTAGGACAGGCCAGAGC
CAAGCTTGAGTCATGGAAGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATCACAGAAACCAAG

Hybrid exon 49-52

Sa-gRNAs 6+10

GAAACTGAAATAGCAGTTCAAGCTAAACAACCGGATGTGGAAGAGATTTTGTCTAAAGGGCAGCATTTGTACAAGGAAA
AACCAGCCACTCAGCCA/ATTACGGATCGAA

Hybrid exon 49-53

Sa-gRNAs 6+11

GAAACTGAAATAGCAGTTCAAGCTAAACAACCGGATGTGGAAGAGATTTTGTCTAAAGGGCAGCATTTGTACAAGGAAA
AACCAGCCACTCAGCCA/CAGTGGGATGAAGTACAAGAACACCTTCAGAACCGGAGGCAACAGTTGAATGAAATGTTAA
AGGATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCTTAGGACAGGCCAGAGCCAAGCTTGAGTCAT
GGAAGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATCACAGAAACCAAG

Hybrid exons 47-58

Sa-gRNAs 3+16

TTACTGG/AGAGGGAATTGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA
GCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAGAG

Sa-gRNAs 5+18

TTACTGGTGGAAGAGTTGCCCCTGCGCCAGGGAATTCTCAAACAATTAAATGAAACTGGAGGACCCGTGCT/GGAGCCC
AGAG

FIG. 17

a

Combinations of gRNA	Hybrid exon	Number of sequenced clones containing the expected hybrid exon / Number of sequenced clones	% of expected junction
Sa gRNA 1 + 7	46-51	2/4	50
Sa gRNA 1 + 8	46-51	3/5	60
Sa gRNA 1 + 12	46-53	1/5	20
Sa gRNA 1 + 13	46-53	1/4	25
Sa gRNA 6 + 10	49-52	3/4	75
Sa gRNA 6 + 11	49-53	1/4	25
Sa gRNA 3 + 16	47-58	12/16	75
Sa gRNA 5 + 18	47-58	7/15	46,5

b

DMD mutation	Δ 49-50	Δ 51-53	Δ 51-56
Sa gRNA 3+16	7/10	4/10	8/9
Sa gRNA 5+18	7/10	5/10	7/11

FIG. 18

```
   1 ATGCTTTGGTGGGAAGAAGTAGAGGACTGTTatgaaagagaagatgttcaaaagaaaaca    60
   1 ATGCTTTGGTGGGAAGAAGTAGAGGACTGTTATGAAAGAGAAGATGTTCAAAAGAAAACA    60
   1 -M--L--W--W--E--E--V--E--D--C--Y--E--R--E--D--V--Q--K--K--T-    20

  61 ttcacaaaatgggtaaatgcacaattttctaagTTTGGGAAGCAGCATATTGAGAACCTC   120
  61 TTCACAAAATGGGTAAATGCACAATTTTCTAAGTTTGGGAAGCAGCATATTGAGAACCTC   120
  21 -F--T--K--W--V--N--A--Q--F--S--K--F--G--K--Q--H--I--E--N--L-    40

 121 TTCAGTGACCTACAGGATGGGAGGCGCCTCCTAGACCTCCTCGAAGGCCTGACAGGGCAA   180
 121 TTCAGTGACCTACAGGATGGGAGGCGCCTCCTAGACCTCCTCGAAGGCCTGACAGGGCAA   180
  41 -F--S--D--L--Q--D--G--R--R--L--L--D--L--L--E--G--L--T--G--Q-    60

 181 AAACTGccaaaagaaaaaggatccacaagagttcatgccctgaacaatgtcaacaaggca   240
 181 AAACTGCCAAAAGAAAAAGGATCCACAAGAGTTCATGCCCTGAACAATGTCAACAAGGCA   240
  61 -K--L--P--K--E--K--G--S--T--R--V--H--A--L--N--N--V--N--K--A-    80

 241 ctgcgggttttgcagaacaataatGTTGATTTAGTGAATATTGGAAGTACTGACATCGTA   300
 241 CTGCGGGTTTTGCAGAACAATAATGTTGATTTAGTGAATATTGGAAGTACTGACATCGTA   300
  81 -L--R--V--L--Q--N--N--N--V--D--L--V--N--I--G--S--T--D--I--V-   100

 301 GATGGAAATCATAAACTGACTCTTGGTTTGATTTGGAATATAATCCTCCACTGGCAGgtc   360
 301 GATGGAAATCATAAACTGACTCTTGGTTTGATTTGGAATATAATCCTCCACTGGCAGGTC   360
 101 -D--G--N--H--K--L--T--L--G--L--I--W--N--I--I--L--H--W--Q--V-   120

 361 aaaaatgtaatgaaaaatatcatggctggattgcaacaaaccaacagtgaaaagattctc   420
 361 AAAAATGTAATGAAAAATATCATGGCTGGATTGCAACAAACCAACAGTGAAAAGATTCTC   420
 121 -K--N--V--M--K--N--I--M--A--G--L--Q--Q--T--N--S--E--K--I--L-   140

 421 ctgagctgggtccgacaatcaactcgtaattatccacaggttaatgtaatcaacttcacc   480
 421 CTGAGCTGGGTCCGACAATCAACTCGTAATTATCCACAGGTTAATGTAATCAACTTCACC   480
 141 -L--S--W--V--R--Q--S--T--R--N--Y--P--Q--V--N--V--I--N--F--T-   160

 481 accagctggtctgatggcctggctttgaatgctctcatccatagtcatagGCCAGACCTA   540
 481 ACCAGCTGGTCTGATGGCCTGGCTTTGAATGCTCTCATCCATAGTCATAGGCCAGACCTA   540
 161 -T--S--W--S--D--G--L--A--L--N--A--L--I--H--S--H--R--P--D--L-   180

 541 TTTGACTGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACAACGACTGGAACATGCATTC   600
 541 TTTGACTGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACAACGACTGGAACATGCATTC   600
 181 -F--D--W--N--S--V--V--C--Q--Q--S--A--T--Q--R--L--E--H--A--F-   200

 601 AACATCGCCAGATATCAATTAGGCATAGAGAAACTACTCGATCCTGAAGatgttgatacc   660
 601 AACATCGCCAGATATCAATTAGGCATAGAGAAACTACTCGATCCTGAAGATGTTGATACC   660
 201 -N--I--A--R--Y--Q--L--G--I--E--K--L--L--D--P--E--D--V--D--T-   220

 661 acctatccagataagaagtccatcttaatgtacatcacatcactcttccaagttttgcct   720
 661 ACCTATCCAGATAAGAAGTCCATCTTAATGTACATCACATCACTCTTCCAAGTTTTGCCT   720
 221 -T--Y--P--D--K--K--S--I--L--M--Y--I--T--S--L--F--Q--V--L--P-   240

 721 caacaagtgagcattgaagccatccaggaagtggaaatgttgccaaggccacctaaagtg   780
 721 CAACAAGTGAGCATTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCCACCTAAAGTG   780
 241 -Q--Q--V--S--I--E--A--I--Q--E--V--E--M--L--P--R--P--P--K--V-   260

 781 actaaagaagaacattttcagttacatcatcaaatgcactattctcaacagATCACGGTC   840
 781 ACTAAAGAAGAACATTTTCAGTTACATCATCAAATGCACTATTCTCAACAGATCACGGTC   840
 261 -T--K--E--E--H--F--Q--L--H--H--Q--M--H--Y--S--Q--Q--I--T--V-   280

 841 AGTCTAGCACAGGGATATGAGAGAACTTCTTCCCCTAAGCCTCGATTCAAGAGCTATGCC   900
 841 AGTCTAGCACAGGGATATGAGAGAACTTCTTCCCCTAAGCCTCGATTCAAGAGCTATGCC   900
 281 -S--L--A--Q--G--Y--E--R--T--S--S--P--K--P--R--F--K--S--Y--A-   300

 901 TACACACAGGCTGCTTATGTCACCACCTCTGACCCTACACGGAGCCCATTTCCTTCACAG   960
 901 TACACACAGGCTGCTTATGTCACCACCTCTGACCCTACACGGAGCCCATTTCCTTCACAG   960
 301 -Y--T--Q--A--A--Y--V--T--T--S--D--P--T--R--S--P--F--P--S--Q-   320

 961 catttggaagctcctgaagacaagtcatttggcagttcattgatggagagtgaagtaaac  1020
```

FIG. 19A

```
 961 CATTTGGAAGCTCCTGAAGACAAGTCATTTGGCAGTTCATTGATGGAGAGTGAAGTAAAC  1020
 321 -H--L--E--A--P--E--D--K--S--F--G--S--S--L--M--E--S--E--V--N-   340

1021 ctggaccgttatcaaacagctttagaagaagtattatcgtggcttctttctgctgaggac  1080
1021 CTGGACCGTTATCAAACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTTCTGCTGAGGAC  1080
 341 -L--D--R--Y--Q--T--A--L--E--E--V--L--S--W--L--L--S--A--E--D-   360

1081 acattgcaagcacaaggagagatttctaatgatgtggaagtggtgaaagaccagtttcat  1140
1081 ACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAAGACCAGTTTCAT  1140
 361 -T--L--Q--A--Q--G--E--I--S--N--D--V--E--V--V--K--D--Q--F--H-   380

1141 actcatgagGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGGTTGGTAATATTCTA  1200
1141 ACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGGTTGGTAATATTCTA  1200
 381 -T--H--E--G--Y--M--M--D--L--T--A--H--Q--G--R--V--G--N--I--L-   400

1201 CAATTGGGAAGTAAGCTGATTGGAACAGGAAAATTATCAGAAGATGAAGAAACTGAAGTA  1260
1201 CAATTGGGAAGTAAGCTGATTGGAACAGGAAAATTATCAGAAGATGAAGAAACTGAAGTA  1260
 401 -Q--L--G--S--K--L--I--G--T--G--K--L--S--E--D--E--E--T--E--V-   420

1261 CAAGAGCAGATGAATCTCCTAAATTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAA  1320
1261 CAAGAGCAGATGAATCTCCTAAATTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAA  1320
 421 -Q--E--Q--M--N--L--L--N--S--R--W--E--C--L--R--V--A--S--M--E-   440

1321 AAACAAAGCAAtttacatagagttttaatggatctccagaatcagaaactgaaagagttg  1380
1321 AAACAAAGCAATTTACATAGAGTTTTAATGGATCTCCAGAATCAGAAACTGAAAGAGTTG  1380
 441 -K--Q--S--N--L--H--R--V--L--M--D--L--Q--N--Q--K--L--K--E--L-   460

1381 aatgactggctaacaaaaacagaagaaagaacaaggaaaatggaggaagagcctcttgga  1440
1381 AATGACTGGCTAACAAAAACAGAAGAAAGAACAAGGAAAATGGAGGAAGAGCCTCTTGGA  1440
 461 -N--D--W--L--T--K--T--E--E--R--T--R--K--M--E--E--E--P--L--G-   480

1441 cctgatcttgaagacctaaaacgccaagtacaacaacataagGTGCTTCAAGAAGATCTA  1500
1441 CCTGATCTTGAAGACCTAAAACGCCAAGTACAACAACATAAGGTGCTTCAAGAAGATCTA  1500
 481 -P--D--L--E--D--L--K--R--Q--V--Q--Q--H--K--V--L--Q--E--D--L-   500

1501 GAACAAGAACAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTGGTAGTTGATGAATCT  1560
1501 GAACAAGAACAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTGGTAGTTGATGAATCT  1560
 501 -E--Q--E--Q--V--R--V--N--S--L--T--H--M--V--V--V--V--D--E--S-   520

1561 AGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGgtattgggagatcgatgg  1620
1561 AGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGGAGATCGATGG  1620
 521 -S--G--D--H--A--T--A--A--L--E--E--Q--L--K--V--L--G--D--R--W-   540

1621 gcaaacatctgtagatggacagaagaccgctgggttcttttacaagacatccttctcaaa  1680
1621 GCAAACATCTGTAGATGGACAGAAGACCGCTGGGTTCTTTTACAAGACATCCTTCTCAAA  1680
 541 -A--N--I--C--R--W--T--E--D--R--W--V--L--L--Q--D--I--L--L--K-   560

1681 tggcaacgtcttactgaagaacagTGCCTTTTTAGTGCATGGCTTTCAGAAAAAGAAGAT  1740
1681 TGGCAACGTCTTACTGAAGAACAGTGCCTTTTTAGTGCATGGCTTTCAGAAAAAGAAGAT  1740
 561 -W--Q--R--L--T--E--E--Q--C--L--F--S--A--W--L--S--E--K--E--D-   580

1741 GCAGTGAACAAGATTCACACAACTGGCTTTAAAGATCAAAATGAAATGTTATCAAGTCTT  1800
1741 GCAGTGAACAAGATTCACACAACTGGCTTTAAAGATCAAAATGAAATGTTATCAAGTCTT  1800
 581 -A--V--N--K--I--H--T--T--G--F--K--D--Q--N--E--M--L--S--S--L-   600

1801 CAAAAACTGGCCgttttaaaagcggatctagaaaagaaaaagcaatccatgggcaaactg  1860
1801 CAAAAACTGGCCGTTTTAAAAGCGGATCTAGAAAAGAAAAAGCAATCCATGGGCAAACTG  1860
 601 -Q--K--L--A--V--L--K--A--D--L--E--K--K--K--Q--S--M--G--K--L-   620

1861 tattcactcaaacaagatcttctttcaacactgaagaataagtcagtgacccagaagacg  1920
1861 TATTCACTCAAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGAAGACG  1920
 621 -Y--S--L--K--Q--D--L--L--S--T--L--K--N--K--S--V--T--Q--K--T-   640

1921 gaagcatggctggataactttgcccggtgttgggataatttagtccaaaaacttgaaaag  1980
1921 GAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAAACTTGAAAAG  1980
 641 -E--A--W--L--D--N--F--A--R--C--W--D--N--L--V--Q--K--L--E--K-   660

1981 agtacagcacagATTTCACAGGCTGTCACCACCACTCAGCCATCACTAACACAGACAACT  2040
```

FIG. 19B

```
1981 AGTACAGCACAGATTTCACAGGCTGTCACCACCACTCAGCCATCACTAACACAGACAACT  2040
 661 -S--T--A--Q--I--S--Q--A--V--T--T--T--Q--P--S--L--T--Q--T--T-   680

2041 GTAATGGAAACAGTAACTACGGTGACCACAAGGGAACAGATCCTGGTAAAGCATGCTCAA  2100
2041 GTAATGGAAACAGTAACTACGGTGACCACAAGGGAACAGATCCTGGTAAAGCATGCTCAA  2100
 681 -V--M--E--T--V--T--T--V--T--T--R--E--Q--I--L--V--K--H--A--Q-   700

2101 GAGGAACTTCCACCACCACCTCCCCAAAAGAAGAGGCAGATTACTGTGGATTCTGAAATT  2160
2101 GAGGAACTTCCACCACCACCTCCCCAAAAGAAGAGGCAGATTACTGTGGATTCTGAAATT  2160
 701 -E--E--L--P--P--P--P--P--Q--K--K--R--Q--I--T--V--D--S--E--I-   720

2161 AGGAAAAGgttggatgttgatataactgaacttcacagctggattactcgctcagaagct  2220
2161 AGGAAAAGGTTGGATGTTGATATAACTGAACTTCACAGCTGGATTACTCGCTCAGAAGCT  2220
 721 -R--K--R--L--D--V--D--I--T--E--L--H--S--W--I--T--R--S--E--A-   740

2221 gtgttgcagagtcctgaatttgcaatctttcggaaggaaggcaacttctcagacttaaaa  2280
2221 GTGTTGCAGAGTCCTGAATTTGCAATCTTTCGGAAGGAAGGCAACTTCTCAGACTTAAAA  2280
 741 -V--L--Q--S--P--E--F--A--I--F--R--K--E--G--N--F--S--D--L--K-   760

2281 gaaaaagtcaatGCCATAGAGCGAGAAAAAGCTGAGAAGTTCAGAAAACTGCAAGATGCC  2340
2281 GAAAAAGTCAATGCCATAGAGCGAGAAAAAGCTGAGAAGTTCAGAAAACTGCAAGATGCC  2340
 761 -E--K--V--N--A--I--E--R--E--K--A--E--K--F--R--K--L--Q--D--A-   780

2341 AGCAGATCAGCTCAGGCCCTGGTGGAACAGATGGTGAATGagggtgttaatgcagatagc  2400
2341 AGCAGATCAGCTCAGGCCCTGGTGGAACAGATGGTGAATGAGGGTGTTAATGCAGATAGC  2400
 781 -S--R--S--A--Q--A--L--V--E--Q--M--V--N--E--G--V--N--A--D--S-   800

2401 atcaaacaagcctcagaacaactgaacagccggtggatcgaattctgccagttgctaagt  2460
2401 ATCAAACAAGCCTCAGAACAACTGAACAGCCGGTGGATCGAATTCTGCCAGTTGCTAAGT  2460
 801 -I--K--Q--A--S--E--Q--L--N--S--R--W--I--E--F--C--Q--L--L--S-   820

2461 gagagacttaactggctggagtatcagaacaacatcatcgctttctataatcagctacaa  2520
2461 GAGAGACTTAACTGGCTGGAGTATCAGAACAACATCATCGCTTTCTATAATCAGCTACAA  2520
 821 -E--R--L--N--W--L--E--Y--Q--N--N--I--I--A--F--Y--N--Q--L--Q-   840

2521 caattggagcagatgacaactactgctgaaaactggttgaaaatccaacccaccacccca  2580
2521 CAATTGGAGCAGATGACAACTACTGCTGAAAACTGGTTGAAAATCCAACCCACCACCCCA  2580
 841 -Q--L--E--Q--M--T--T--T--A--E--N--W--L--K--I--Q--P--T--T--P-   860

2581 tcagagccaacagcaattaaaagtcagttaaaaatttgtaagGATGAAGTCAACCGGCTA  2640
2581 TCAGAGCCAACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGGATGAAGTCAACCGGCTA  2640
 861 -S--E--P--T--A--I--K--S--Q--L--K--I--C--K--D--E--V--N--R--L-   880

2641 TCAGATCTTCAACCTCAAATTGAACGATTAAAAATTCAAAGCATAGCCCTGAAAGAGAAA  2700
2641 TCAGATCTTCAACCTCAAATTGAACGATTAAAAATTCAAAGCATAGCCCTGAAAGAGAAA  2700
 881 -S--D--L--Q--P--Q--I--E--R--L--K--I--Q--S--I--A--L--K--E--K-   900

2701 GGACAAGGACCCATGTTCCTGGATGCAGACTTTGTGGCCTTTACAAATCATTTTAAGCAA  2760
2701 GGACAAGGACCCATGTTCCTGGATGCAGACTTTGTGGCCTTTACAAATCATTTTAAGCAA  2760
 901 -G--Q--G--P--M--F--L--D--A--D--F--V--A--F--T--N--H--F--K--Q-   920

2761 GTCTTTTCTGATGTGCAGGCCAGAGAGAAAGAGCTACAGACAAttttgacactttgcca  2820
2761 GTCTTTTCTGATGTGCAGGCCAGAGAGAAAGAGCTACAGACAATTTTTGACACTTTGCCA  2820
 921 -V--F--S--D--V--Q--A--R--E--K--E--L--Q--T--I--F--D--T--L--P-   940

2821 ccaatgcgctatcaggagaccatgagtgccatcaggacatgggtccagcagtcagaaacc  2880
2821 CCAATGCGCTATCAGGAGACCATGAGTGCCATCAGGACATGGGTCCAGCAGTCAGAAACC  2880
 941 -P--M--R--Y--Q--E--T--M--S--A--I--R--T--W--V--Q--Q--S--E--T-   960

2881 aaactctccatacctcaacttagtgtcaccgactatgaaatcatggagcagagactcggg  2940
2881 AAACTCTCCATACCTCAACTTAGTGTCACCGACTATGAAATCATGGAGCAGAGACTCGGG  2940
 961 -K--L--S--I--P--Q--L--S--V--T--D--Y--E--I--M--E--Q--R--L--G-   980

2941 gaattgcagGCTTTACAAAGTTCTCTGCAAGAGCAACAAAGTGGCCTATACTATCTCAGC  3000
2941 GAATTGCAGGCTTTACAAAGTTCTCTGCAAGAGCAACAAAGTGGCCTATACTATCTCAGC  3000
 981 -E--L--Q--A--L--Q--S--S--L--Q--E--Q--Q--S--G--L--Y--Y--L--S-  1000

3001 ACCACTGTGAAAGAGATGTCGAAGAAAGCGCCCTCTGAAATTAGCCGGAAATATCAATCA  3060
```

FIG. 19C

```
3001 ACCACTGTGAAAGAGATGTCGAAGAAAGCGCCCTCTGAAATTAGCCGGAAATATCAATCA  3060
1001 -T--T--V--K--E--M--S--K--K--A--P--S--E--I--S--R--K--Y--Q--S-  1020

3061 GAATTTGAAGAAATTGAGGGACGCTGGAAGAAGCTCTCCTCCCAGCTGGTTGAGCATTGT  3120
3061 GAATTTGAAGAAATTGAGGGACGCTGGAAGAAGCTCTCCTCCCAGCTGGTTGAGCATTGT  3120
1021 -E--F--E--E--I--E--G--R--W--K--K--L--S--S--Q--L--V--E--H--C-  1040

3121 CAAAAGCTAGAGGAGCAAATGAATAAACTCCGAAAAATTCAGaatcacatacaaaccctg  3180
3121 CAAAAGCTAGAGGAGCAAATGAATAAACTCCGAAAAATTCAGAATCACATACAAACCCTG  3180
1041 -Q--K--L--E--E--Q--M--N--K--L--R--K--I--Q--N--H--I--Q--T--L-  1060

3181 aagaaatggatggctgaagttgatgtttttctgaaggaggaatggcctgcccttggggat  3240
3181 AAGAAATGGATGGCTGAAGTTGATGTTTTTCTGAAGGAGGAATGGCCTGCCCTTGGGGAT  3240
1061 -K--K--W--M--A--E--V--D--V--F--L--K--E--E--W--P--A--L--G--D-  1080

3241 tcagaaattctaaaaaagcagctgaaacagtgcagaCTTTTAGTCAGTGATATTCAGACA  3300
3241 TCAGAAATTCTAAAAAAGCAGCTGAAACAGTGCAGACTTTTAGTCAGTGATATTCAGACA  3300
1081 -S--E--I--L--K--K--Q--L--K--Q--C--R--L--L--V--S--D--I--Q--T-  1100

3301 ATTCAGCCCAGTCTAAACAGTGTCAATGAAGGTGGGCAGAAGATAAAGAATGAAGCAGAG  3360
3301 ATTCAGCCCAGTCTAAACAGTGTCAATGAAGGTGGGCAGAAGATAAAGAATGAAGCAGAG  3360
1101 -I--Q--P--S--L--N--S--V--N--E--G--G--Q--K--I--K--N--E--A--E-  1120

3361 CCAGAGTTTGCTTCGAGACTTGAGACAGAACTCAAAGAACTTAACACTCAGTGGGATCAC  3420
3361 CCAGAGTTTGCTTCGAGACTTGAGACAGAACTCAAAGAACTTAACACTCAGTGGGATCAC  3420
1121 -P--E--F--A--S--R--L--E--T--E--L--K--E--L--N--T--Q--W--D--H-  1140

3421 ATGTGCCAACAGgtctatgccagaaaggaggccttgaagggaggtttggagaaaactgta  3480
3421 ATGTGCCAACAGGTCTATGCCAGAAAGGAGGCCTTGAAGGGAGGTTTGGAGAAAACTGTA  3480
1141 -M--C--Q--Q--V--Y--A--R--K--E--A--L--K--G--G--L--E--K--T--V-  1160

3481 agcctccagaaagatctatcagagatgcacgaatggatgacacaagctgaagaagagtat  3540
3481 AGCCTCCAGAAAGATCTATCAGAGATGCACGAATGGATGACACAAGCTGAAGAAGAGTAT  3540
1161 -S--L--Q--K--D--L--S--E--M--H--E--W--M--T--Q--A--E--E--E--Y-  1180

3541 cttgagagagattttgaatataaaactccagatgaattacagaaagcagttgaagagatg  3600
3541 CTTGAGAGAGATTTTGAATATAAAACTCCAGATGAATTACAGAAAGCAGTTGAAGAGATG  3600
1181 -L--E--R--D--F--E--Y--K--T--P--D--E--L--Q--K--A--V--E--E--M-  1200

3601 aagAGAGCTAAAGAAGAGGCCCAACAAAAAGAAGCGAAAGTGAAACTCCTTACTGAGTCT  3660
3601 AAGAGAGCTAAAGAAGAGGCCCAACAAAAAGAAGCGAAAGTGAAACTCCTTACTGAGTCT  3660
1201 -K--R--A--K--E--E--A--Q--Q--K--E--A--K--V--K--L--L--T--E--S-  1220

3661 GTAAATAGTGTCATAGCTCAAGCTCCACCTGTAGCACAAGAGGCCTTAAAAAAGGAACTT  3720
3661 GTAAATAGTGTCATAGCTCAAGCTCCACCTGTAGCACAAGAGGCCTTAAAAAAGGAACTT  3720
1221 -V--N--S--V--I--A--Q--A--P--P--V--A--Q--E--A--L--K--K--E--L-  1240

3721 GAAACTCTAACCACCAACTACCAGTGGCTCTGCACTAGGCTGAATGGGAAATGCAAGACT  3780
3721 GAAACTCTAACCACCAACTACCAGTGGCTCTGCACTAGGCTGAATGGGAAATGCAAGACT  3780
1241 -E--T--L--T--T--N--Y--Q--W--L--C--T--R--L--N--G--K--C--K--T-  1260

3781 TTGGAAgaagtttgggcatgttggcatgagttattgtcatacttggagaaagcaaacaag  3840
3781 TTGGAAGAAGTTTGGGCATGTTGGCATGAGTTATTGTCATACTTGGAGAAAGCAAACAAG  3840
1261 -L--E--E--V--W--A--C--W--H--E--L--L--S--Y--L--E--K--A--N--K-  1280

3841 tggctaaatgaagtagaatttaaacttaaaaccactgaaaacattcctggcggagctgag  3900
3841 TGGCTAAATGAAGTAGAATTTAAACTTAAAACCACTGAAAACATTCCTGGCGGAGCTGAG  3900
1281 -W--L--N--E--V--E--F--K--L--K--T--T--E--N--I--P--G--G--A--E-  1300

3901 gaaatctctgaggtgctagatTCACTTGAAAATTTGATGCGACATTCAGAGGATAACCCA  3960
3901 GAAATCTCTGAGGTGCTAGATTCACTTGAAAATTTGATGCGACATTCAGAGGATAACCCA  3960
1301 -E--I--S--E--V--L--D--S--L--E--N--L--M--R--H--S--E--D--N--P-  1320

3961 AATCAGATTCGCATATTGGCACAGACCCTAACAGATGGCGGAGTCATGGATGAGCTAATC  4020
3961 AATCAGATTCGCATATTGGCACAGACCCTAACAGATGGCGGAGTCATGGATGAGCTAATC  4020
1321 -N--Q--I--R--I--L--A--Q--T--L--T--D--G--G--V--M--D--E--L--I-  1340

4021 AATGAGGAACTTGAGACATTTAATTCTCGTTGGAGGGAACTACATGAAGAGgctgtaagg  4080
```

FIG. 19D

```
4021 AATGAGGAACTTGAGACATTTAATTCTCGTTGGAGGGAACTACATGAAGAGGCTGTAAGG  4080
 1341 -N--E--E--L--E--T--F--N--S--R--W--R--E--L--H--E--E--A--V--R-  1360

 4081 aggcaaaagttgcttgaacagagcatccagtctgcccaggagactgaaaaatccttacac  4140
 4081 AGGCAAAAGTTGCTTGAACAGAGCATCCAGTCTGCCCAGGAGACTGAAAAATCCTTACAC  4140
 1361 -R--Q--K--L--L--E--Q--S--I--Q--S--A--Q--E--T--E--K--S--L--H-  1380

 4141 ttaatccaggagtccctcacattcattgacaagcagttggcagcttatattgcagacaag  4200
 4141 TTAATCCAGGAGTCCCTCACATTCATTGACAAGCAGTTGGCAGCTTATATTGCAGACAAG  4200
 1381 -L--I--Q--E--S--L--T--F--I--D--K--Q--L--A--A--Y--I--A--D--K-  1400

4201 gtggacgcagctcaaatgcctcaggaagcccagAAAATCCAATCTGATTTGACAAGTCAT  4260
4201 GTGGACGCAGCTCAAATGCCTCAGGAAGCCCAGAAAATCCAATCTGATTTGACAAGTCAT  4260
1401 -V--D--A--A--Q--M--P--Q--E--A--Q--K--I--Q--S--D--L--T--S--H-  1420

4261 GAGATCAGTTTAGAAGAAATGAAGAAACATAATCAGGGGAAGGAGGCTGCCCAAAGAGTC  4320
4261 GAGATCAGTTTAGAAGAAATGAAGAAACATAATCAGGGGAAGGAGGCTGCCCAAAGAGTC  4320
1421 -E--I--S--L--E--E--M--K--K--H--N--Q--G--K--E--A--A--Q--R--V-  1440

4321 CTGTCTCAGATTGATGTTGCACAGaaaaaattacaagatgtctccatgaagtttcgatta  4380
4321 CTGTCTCAGATTGATGTTGCACAGAAAAAATTACAAGATGTCTCCATGAAGTTTCGATTA  4380
1441 -L--S--Q--I--D--V--A--Q--K--K--L--Q--D--V--S--M--K--F--R--L-  1460

4381 ttccagaaaccagccaattttgagcagcgtctacaagaaagtaagatgattttagatgaa  4440
4381 TTCCAGAAACCAGCCAATTTTGAGCAGCGTCTACAAGAAAGTAAGATGATTTTAGATGAA  4440
1461 -F--Q--K--P--A--N--F--E--Q--R--L--Q--E--S--K--M--I--L--D--E-  1480

4441 gtgaagatgcacttgcctgcattggaaacaaagagtgtggaacaggaagtagtacagtca  4500
4441 GTGAAGATGCACTTGCCTGCATTGGAAACAAAGAGTGTGGAACAGGAAGTAGTACAGTCA  4500
1481 -V--K--M--H--L--P--A--L--E--T--K--S--V--E--Q--E--V--V--Q--S-  1500

4501 cagctaaatcattgtgtgAACTTGTATAAAAGTCTGAGTGAAGTGAAGTCTGAAGTGGAA  4560
4501 CAGCTAAATCATTGTGTGAACTTGTATAAAAGTCTGAGTGAAGTGAAGTCTGAAGTGGAA  4560
1501 -Q--L--N--H--C--V--N--L--Y--K--S--L--S--E--V--K--S--E--V--E-  1520

4561 ATGGTGATAAAGACTGGACGTCAGATTGTACAGAAAAAGCAGACGGAAAATCCCAAAGAA  4620
4561 ATGGTGATAAAGACTGGACGTCAGATTGTACAGAAAAAGCAGACGGAAAATCCCAAAGAA  4620
1521 -M--V--I--K--T--G--R--Q--I--V--Q--K--K--Q--T--E--N--P--K--E-  1540

4621 CTTGATGAAAGAGTAACAGCTTTGAAATTGCATTATAATGAGCTGGGAGCAAAGgtaaca  4680
4621 CTTGATGAAAGAGTAACAGCTTTGAAATTGCATTATAATGAGCTGGGAGCAAAGGTAACA  4680
1541 -L--D--E--R--V--T--A--L--K--L--H--Y--N--E--L--G--A--K--V--T-  1560

4681 gaaagaaagcaacagttggagaaatgcttgaaattgtcccgtaagatgcgaaaggaaatg  4740
4681 GAAAGAAAGCAACAGTTGGAGAAATGCTTGAAATTGTCCCGTAAGATGCGAAAGGAAATG  4740
1561 -E--R--K--Q--Q--L--E--K--C--L--K--L--S--R--K--M--R--K--E--M-  1580

4741 aatgtcttgacagaatggctggcagctacagatatggaattgacaaagagatcagcagtt  4800
4741 AATGTCTTGACAGAATGGCTGGCAGCTACAGATATGGAATTGACAAAGAGATCAGCAGTT  4800
1581 -N--V--L--T--E--W--L--A--A--T--D--M--E--L--T--K--R--S--A--V-  1600

4801 gaaggaatgcctagtaatttggattctgaagttgcctggggaaagGCTACTCAAAAAGAG  4860
4801 GAAGGAATGCCTAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAGGCTACTCAAAAAGAG  4860
1601 -E--G--M--P--S--N--L--D--S--E--V--A--W--G--K--A--T--Q--K--E-  1620

4861 ATTGAGAAACAGAAGGTGCACCTGAAGAGTATCACAGAGGTAGGAGAGGCCTTGAAAACA  4920
1621 -I--E--K--Q--K--V--H--L--K--S--I--T--E--V--G--E--A--L--K--T-  1640

4921 GTTTTGGGCAAGAAGGAGACGTTGGTGGAAGATAAACTCAGTCTTCTGAATAGTAACTGG  4980
4921 GTTTTGGGCAAGAAGGAGACGTTGGTGGAAGATAAACTCAGTCTTCTGAATAGTAACTGG  4980
1641 -V--L--G--K--K--E--T--L--V--E--D--K--L--S--L--L--N--S--N--W-  1660

4981 ATAGCTGTCACCTCCCGAGCAGAAGAGTGGTTAAATCTTTTGTTGgaataccagaaacac  5040
4981 ATAGCTGTCACCTCCCGAGCAGAAGAGTGGTTAAATCTTTTGTTGGAATACCAGAAACAC  5040
1661 -I--A--V--T--S--R--A--E--E--W--L--N--L--L--L--E--Y--Q--K--H-  1680

5041 atggaaacttttgaccagaatgtggaccacatcacaaagtggatcattcaggctgacaca  5100
```

FIG. 19E

```
5041 ATGGAAACTTTTGACCAGAATGTGGACCACATCACAAAGTGGATCATTCAGGCTGACACA  5100
1681 -M--E--T--F--D--Q--N--V--D--H--I--T--K--W--I--I--Q--A--D--T-  1700

5101 cttttggatgaatcagagaaaaagaaaccccagcaaaaagaagacgtgcttaagCGTTTA  5160
5101 CTTTTGGATGAATCAGAGAAAAAGAAACCCCAGCAAAAAGAAGACGTGCTTAAGCGTTTA  5160
1701 -L--L--D--E--S--E--K--K--K--P--Q--Q--K--E--D--V--L--K--R--L-  1720

5161 AAGGCAGAACTGAATGACATACGCCCAAAGGTGGACTCTACACGTGACCAAGCAGCAAAC  5220
5161 AAGGCAGAACTGAATGACATACGCCCAAAGGTGGACTCTACACGTGACCAAGCAGCAAAC  5220
1721 -K--A--E--L--N--D--I--R--P--K--V--D--S--T--R--D--Q--A--A--N-  1740

5221 TTGATGGCAAACCGCGGTGACCACTGCAGGAAATTAGTAGAGCCCCAAATCTCAGAGCTC  5280
5221 TTGATGGCAAACCGCGGTGACCACTGCAGGAAATTAGTAGAGCCCCAAATCTCAGAGCTC  5280
1741 -L--M--A--N--R--G--D--H--C--R--K--L--V--E--P--Q--I--S--E--L-  1760

5281 AACCATCGATTTGCAGCCATTTCACACAGAATTAAGACTGGAAAGgcctccattcctttg  5340
5281 AACCATCGATTTGCAGCCATTTCACACAGAATTAAGACTGGAAAGGCCTCCATTCCTTTG  5340
1761 -N--H--R--F--A--A--I--S--H--R--I--K--T--G--K--A--S--I--P--L-  1780

5341 aaggaattggagcagtttaactcagatatacaaaaattgcttgaaccactggaggctgaa  5400
5341 AAGGAATTGGAGCAGTTTAACTCAGATATACAAAAATTGCTTGAACCACTGGAGGCTGAA  5400
1781 -K--E--L--E--Q--F--N--S--D--I--Q--K--L--L--E--P--L--E--A--E-  1800

5401 attcagcaggggtgaatctgaaagaggaagacttcaataaagatatgAATGAAGACAAT  5460
5401 ATTCAGCAGGGGGTGAATCTGAAAGAGGAAGACTTCAATAAAGATATGAATGAAGACAAT  5460
1801 -I--Q--Q--G--V--N--L--K--E--E--D--F--N--K--D--M--N--E--D--N-  1820

5461 GAGGGTACTGTAAAAGAATTGTTGCAAAGAGGAGACAACTTACAACAAAGAATCACAGAT  5520
5461 GAGGGTACTGTAAAAGAATTGTTGCAAAGAGGAGACAACTTACAACAAAGAATCACAGAT  5520
1821 -E--G--T--V--K--E--L--L--Q--R--G--D--N--L--Q--Q--R--I--T--D-  1840

5521 GAGAGAAAGCGAGAGGAAATAAAGATAAAACAGCAGCTGTTACAGACAAAACATAATGCT  5580
5521 GAGAGAAAGCGAGAGGAAATAAAGATAAAACAGCAGCTGTTACAGACAAAACATAATGCT  5580
1841 -E--R--K--R--E--E--I--K--I--K--Q--Q--L--L--Q--T--K--H--N--A-  1860

5581 CTCAAGgatttgaggtctcaaagaagaaaaaaggctctagaaatttctcatcagtggtat  5640
5581 CTCAAGGATTTGAGGTCTCAAAGAAGAAAAAAGGCTCTAGAAATTTCTCATCAGTGGTAT  5640
1861 -L--K--D--L--R--S--Q--R--R--K--K--A--L--E--I--S--H--Q--W--Y-  1880

5641 cagtacaagaggcaggctgatgatctcctgaaatgcttggatgacattgaaaaaaaatta  5700
5641 CAGTACAAGAGGCAGGCTGATGATCTCCTGAAATGCTTGGATGACATTGAAAAAAAATTA  5700
1881 -Q--Y--K--R--Q--A--D--D--L--L--K--C--L--D--D--I--E--K--K--L-  1900

5701 gccagcctacctgagcccagagatgaaaggaaaataaagGAAATTGATCGGGAATTGCAG  5760
5701 GCCAGCCTACCTGAGCCCAGAGATGAAAGGAAAATAAAGGAAATTGATCGGGAATTGCAG  5760
1901 -A--S--L--P--E--P--R--D--E--R--K--I--K--E--I--D--R--E--L--Q-  1920

5761 AAGAAGAAAGAGGAGCTGAATGCAGTGCGTAGGCAAGCTGAGGGCTTGTCTGAGGATGGG  5820
5761 AAGAAGAAAGAGGAGCTGAATGCAGTGCGTAGGCAAGCTGAGGGCTTGTCTGAGGATGGG  5820
1921 -K--K--K--E--E--L--N--A--V--R--R--Q--A--E--G--L--S--E--D--G-  1940

5821 GCCGCAATGGCAGTGGAGCCAACTCAGATCCAGCTCAGCAAGCGCTGGCGGGAAATTGAG  5880
5821 GCCGCAATGGCAGTGGAGCCAACTCAGATCCAGCTCAGCAAGCGCTGGCGGGAAATTGAG  5880
1941 -A--A--M--A--V--E--P--T--Q--I--Q--L--S--K--R--W--R--E--I--E-  1960

5881 AGCAAATTTGCTCAGTTTCGAAGACTCAACTTTGCACAAATTcacactgtccgtgaagaa  5940
5881 AGCAAATTTGCTCAGTTTCGAAGACTCAACTTTGCACAAATTCACACTGTCCGTGAAGAA  5940
1961 -S--K--F--A--Q--F--R--R--L--N--F--A--Q--I--H--T--V--R--E--E-  1980

5941 acgatgatggtgatgactgaagacatgcctttggaaatttcttatgtgccttctacttat  6000
5941 ACGATGATGGTGATGACTGAAGACATGCCTTTGGAAATTTCTTATGTGCCTTCTACTTAT  6000
1981 -T--M--M--V--M--T--E--D--M--P--L--E--I--S--Y--V--P--S--T--Y-  2000

6001 ttgactgaaatcactcatgtctcacaagccctattagaagtggaacaacttctcaatgct  6060
6001 TTGACTGAAATCACTCATGTCTCACAAGCCCTATTAGAAGTGGAACAACTTCTCAATGCT  6060
2001 -L--T--E--I--T--H--V--S--Q--A--L--L--E--V--E--Q--L--L--N--A-  2020

6061 cctgacctctgtgctaaggactttgaagatctctttaagcaagaggagtctctgaagAAT  6120
```

Figure 19F

```
6061 CCTGACCTCTGTGCTAAGGACTTTGAAGATCTCTTTAAGCAAGAGGAGTCTCTGAAGAAT  6120
 2021 -P--D--L--C--A--K--D--F--E--D--L--F--K--Q--E--E--S--L--K--N-  2040

 6121 ATAAAAGATAGTCTACAACAAAGCTCAGGTCGGATTGACATTATTCATAGCAAGAAGACA  6180
 6121 ATAAAAGATAGTCTACAACAAAGCTCAGGTCGGATTGACATTATTCATAGCAAGAAGACA  6180
 2041 -I--K--D--S--L--Q--Q--S--S--G--R--I--D--I--I--H--S--K--K--T-  2060

 6181 GCAGCATTGCAAAGTGCAACGCCTGTGGAAAGGGTGAAGCTACAGGAAGCTCTCTCCCAG  6240
 6181 GCAGCATTGCAAAGTGCAACGCCTGTGGAAAGGGTGAAGCTACAGGAAGCTCTCTCCCAG  6240
 2061 -A--A--L--Q--S--A--T--P--V--E--R--V--K--L--Q--E--A--L--S--Q-  2080

6241 CTTGATTTCCAATGGGAAAAAGTTAACAAAATGTACAAGGACCGACAAGGcgatttgac  6300
6241 CTTGATTTCCAATGGGAAAAAGTTAACAAAATGTACAAGGACCGACAAGGGCGATTTGAC  6300
2081 -L--D--F--Q--W--E--K--V--N--K--M--Y--K--D--R--Q--G--R--F--D-  2100

6301 agatctgttgagaaatggcggcgttttcattatgatataaagatatttaatcagtggcta  6360
6301 AGATCTGTTGAGAAATGGCGGCGTTTTCATTATGATATAAAGATATTTAATCAGTGGCTA  6360
2101 -R--S--V--E--K--W--R--R--F--H--Y--D--I--K--I--F--N--Q--W--L-  2120

6361 acagaagctgaacagtttctcagaaagacacaaattcctgagaattgggaacatgctaaa  6420
6361 ACAGAAGCTGAACAGTTTCTCAGAAAGACACAAATTCCTGAGAATTGGGAACATGCTAAA  6420
2121 -T--E--A--E--Q--F--L--R--K--T--Q--I--P--E--N--W--E--H--A--K-  2140

6421 tacaaatggtatcttaagGAACTCCAGGATGGCATTGGGCAGCGGCAAACTGTTGTCAGA  6480
6421 TACAAATGGTATCTTAAGGAACTCCAGGATGGCATTGGGCAGCGGCAAACTGTTGTCAGA  6480
2141 -Y--K--W--Y--L--K--E--L--Q--D--G--I--G--Q--R--Q--T--V--V--R-  2160

6481 ACATTGAATGCAACTGGGGAAGAAATAATTCAGCAATCCTCAAAAACAGATGCCAGTATT  6540
6481 ACATTGAATGCAACTGGGGAAGAAATAATTCAGCAATCCTCAAAAACAGATGCCAGTATT  6540
2161 -T--L--N--A--T--G--E--E--I--I--Q--Q--S--S--K--T--D--A--S--I-  2180

6541 CTACAGGAAAAATTGGGAAGCCTGAATCTGCGGTGGCAGGAGGTCTGCAAACAGCTGTCA  6600
6541 CTACAGGAAAAATTGGGAAGCCTGAATCTGCGGTGGCAGGAGGTCTGCAAACAGCTGTCA  6600
2181 -L--Q--E--K--L--G--S--L--N--L--R--W--Q--E--V--C--K--Q--L--S-  2200

6601 GACAGAAAAAAGAGgctagaagaacaaaagaatatcttgtcagaatttcaaagagattta  6660
6601 GACAGAAAAAAGAGGCTAGAAGAACAAAAGAATATCTTGTCAGAATTTCAAAGAGATTTA  6660
2201 -D--R--K--K--R--L--E--E--Q--K--N--I--L--S--E--F--Q--R--D--L-  2220

6661 aatgaatttgttttatggttggaggaagcagataacattgctagtatcccacttgaacct  6720
6661 AATGAATTTGTTTTATGGTTGGAGGAAGCAGATAACATTGCTAGTATCCCACTTGAACCT  6720
2221 -N--E--F--V--L--W--L--E--E--A--D--N--I--A--S--I--P--L--E--P-  2240

6721 ggaaaagagcagcaactaaaagaaaagcttgagcaagtcaagTTACTGGTGGAAGAGTTG  6780
6721 GGAAAAGAGCAGCAACTAAAAGAAAAGCTTGAGCAAGTCAAGTTACTGGTGGAAGAGTTG  6780
2241 -G--K--E--Q--Q--L--K--E--K--L--E--Q--V--K--L--L--V--E--E--L-  2260

6781 CCCCTGCGCCAGGGAATTCTCAAACAATTAAATGAAACTGGAGGACCCGTGCTTGTAAGT  6840
6781 CCCCTGCGCCAGGGAATTCTCAAACAATTAAATGAAACTGGAGGACCCGTGCTTGTAAGT  6840
2261 -P--L--R--Q--G--I--L--K--Q--L--N--E--T--G--G--P--V--L--V--S-  2280

6841 GCTCCCATAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAGCTCAAGCAGACAAATCTC  6900
6841 GCTCCCATAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAGCTCAAGCAGACAAATCTC  6900
2281 -A--P--I--S--P--E--E--Q--D--K--L--E--N--K--L--K--Q--T--N--L-  2300

6901 CAGTGGATAAAGgtttccagagctttacctgagaaacaaggagaaattgaagctcaaata  6960
6901 CAGTGGATAAAGGTTTCCAGAGCTTTACCTGAGAAACAAGGAGAAATTGAAGCTCAAATA  6960
2301 -Q--W--I--K--V--S--R--A--L--P--E--K--Q--G--E--I--E--A--Q--I-  2320

6961 aaagaccttgggcagcttgaaaaaaagcttgaagaccttgaagagcagttaaatcatctg  7020
6961 AAAGACCTTGGGCAGCTTGAAAAAAAGCTTGAAGACCTTGAAGAGCAGTTAAATCATCTG  7020
2321 -K--D--L--G--Q--L--E--K--K--L--E--D--L--E--E--Q--L--N--H--L-  2340

7021 ctgctgtggttatctcctattaggaatcagttggaaatttataaccaaccaaaccaagaa  7080
7021 CTGCTGTGGTTATCTCCTATTAGGAATCAGTTGGAAATTTATAACCAACCAAACCAAGAA  7080
2341 -L--L--W--L--S--P--I--R--N--Q--L--E--I--Y--N--Q--P--N--Q--E-  2360

7081 ggaccatttgacgttaagGAAACTGAAATAGCAGTTCAAGCTAAACAACCGGATGTGGAA  7140
```

FIG. 19G

```
7081 GGACCATTTGACGTTAAGGAAACTGAAATAGCAGTTCAAGCTAAACAACCGGATGTGGAA  7140
 2361 -G--P--F--D--V--K--E--T--E--I--A--V--Q--A--K--Q--P--D--V--E-  2380

 7141 GAGATTTTGTCTAAAGGGCAGCATTTGTACAAGGAAAAACCAGCCACTCAGCCAGTGAAG  7200
 7141 GAGATTTTGTCTAAAGGGCAGCATTTGTACAAGGAAAAACCAGCCACTCAGCCAGTGAAG  7200
 2381 -E--I--L--S--K--G--Q--H--L--Y--K--E--K--P--A--T--Q--P--V--K-  2400

 7201 aggaagttagaagatctgagctctgagtggaaggcggtaaaccgtttacttcaagagctg  7260
 7201 AGGAAGTTAGAAGATCTGAGCTCTGAGTGGAAGGCGGTAAACCGTTTACTTCAAGAGCTG  7260
 2401 -R--K--L--E--D--L--S--S--E--W--K--A--V--N--R--L--L--Q--E--L-  2420

 7261 agggcaaagcagcctgacctagctcctggactgaccactattggagcctCTCCTACTCAG  7320
 7261 AGGGCAAAGCAGCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCTCCTACTCAG  7320
 2421 -R--A--K--Q--P--D--L--A--P--G--L--T--T--I--G--A--S--P--T--Q-  2440

 7321 ACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAA  7380
 7321 ACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAA  7380
 2441 -T--V--T--L--V--T--Q--P--V--V--T--K--E--T--A--I--S--K--L--E-  2460

 7381 ATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCTGGCAGATTTCAACCGGGCTTGGACA  7440
 7381 ATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCTGGCAGATTTCAACCGGGCTTGGACA  7440
 2461 -M--P--S--S--L--M--L--E--V--P--A--L--A--D--F--N--R--A--W--T-  2480

 7441 GAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATCACAGAGGGTGATGGTG  7500
 7441 GAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATCACAGAGGGTGATGGTG  7500
 2481 -E--L--T--D--W--L--S--L--L--D--Q--V--I--K--S--Q--R--V--M--V-  2500

 7501 GGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAGgcaacaatgcaggatttg  7560
 7501 GGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAGGCAACAATGCAGGATTTG  7560
 2501 -G--D--L--E--D--I--N--E--M--I--I--K--Q--K--A--T--M--Q--D--L-  2520

 7561 gaacagaggcgtccccagttggaagaactcattaccgctgcccaaaatttgaaaaacaag  7620
 7561 GAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAG  7620
 2521 -E--Q--R--R--P--Q--L--E--E--L--I--T--A--A--Q--N--L--K--N--K-  2540

 7621 accagcaatcaagaggctagaacaatcattacggatcgaaTTGAAAGAATTCAGAATCAG  7680
 7621 ACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAATTGAAAGAATTCAGAATCAG  7680
 2541 -T--S--N--Q--E--A--R--T--I--I--T--D--R--I--E--R--I--Q--N--Q-  2560

 7681 TGGGATGAAGTACAAGAACACCTTCAGAACCGGAGGCAACAGTTGAATGAAATGTTAAAG  7740
 7681 TGGGATGAAGTACAAGAACACCTTCAGAACCGGAGGCAACAGTTGAATGAAATGTTAAAG  7740
 2561 -W--D--E--V--Q--E--H--L--Q--N--R--R--Q--Q--L--N--E--M--L--K-  2580

 7741 GATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCTTAGGACAGGCCAGA  7800
 7741 GATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCTTAGGACAGGCCAGA  7800
 2581 -D--S--T--Q--W--L--E--A--K--E--E--A--E--Q--V--L--G--Q--A--R-  2600

 7801 GCCAAGCTTGAGTCATGGAAGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATC  7860
 7801 GCCAAGCTTGAGTCATGGAAGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATC  7860
 2601 -A--K--L--E--S--W--K--E--G--P--Y--T--V--D--A--I--Q--K--K--I-  2620

 7861 ACAGAAACCAAGcagttggccaaagacctccgccagtggcagacaaatgtagatgtggca  7920
 7861 ACAGAAACCAAGCAGTTGGCCAAAGACCTCCGCCAGTGGCAGACAAATGTAGATGTGGCA  7920
 2621 -T--E--T--K--Q--L--A--K--D--L--R--Q--W--Q--T--N--V--D--V--A-  2640

 7921 aatgacttggccctgaaacttctccgggattattctgcagatgataccagaaaagtccac  7980
 7921 AATGACTTGGCCCTGAAACTTCTCCGGGATTATTCTGCAGATGATACCAGAAAAGTCCAC  7980
 2641 -N--D--L--A--L--K--L--L--R--D--Y--S--A--D--D--T--R--K--V--H-  2660

 7981 atgataacagagaatatcaatgcctcttggagaagcattcataaaagGGTGAGTGAGCGA  8040
 7981 ATGATAACAGAGAATATCAATGCCTCTTGGAGAAGCATTCATAAAAGGGTGAGTGAGCGA  8040
 2661 -M--I--T--E--N--I--N--A--S--W--R--S--I--H--K--R--V--S--E--R-  2680

 8041 GAGGCTGCTTTGGAAGAAACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAAG  8100
 8041 GAGGCTGCTTTGGAAGAAACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAAG  8100
 2681 -E--A--A--L--E--E--T--H--R--L--L--Q--Q--F--P--L--D--L--E--K-  2700

 8101 TTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCCTACAGGATGCTACCCGT  8160
```

FIG. 19H

```
8101 TTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCCTACAGGATGCTACCCGT  8160
 2701 -F--L--A--W--L--T--E--A--E--T--T--A--N--V--L--Q--D--A--T--R-  2720

 8161 AAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAgac  8220
 8161 AAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAGAC  8220
 2721 -K--E--R--L--L--E--D--S--K--G--V--K--E--L--M--K--Q--W--Q--D-  2740

 8221 ctccaaggtgaaattgaagctcacacagatgtttatcacaacctggatgaaaacagccaa  8280
 8221 CTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAAACAGCCAA  8280
 2741 -L--Q--G--E--I--E--A--H--T--D--V--Y--H--N--L--D--E--N--S--Q-  2760

 8281 aaaatcctgagatccctggaaggttccgatgatgcagtcctgttacaaagacgtttggat  8340
 8281 AAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCTGTTACAAAGACGTTTGGAT  8340
 2761 -K--I--L--R--S--L--E--G--S--D--D--A--V--L--L--Q--R--R--L--D-  2780

 8341 aacatgaacttcaagtggagtgaacttcggaaaaagtctctcaacattagGTCCCATTTG  8400
 8341 AACATGAACTTCAAGTGGAGTGAACTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTG  8400
 2781 -N--M--N--F--K--W--S--E--L--R--K--K--S--L--N--I--R--S--H--L-  2800

 8401 GAAGCCAGTTCTGACCAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGTGG  8460
 8401 GAAGCCAGTTCTGACCAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGTGG  8460
 2801 -E--A--S--S--D--Q--W--K--R--L--H--L--S--L--Q--E--L--L--V--W-  2820

 8461 CTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGCA  8520
 8461 CTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGCA  8520
 2821 -L--Q--L--K--D--D--E--L--S--R--Q--A--P--I--G--G--D--F--P--A-  2840

 8521 GTTCAGAAGCAGAACGATGTACATAGGgccttcaagagggaattgaaaactaaagaacct  8580
 8521 GTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAAAACTAAAGAACCT  8580
 2841 -V--Q--K--Q--N--D--V--H--R--A--F--K--R--E--L--K--T--K--E--P-  2860

 8581 gtaatcatgagtactcttgagactgtacgaatatttctgacagagcagcctttggaagga  8640
 8581 GTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGA  8640
 2861 -V--I--M--S--T--L--E--T--V--R--I--F--L--T--E--Q--P--L--E--G-  2880

 8641 ctagagaaactctaccaggagcccagagAGCTGCCTCCTGAGGAGAGAGCCCAGAATGTC  8700
 8641 CTAGAGAAACTCTACCAGGAGCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCAGAATGTC  8700
 2881 -L--E--K--L--Y--Q--E--P--R--E--L--P--P--E--E--R--A--Q--N--V-  2900

 8701 ACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCTG  8760
 8701 ACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCTG  8760
 2901 -T--R--L--L--R--K--Q--A--E--E--V--N--T--E--W--E--K--L--N--L-  2920

 8761 CACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCTTGAAAGACTCCGGGAACTTCAA  8820
 8761 CACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCTTGAAAGACTCCGGGAACTTCAA  8820
 2921 -H--S--A--D--W--Q--R--K--I--D--E--T--L--E--R--L--R--E--L--Q-  2940

 8821 GAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATCCTGG  8880
 8821 GAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATCCTGG  8880
 2941 -E--A--T--D--E--L--D--L--K--L--R--Q--A--E--V--I--K--G--S--W-  2960

 8881 CAGCCCGTGGGCGATCTCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGgca  8940
 8881 CAGCCCGTGGGCGATCTCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCA  8940
 2961 -Q--P--V--G--D--L--L--I--D--S--L--Q--D--H--L--E--K--V--K--A-  2980

 8941 cttcgaggagaaattgcgcctctgaaagagaacgtgagccacgtcaatgaccttgctcgc  9000
 8941 CTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGC  9000
 2981 -L--R--G--E--I--A--P--L--K--E--N--V--S--H--V--N--D--L--A--R-  3000

 9001 cagcttaccactttgggcattcagctctcaccgtataacctcagcactctggaagacctg  9060
 9001 CAGCTTACCACTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTG  9060
 3001 -Q--L--T--T--L--G--I--Q--L--S--P--Y--N--L--S--T--L--E--D--L-  3020

 9061 aacaccagatggaagcttctgcagGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAA  9120
 9061 AACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAA  9120
 3021 -N--T--R--W--K--L--L--Q--V--A--V--E--D--R--V--R--Q--L--H--E-  3040

 9121 GCCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCAcgtctgtccagggtccc  9180
```

FIG. 19I

```
9121 GCCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCC  9180
3041 -A--H--R--D--F--G--P--A--S--Q--H--F--L--S--T--S--V--Q--G--P-  3060

9181 tgggagagagccatctcgccaaacaaagtgccctactatatcaaCCACGAGACTCAAACA  9240
9181 TGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAACA  9240
3061 -W--E--R--A--I--S--P--N--K--V--P--Y--Y--I--N--H--E--T--Q--T-  3080

9241 ACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGctgacctgaataat  9300
9241 ACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT  9300
3081 -T--C--W--D--H--P--K--M--T--E--L--Y--Q--S--L--A--D--L--N--N-  3100

9301 gtcagattctcagcttataggactgccatgaaactccgaagactgcagaaggccctttgc  9360
9301 GTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGC  9360
3101 -V--R--F--S--A--Y--R--T--A--M--K--L--R--R--L--Q--K--A--L--C-  3120

9361 tTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGACCAGCACAACCTCAAGCAA  9420
9361 TTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGACCAGCACAACCTCAAGCAA  9420
3121 -L--D--L--L--S--L--S--A--A--C--D--A--L--D--Q--H--N--L--K--Q-  3140

9421 AATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGC  9480
9421 AATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGC  9480
3141 -N--D--Q--P--M--D--I--L--Q--I--I--N--C--L--T--T--I--Y--D--R-  3160

9481 CTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAAC  9540
9481 CTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAAC  9540
3161 -L--E--Q--E--H--N--N--L--V--N--V--P--L--C--V--D--M--C--L--N-  3180

9541 TGGCTGCTGAATGTTTATGATACgggacgaacagggaggatccgtgtcctgtcttttaaa  9600
9541 TGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCCTGTCTTTTAAA  9600
3181 -W--L--L--N--V--Y--D--T--G--R--T--G--R--I--R--V--L--S--F--K-  3200

9601 actggcatcatttccctgtgtaaagcacatttggaagacaagtacagatACCTTTTCAAG  9660
9601 ACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAG  9660
3201 -T--G--I--I--S--L--C--K--A--H--L--E--D--K--Y--R--Y--L--F--K-  3220

9661 CAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTTCTGCATGAT  9720
9661 CAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTTCTGCATGAT  9720
3221 -Q--V--A--S--S--T--G--F--C--D--Q--R--R--L--G--L--L--L--H--D-  3240

9721 TCTATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCAGTAACATTGAG  9780
9721 TCTATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCAGTAACATTGAG  9780
3241 -S--I--Q--I--P--R--Q--L--G--E--V--A--S--F--G--G--S--N--I--E-  3260

9781 CCAAGTGTCCGGAGCTGCTTCCAATTTgctaataataagccagagatcgaagcggccctc  9840
9781 CCAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGATCGAAGCGGCCCTC  9840
3261 -P--S--V--R--S--C--F--Q--F--A--N--N--K--P--E--I--E--A--A--L-  3280

9841 ttcctagactggatgagactggaaccccagtccatggtgtggctgcccgtcctgcacaga  9900
9841 TTCCTAGACTGGATGAGACTGGAACCCCAGTCCATGGTGTGGCTGCCCGTCCTGCACAGA  9900
3281 -F--L--D--W--M--R--L--E--P--Q--S--M--V--W--L--P--V--L--H--R-  3300

9901 gtggctgctgcagaaactgccaagcatcaggccaaatgtaacatctgcaaagagtgtcca  9960
9901 GTGGCTGCTGCAGAAACTGCCAAGCATCAGGCCAAATGTAACATCTGCAAAGAGTGTCCA  9960
3301 -V--A--A--A--E--T--A--K--H--Q--A--K--C--N--I--C--K--E--C--P-  3320

9961 atcattggattcagGTACAGGAGTCTAAAGCACTTTAATTATGACATCTGCCAAAGCTGC 10020
9961 ATCATTGGATTCAGGTACAGGAGTCTAAAGCACTTTAATTATGACATCTGCCAAAGCTGC 10020
3321 -I--I--G--F--R--Y--R--S--L--K--H--F--N--Y--D--I--C--Q--S--C-  3340

10021 TTTTTTTCTGGTCGAGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATATTGC 10080
10021 TTTTTTTCTGGTCGAGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATATTGC 10080
 3341 -F--F--S--G--R--V--A--K--G--H--K--M--H--Y--P--M--V--E--Y--C-  3360

10081 ACTCCGactacatcaggagaagatgttcgagactttgccaaggtactaaaaaacaaattt 10140
10081 ACTCCGACTACATCAGGAGAAGATGTTCGAGACTTTGCCAAGGTACTAAAAAACAAATTT 10140
 3361 -T--P--T--T--S--G--E--D--V--R--D--F--A--K--V--L--K--N--K--F-  3380

10141 cgaaccaaaaggtattttgcgaagcatccccgaatgggctacctgccagtgcagactgtc 10200
```

FIG. 19J

```
10141 CGAACCAAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTACCTGCCAGTGCAGACTGTC 10200
 3381 -R--T--K--R--Y--F--A--K--H--P--R--M--G--Y--L--P--V--Q--T--V-  3400

10201 ttagaggggacaacatggaaacTCCCGTTACTCTGATCAACTTCTGGCCAGTAGATTCT 10260
10201 TTAGAGGGGGACAACATGGAAACTCCCGTTACTCTGATCAACTTCTGGCCAGTAGATTCT 10260
 3401 -L--E--G--D--N--M--E--T--P--V--T--L--I--N--F--W--P--V--D--S-  3420

10261 GCgcctgcctcgtcccctcagctttcacacgatgatactcattcacgcattgaacattat 10320
10261 GCGCCTGCCTCGTCCCCTCAGCTTTCACACGATGATACTCATTCACGCATTGAACATTAT 10320
 3421 -A--P--A--S--S--P--Q--L--S--H--D--D--T--H--S--R--I--E--H--Y-  3440

10321 gctagcagGCTAGCAGAAATGGAAAACAGCAATGGATCTTATCTAAATGATAGCATCTCT 10380
10321 GCTAGCAGGCTAGCAGAAATGGAAAACAGCAATGGATCTTATCTAAATGATAGCATCTCT 10380
 3441 -A--S--R--L--A--E--M--E--N--S--N--G--S--Y--L--N--D--S--I--S-  3460

10381 CCTAATGAGAGCATagatgatgaacatttgttaatccagcattactgccaaagtttgaac 10440
10381 CCTAATGAGAGCATAGATGATGAACATTTGTTAATCCAGCATTACTGCCAAAGTTTGAAC 10440
 3461 -P--N--E--S--I--D--D--E--H--L--L--I--Q--H--Y--C--Q--S--L--N-  3480

10441 caggactccccctgagccagcctcgtagtcctgcccagatcttgatttccttagagagt 10500
10441 CAGGACTCCCCCCTGAGCCAGCCTCGTAGTCCTGCCCAGATCTTGATTTCCTTAGAGAGT 10500
 3481 -Q--D--S--P--L--S--Q--P--R--S--P--A--Q--I--L--I--S--L--E--S-  3500

10501 gaggaaagaggggagctagagagaatcctagcagatcttgaggaagaaaacagGAATCTG 10560
10501 GAGGAAAGAGGGGAGCTAGAGAGAATCCTAGCAGATCTTGAGGAAGAAAACAGGAATCTG 10560
 3501 -E--E--R--G--E--L--E--R--I--L--A--D--L--E--E--E--N--R--N--L-  3520

10561 CAAGCAGAATATGACCGTCTAAAGCAGCAGCACGAACATAAAGGCCTGTCCCCACTGCCG 10620
10561 CAAGCAGAATATGACCGTCTAAAGCAGCAGCACGAACATAAAGGCCTGTCCCCACTGCCG 10620
 3521 -Q--A--E--Y--D--R--L--K--Q--Q--H--E--H--K--G--L--S--P--L--P-  3540

10621 TCCCCTCCTGAAATGATGCCCACCTCTCCCCAGAGTCCCCGGGATGCTGAGCTCATTGCT 10680
10621 TCCCCTCCTGAAATGATGCCCACCTCTCCCCAGAGTCCCCGGGATGCTGAGCTCATTGCT 10680
 3541 -S--P--P--E--M--M--P--T--S--P--Q--S--P--R--D--A--E--L--I--A-  3560

10681 GAGGCCAAGCTACTGCGTCAACACAAAGGCCGCCTGGAAGCCAGGATGCAAATCCTGGAA 10740
10681 GAGGCCAAGCTACTGCGTCAACACAAAGGCCGCCTGGAAGCCAGGATGCAAATCCTGGAA 10740
 3561 -E--A--K--L--L--R--Q--H--K--G--R--L--E--A--R--M--Q--I--L--E-  3580

10741 GACCACAATAAACAGCTGGAGTCACAGTTACACAGGCTAAGGCAGCTGCTGGAGCAAccc 10800
10741 GACCACAATAAACAGCTGGAGTCACAGTTACACAGGCTAAGGCAGCTGCTGGAGCAACCC 10800
 3581 -D--H--N--K--Q--L--E--S--Q--L--H--R--L--R--Q--L--L--E--Q--P-  3600

10801 caggcagaggccaaagtgaatggcacaacggtgtcctctccttctacctctctacagagg 10860
10801 CAGGCAGAGGCCAAAGTGAATGGCACAACGGTGTCCTCTCCTTCTACCTCTCTACAGAGG 10860
 3601 -Q--A--E--A--K--V--N--G--T--T--V--S--S--P--S--T--S--L--Q--R-  3620

10861 tccgacagcagtcagcctatgctgctccgagtggttggcagtcaaacttcggactccatg 10920
10861 TCCGACAGCAGTCAGCCTATGCTGCTCCGAGTGGTTGGCAGTCAAACTTCGGACTCCATG 10920
 3621 -S--D--S--S--Q--P--M--L--L--R--V--V--G--S--Q--T--S--D--S--M-  3640

10921 gGTGAGGAAGATCTTCTCAGTCCTCCCCAGGACACAAGCACAGGGTTAGAGGAGGTGATG 10980
10921 GGTGAGGAAGATCTTCTCAGTCCTCCCCAGGACACAAGCACAGGGTTAGAGGAGGTGATG 10980
 3641 -G--E--E--D--L--L--S--P--P--Q--D--T--S--T--G--L--E--E--V--M-  3660

10981 GAGCAACTCAACAACTCCTTCCCTAGTTCAAGAGgaagaaataccccctggaaagccaatg 11040
10981 GAGCAACTCAACAACTCCTTCCCTAGTTCAAGAGGAAGAAATACCCCTGGAAAGCCAATG 11040
 3661 -E--Q--L--N--N--S--F--P--S--S--R--G--R--N--T--P--G--K--P--M-  3680

11041 agagagGACACAATGTAG                                           11058
11041 AGAGAGGACACAATGTAG                                           11058
 3681 -R--E--D--T--M--*-                                           11058
```

FIG. 19K

```
   1 gctgccttgatatacacttttcaaaATGCTTTGGTGGGAAGAAGTAGAGGACTGTT      31

     gtaagtacaaagtaactaaaaatat.........tttatttttttattttgcattttag

  32 ATGAAAGAGAAGATGTTCAAAAGAAAACATTCACAAAATGGGTAAATGCACAATTTTCTA    91
  92 AG                                                             93

     gtaagaatggtttgttactttactt.........ttgagtgtatttttttaatttcag

  94 TTTGGGAAGCAGCATATTGAGAACCTCTTCAGTGACCTACAGGATGGGAGGCGCCTCCTA   153

 154 GACCTCCTCGAAGGCCTGACAGGGCAAAAACTG                             186

     gtatgtgacttatttttaagaaagt.........gaacactcttttgttttgttctcag

 187 CCAAAAGAAAAAGGATCCACAAGAGTTCATGCCCTGAACAATGTCAACAAGGCACTGCGG   246
 247 GTTTTGCAGAACAATAAT                                           264

     gtaagtagtaccctggacaaggtct.........aatgttttacccctttctttaacag

 265 GTTGATTTAGTGAATATTGGAAGTACTGACATCGTAGATGGAAATCATAAACTGACTCTT   324
 325 GGTTTGATTTGGAATATAATCCTCCACTGGCAG                            357

     gtaagaatcctgatgaatggtttcc.........tatgaaaatttatttccacatgtag

 358 GTCAAAAATGTAATGAAAAATATCATGGCTGGATTGCAACAAACCAACAGTGAAAAGATT   417
 418 CTCCTGAGCTGGGTCCGACAATCAACTCGTAATTATCCACAGGTTAATGTAATCAACTTC   477
 478 ACCACCAGCTGGTCTGATGGCCTGGCTTTGAATGCTCTCATCCATAGTCATAG         530

     gtaagaagattactgagacattaaa.........atgtgtgtatgtgtatgtgttttag

 531 GCCAGACCTATTTGACTGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACAACGACTGGA   590

 591 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAAACTACTCGATCCTGAAG    649

     gttggtaaatttctggactaccact.........atgtgtagtgttaatgtgcttacag

 650 ATGTTGATACCACCTATCCAGATAAGAAGTCCATCTTAATGTACATCACATCACTCTTCC    709
 710 AAGTTTTGCCTCAACAAGTGAGCATTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGC   769
 770 CACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAAATGCACTATTCTCAAC   829
 830 AG                                                            831

     gtaaagtgtgtaaaggacagctact.........cactcccccaaaccccttctctgcag

 832 ATCACGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCCTAAGCCTCGATTCAAG   891
 892 AGCTATGCCTACACACAGGCTGCTTATGTCACCACCTCTGACCCTACACGGAGCCCATTT   951
 952 CCTTCACAG                                                     960

     gtctgtcaacatttactctctgttg.........acacccaatttattttattgtgcag

 961 CATTTGGAAGCTCCTGAAGACAAGTCATTTGGCAGTTCATTGATGGAGAGTGAAGTAAAC  1020
1021 CTGGACCGTTATCAAACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTTCTGCTGAGGAC  1080
1081 ACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAAGACCAGTTTCAT  1140
1141 ACTCATGAG                                                     1149

     gtaaactaaaacgttaatttacaaa.........aattgttaacttccttctttgtcag

1150 GGGTACATGATGGATTTGACAGCCCATCAGGGCCGGGTTGGTAATATTCTACAATTGGGA  1209
1210 AGTAAGCTGATTGGAACAGGAAAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAG  1269
1270 ATGAATCTCCTAAATTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGC  1329
1330 AA                                                            1331

     gtaagtccttatttgtttttaatta.........taataggcttctttcaaattttcag

1332 TTTACATAGAGTTTTAATGGATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCT  1391
1392 AACAAAAACAGAAGAAAGAACAAGGAAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGA  1451
```

FIG. 20A

```
1452 AGACCTAAAACGCCAAGTACAACAACATAAG                                  1482

     gtaggtgtatcttatgttgcgtgct..........tcctttaaaacattttatctttcag

 1483 GTGCTTCAAGAAGATCTAGAACAAGAACAAGTCAGGGTCAATTCTCTCACTCACATGGTG  1542
 1543 GTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAG  1602

     gtcagattattttgcttagtaaact..........ctgtgcttgattgtctcttctccag

 1603 GTATTGGGAGATCGATGGGCAAACATCTGTAGATGGACAGAAGACCGCTGGGTTCTTTTA  1662
 1663 CAAGACATCCTTCTCAAATGGCAACGTCTTACTGAAGAACAG                    1704

     gtgtgtcatgtgtgagaaactagct..........cttggaattctttaatgtcttgcag

 1705 TGCCTTTTTAGTGCATGGCTTTCAGAAAAAGAAGATGCAGTGAACAAGATTCACACAACT  1764
 1765 GGCTTTAAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACTGGCC              1812

     gtatgtactttctagctttcaatgg..........tctgtgatctttcttgttttaacag

 1813 GTTTTAAAAGCGGATCTAGAAAAGAAAAAGCAATCCATGGGCAAACTGTATTCACTCAAA  1872
 1873 CAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGAAGACGGAAGCATGGCTG  1932
 1933 GATAACTTTGCCCCGGTGTTGGGATAATTTAGTCCAAAAACTTGAAAAGAGTACAGCACAG  1992

     gttagtgataccaattatcatgcta..........acctctgtttcaatacttctcacag

 1993 ATTTCACAGGCTGTCACCACCACTCAGCCATCACTAACACAGACAACTGTAATGGAAACA  2052
 2053 GTAACTACGGTGACCACAAGGGAACAGATCCTGGTAAAGCATGCTCAAGAGGAACTTCCA  2112
 2113 CCACCACCTCCCCAAAAGAAGAGGCAGATTACTGTGGATTCTGAAATTAGGAAAAG      2168

     gtgagagcatcttaagctttatct..........tgacttttatttttgctgtcttag

 2169 GTTGGATGTTGATATAACTGAACTTCACAGCTGGATTACTCGCTCAGAAGCTGTGTTGCA  2228
 2229 GAGTCCTGAATTTGCAATCTTTCGGAAGGAAGGCAACTTCTCAGACTTAAAAGAAAAAGT  2288
 2289 CAAT                                                          2292

     gtaggttatgcattaattttatat..........actcatctttgctctcatgctgcag

 2293 GCCATAGAGCGAGAAAAAGCTGAGAAGTTCAGAAAACTGCAAGATGCCAGCAGATCAGCT  2352
 2353 CAGGCCCTGGTGGAACAGATGGTGAATG                                  2380

     gtaattacacgagttgatttagata..........tatttaattatttttttctttctag

 2381 AGGGTGTTAATGCAGATAGCATCAAACAAGCCTCAGAACAACTGAACAGCCGGTGGATCG  2440
 2441 AATTCTGCCAGTTGCTAAGTGAGAGACTTAACTGGCTGGAGTATCAGAACAACATCATCG  2500
 2501 CTTTCTATAATCAGCTACAACAATTGGAGCAGATGACAACTACTGCTGAAAACTGGTTGA  2560
 2561 AAATCCAACCCACCACCCCATCAGAGCCAACAGCAATTAAAAGTCAGTTAAAAATTTGTA  2620
 2621 AG                                                            2622

     gtaagaatctcttctccttccattt..........ttactttccatactctatggcacag

 2623 GATGAAGTCAACCGGCTATCAGATCTTCAACCTCAAATTGAACGATTAAAAATTCAAAGC  2682
 2683 ATAGCCCTGAAAGAGAAAGGACAAGGACCCATGTTCCTGGATGCAGACTTTGTGGCCTTT  2742
 2743 ACAAATCATTTTAAGCAAGTCTTTTCTGATGTGCAGGCCAGAGAGAAAGAGCTACAGACA  2802
 2803 A                                                             2803

     gtaagtaaaaagcctaaaatggcta..........cattctttttccctttgataaag

 2804 TTTTTGACACTTTGCCACCAATGCGCTATCAGGAGACCATGAGTGCCATCAGGACATGGG  2863
 2864 TCCAGCAGTCAGAAACCAAACTCTCCATACCTCAACTTAGTGTCACCGACTATGAAATCA  2923
 2924 TGGAGCAGAGACTCGGGGAATTGCAG                                    2949

     gtctgtgaatatttgaatgtcaaaa..........atgtatttaaaaaattgtttttag

 2950 GCTTTACAAAGTTCTCTGCAAGAGCAACAAAGTGGCCTATACTATCTCAGCACCACTGTG  3009
 3010 AAAGAGATGTCGAAGAAAGCGCCCTCTGAAATTAGCCGGAAATATCAATCAGAATTTGAA  3069
 3070 GAAATTGAGGGACGCTGGAAGAAGCTCTCCTCCCAGCTGGTTGAGCATTGTCAAAAGCTA  3129
 3130 GAGGAGCAAATGAATAAACTCCGAAAAATTCAG                             3162
```

FIG. 20B

```
     gtaattcaagattttactttctacc.........tgccttataacgggtctcgtttcag

3163 AATCACATACAAACCCTGAAGAAATGGATGGCTGAAGTTGATGTTTTTCTGAAGGAGGAA  3222
3223 TGGCCTGCCCTTGGGGATTCAGAAATTCTAAAAAAGCAGCTGAAACAGTGCAGA        3276

     gtaagattttatatgatgccttta.........ggcttaaattgatttattttcttag

3277 CTTTTAGTCAGTGATATTCAGACAATTCAGCCCAGTCTAAACAGTGTCAATGAAGGTGGG  3336
3337 CAGAAGATAAAGAATGAAGCAGAGCCAGAGTTTGCTTCGAGACTTGAGACAGAACTCAAA  3396
3397 GAACTTAACACTCAGTGGGATCACATGTGCCAACAG                          3432

     gtatagacaatctctttcactgtgg.........gttttgtttgtttgttttgtggaag

3433 GTCTATGCCAGAAAGGAGGCCTTGAAGGGAGGTTTGGAGAAAACTGTAAGCCTCCAGAAA  3492
3493 GATCTATCAGAGATGCACGAATGGATGACACAAGCTGAAGAAGAGTATCTTGAGAGAGAT  3552
3553 TTTGAATATAAAACTCCAGATGAATTACAGAAAGCAGTTGAAGAGATGAAG           3603

     gtaaaaaaaaaaaaagaaaaactaa.........taagagagcattctttatttttcag

3604 AGAGCTAAAGAAGAGGCCCAACAAAAAGAAGCGAAAGTGAAACTCCTTACTGAGTCTGTA  3663
3664 AATAGTGTCATAGCTCAAGCTCCACCTGTAGCACAAGAGGCCTTAAAAAAGGAACTTGAA  3723
3724 ACTCTAACCACCAACTACCAGTGGCTCTGCACTAGGCTGAATGGGAAATGCAAGACTTTG  3783
3784 GAA                                                           3786

     gtcagttgcttttcttggtctttgt.........tctgtgatatatatttctttcttag

3787 GAAGTTTGGGCATGTTGGCATGAGTTATTGTCATACTTGGAGAAAGCAAACAAGTGGCTA  3846
3847 AATGAAGTAGAATTTAAACTTAAAACCACTGAAAACATTCCTGGCGGAGCTGAGGAAATC  3906
3907 TCTGAGGTGCTAGAT                                               3921

     gtaagttgtaaattaagccaaatga.........agtaattattgcaaatgtgtttcag

3922 TCACTTGAAAATTTGATGCGACATTCAGAGGATAACCCAAATCAGATTCGCATATTGGCA  3981
3982 CAGACCCTAACAGATGGCGGAGTCATGGATGAGCTAATCAATGAGGAACTTGAGACATTT  4041
4042 AATTCTCGTTGGAGGGAACTACATGAAGAG                                4071

     gtatgaagataagtgaaaaatctct.........atacactcttattccttctttttag

4072 GCTGTAAGGAGGCAAAAGTTGCTTGAACAGAGCATCCAGTCTGCCCAGGAGACTGAAAAA  4131
4132 TCCTTACACTTAATCCAGGAGTCCCTCACATTCATTGACAAGCAGTTGGCAGCTTATATT  4191
4192 GCAGACAAGGTGGACGCAGCTCAAATGCCTCAGGAAGCCCAG                    4233

     gcaagtacatctgggaatcagcttc.........actaataatgctatcctcccaacag

4234 AAAATCCAATCTGATTTGACAAGTCATGAGATCAGTTTAGAAGAAATGAAGAAACATAAT  4293
4294 CAGGGGAAGGAGGCTGCCCAAAGAGTCCTGTCTCAGATTGATGTTGCACAG          4344

     gtatatgttatttcagaaactaagg.........gtgccttttacactgtccttacag

4345 AAAAAATTACAAGATGTCTCCATGAAGTTTCGATTATTCCAGAAACCAGCCAATTTTGAG  4404
4405 CAGCGTCTACAAGAAAGTAAGATGATTTTAGATGAAGTGAAGATGCACTTGCCTGCATTG  4464
4465 GAAACAAAGAGTGTGGAACAGGAAGTAGTACAGTCACAGCTAAATCATTGTGTG        4518

     gtatgtatttctggtggcaaatacg.........tgttttgttttatgtttaaacttag

4519 AACTTGTATAAAAGTCTGAGTGAAGTGAAGTCTGAAGTGGAAATGGTGATAAAGACTGGA  4578
4579 CGTCAGATTGTACAGAAAAAGCAGACGGAAAATCCCAAAGAACTTGATGAAAGAGTAACA  4638
4639 GCTTTGAAATTGCATTATAATGAGCTGGGAGCAAAG                          4674

     gtgtgtgcatgctgagaccacaaac.........tacatttcattataattcttttcag

4675 GTAACAGAAAGAAAGCAACAGTTGGAGAAATGCTTGAAATTGTCCCGTAAGATGCGAAAG  4734
4735 GAAATGAATGTCTTGACAGAATGGCTGGCAGCTACAGATATGGAATTGACAAAGAGATCA  4794
4795 GCAGTTGAAGGAATGCCTAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAG           4845

     gtaaaacctatatcactgaaggtta.........aaggtcaatgctctcctttcacag

4846 GCTACTCAAAAAGAGATTGAGAAACAGAAGGTGCACCTGAAGAGTATCACAGAGGTAGGA  4905
```

FIG. 20C

```
4906 GAGGCCTTGAAAACAGTTTTGGGCAAGAAGGAGACGTTGGTGGAAGATAAACTCAGTCTT  4965
4966 CTGAATAGTAACTGGATAGCTGTCACCTCCCGAGCAGAAGAGTGGTTAAATCTTTTGTTG  5025

     gtaagagaaaaggctagaagctttt.........catggtatgtctctgtacaattaag

5026 GAATACCAGAAACACATGGAAACTTTTGACCAGAATGTGGACCACATCACAAAGTGGATC  5085
5086 ATTCAGGCTGACACACTTTTGGATGAATCAGAGAAAAAGAAACCCCAGCAAAAAGAAGAC  5145
5146 GTGCTTAAG                                                     5154

     gtagcaaataaaatatgaaaagtaa.........cctatctcttgctcatggaatatag

5155 CGTTTAAAGGCAGAACTGAATGACATACGCCCAAAGGTGGACTCTACACGTGACCAAGCA  5214
5215 GCAAACTTGATGGCAAACCGCGGTGACCACTGCAGGAAATTAGTAGAGCCCCAAATCTCA  5274
5275 GAGCTCAACCATCGATTTGCAGCCATTTCACACAGAATTAAGACTGGAAAG           5325

     gtaggaagatctactccaaggtgga.........aaagtagcactatctttttttttag

5326 GCCTCCATTCCTTTGAAGGAATTGGAGCAGTTTAACTCAGATATACAAAAATTGCTTGAA  5385
5386 CCACTGGAGGCTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAAGACTTCAATAAAGAT  5445
5446 ATG                                                           5448

     gtaaattggttgtgataaaagtgtg.........gactgtacttgttgttttgatcag

5449 AATGAAGACAATGAGGGTACTGTAAAAGAATTGTTGCAAAGAGGAGACAACTTACAACAA  5508
5509 AGAATCACAGATGAGAGAAAGCGAGAGGAAATAAAGATAAAACAGCAGCTGTTACAGACA  5568
5569 AAACATAATGCTCTCAAG                                            5586

     gtattagagctaaaattataatata.........ttaataatgtctgcaccatgaacag

5587 GATTTGAGGTCTCAAAGAAGAAAAAAGGCTCTAGAAATTTCTCATCAGTGGTATCAGTAC  5646
5647 AAGAGGCAGGCTGATGATCTCCTGAAATGCTTGGATGACATTGAAAAAAAATTAGCCAGC  5706
5707 CTACCTGAGCCCAGAGATGAAAGGAAAATAAAG                             5739

     gtaatgttgttttagaatgtcaata.........gccctgtattggttttgctcaatag

5740 GAAATTGATCGGGAATTGCAGAAGAAGAAAGAGGAGCTGAATGCAGTGCGTAGGCAAGCT  5799
5800 GAGGGCTTGTCTGAGGATGGGGCCGCAATGGCAGTGGAGCCAACTCAGATCCAGCTCAGC  5859
5860 AAGCGCTGGCGGGAAATTGAGAGCAAATTTGCTCAGTTTCGAAGACTCAACTTTGCACAA  5919
5920 ATT                                                           5922

     gtgagttgttactggcaaacccacg.........ttgttcttttgtatatctataccag

5923 CACACTGTCCGTGAAGAAACGATGATGGTGATGACTGAAGACATGCCTTTGGAAATTTCT  5982
5983 TATGTGCCTTCTACTTATTTGACTGAAATCACTCATGTCTCACAAGCCCTATTAGAAGTG  6042
6043 GAACAACTTCTCAATGCTCCTGACCTCTGTGCTAAGGACTTTGAAGATCTCTTTAAGCAA  6102
6103 GAGGAGTCTCTGAAG                                               6117

     gtaaaaccaaagcactttcattcgt.........ctgttttaaaatttttatattacag

6118 AATATAAAAGATAGTCTACAACAAAGCTCAGGTCGGATTGACATTATTCATAGCAAGAAG  6177
6178 ACAGCAGCATTGCAAAGTGCAACGCCTGTGGAAAGGGTGAAGCTACAGGAAGCTCTCTCC  6237
6238 CAGCTTGATTTCCAATGGGAAAAAGTTAACAAAATGTACAAGGACCGACAAGG         6290

     gtaggtaacacatatatttttcttg.........ttgatccatatgcttttacctgcag

6291 GCGATTTGACAGATCTGTTGAGAAATGGCGGCGTTTTCATTATGATATAAAGATATTTAA  6350
6351 TCAGTGGCTAACAGAAGCTGAACAGTTTCTCAGAAAGACACAAATTCCTGAGAATTGGGA  6410
6411 ACATGCTAAATACAAATGGTATCTTAAG                                  6438

     gtaagtctttgatttgttttttcga.........gttttgccttttttggtatcttacag

6439 GAACTCCAGGATGGCATTGGGCAGCGGCAAACTGTTGTCAGAACATTGAATGCAACTGGG  6498
6499 GAAGAAATAATTCAGCAATCCTCAAAAACAGATGCCAGTATTCTACAGGAAAAATTGGGA  6558
6559 AGCCTGAATCTGCGGTGGCAGGAGGTCTGCAAACAGCTGTCAGACAGAAAAAAGAG      6614

     gtagggcgacagatctaataggaat.........aacaattttattcttctttctccag

6615 GCTAGAAGAACAAAAGAATATCTTGTCAGAATTTCAAAGAGATTTAAATGAATTTGTTTT  6674
```

FIG. 20D

```
6675 ATGGTTGGAGGAAGCAGATAACATTGCTAGTATCCCACTTGAACCTGGAAAAGAGCAGCA  6734
 6735 ACTAAAAGAAAAGCTTGAGCAAGTCAAG                                  6762

      gtaattttatttctcaaatccccc.....acgttgttgcatttgtctgtttcag

 6763 TTACTGGTGGAAGAGTTGCCCCTGCGCCAGGGAATTCTCAAACAATTAAATGAAACTGGA  6822
 6823 GGACCCGTGCTTGTAAGTGCTCCCATAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAG  6882
 6883 CTCAAGCAGACAAATCTCCAGTGGATAAAG                                  6912

      gttagacattaaccatctcttccgt.....tttttaaaatgtattttcctttcag

 6913 GTTTCCAGAGCTTTACCTGAGAAACAAGGAGAAATTGAAGCTCAAATAAAAGACCTTGGG  6972
 6973 CAGCTTGAAAAAAAGCTTGAAGACCTTGAAGAGCAGTTAAATCATCTGCTGCTGTGGTTA  7032
 7033 TCTCCTATTAGGAATCAGTTGGAAATTTATAACCAACCAAACCAAGAAGGACCATTTGAC  7092
 7093 GTTAAG                                                        7098

      gtagggaacttttgctttaaatat.....gcactatatgggttcttttccccag

 7099 GAAACTGAAATAGCAGTTCAAGCTAAACAACCGGATGTGGAAGAGATTTTGTCTAAAGGG  7158
 7159 CAGCATTTGTACAAGGAAAAACCAGCCACTCAGCCAGTGAAG                    7200

      gtaatgaagcaacctctagcaatat.....taatgtgtatgcttttctgttaaag

 7201 AGGAAGTTAGAAGATCTGAGCTCTGAGTGGAAGGCGGTAAACCGTTTACTTCAAGAGCTG  7260
 7261 AGGGCAAAGCAGCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCT            7309

      gtaagtatactggatcccattctct.....tttgcaaaaacccaaaatattttag

 7310 CTCCTACTCAGACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCT  7369
 7370 CCAAACTAGAAATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCTGGCAGATTTCAACC  7429
 7430 GGGCTTGGACAGAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATCACAGA  7489
 7490 GGGTGATGGTGGGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAG          7542

      gtatgagaaaaaatgataaaagttg.....tactaagggatatttgttcttacag

 7543 GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCC  7602
 7603 CAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA   7660

      gtaagtttttaacaagcatgggac.....atatttatttttccttttattctag

 7661 TTGAAAGAATTCAGAATCAGTGGGATGAAGTACAAGAACACCTTCAGAACCGGAGGCAAC  7720
 7721 AGTTGAATGAAATGTTAAAGGATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGC  7780
 7781 AGGTCTTAGGACAGGCCAGAGCCAAGCTTGAGTCATGGAAGGAGGGTCCCTATACAGTAG  7840
 7841 ATGCAATCCAAAAGAAAATCACAGAAACCAAG                              7872

      gttagtatcaaagatacctttttaa.....ttctctttctcataaaaatctatag

 7873 CAGTTGGCCAAAGACCTCCGCCAGTGGCAGACAAATGTAGATGTGGCAAATGACTTGGCC  7932
 7933 CTGAAACTTCTCCGGGATTATTCTGCAGATGATACCAGAAAAGTCCACATGATAACAGAG  7992
 7993 AATATCAATGCCTCTTGGAGAAGCATTCATAAAAG                            8027

      gtatgaattacattatttctaaaac.....catctgaacatttggtcctttgcag

 8028 GGTGAGTGAGCGAGAGGCTGCTTTGGAAGAAACTCATAGATTACTGCAACAGTTCCCCCT  8087
 8088 GGACCTGGAAAAGTTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCCTACA  8147
 8148 GGATGCTACCCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAA  8207
 8208 ACAATGGCAA                                                    8217

      gtaagtcaggcatttccgctttagc.....tattcttcttcctgctgtcctgtag

 8218 GACCTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAAACAGC  8277
 8278 CAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCTGTTACAAAGACGTTTG  8337
 8338 GATAACATGAACTTCAAGTGGAGTGAACTTCGGAAAAAGTCTCTCAACATTAG         8390

      gtaggaaaagatgtggagcaaaaag.....atggtacgctgctgttctttttcag

 8391 GTCCCATTTGGAAGCCAGTTCTGACCAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACT  8450
 8451 TCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGA  8510
```

FIG. 20E

```
8511 CTTTCCAGCAGTTCAGAAGCAGAACGATGTACATAGG                          8547

     gtaggacatttttaagcctcgtgcc.....cacttcttttcatctcatttcacag

 8548 GCCTTCAAGAGGGAATTGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTA  8607
 8608 CGAATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAGA  8667
 8668 G                                                             8668

     gtaattgaatgtggaactataataa.....aaaccttgtcatattgccaatttag

 8669 AGCTGCCTCCTGAGGAGAGAGCCCAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGG  8728
 8729 AGGTCAATACTGAGTGGGAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAG  8788
 8789 ATGAGACCCTTGAAAGACTCCGGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGC  8848
 8849 TGCGCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATCTCCTCATTGACT  8908
 8909 CTCTCCAAGATCACCTCGAGAAAGTCAAG                                 8937

     gtaccgtctacttctttgcttcagg.....atttgcttttgactattgcacacag

 8938 GCACTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCT  8997
 8998 CGCCAGCTTACCACTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGAC  9057
 9058 CTGAACACCAGATGGAAGCTTCTGCAG                                  9084

     gtaagcacattgtaaacattgttgt.....atcatttctctccttttcctcccag

 9085 GTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTCCAGCA  9144
 9145 TCTCAGCACTTTCTTTCCA                                         9163

     gtaagtcattttcagcttttatcac.....tctttttttcctcccttctttcag

 9164 CGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCA  9223
 9224 A                                                             9224

     gtaagttggaagtatcacattttta.....tctttctttatgttttgtgttttag

 9225 CCACGAGACTCAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT  9284
 9285 AG                                                            9286

     gtaaggacatggccatgtttcctcc.....tgctctttgttttccctcttttcag

 9287 CTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGC  9346
 9347 AGAAGGCCCTTTGCT                                              9361

     gtaagtattggccagtatttgaaga.....ttgtgattttatttgtttttgcag

 9362 TGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGACCAGCACAACCTCAAGCAAA  9421
 9422 ATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGCC  9481
 9482 TGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAACT  9541
 9542 GGCTGCTGAATGTTTATGATAC                                       9563

     gtacgtatggcatgttttatttcc.....tttctgctttgattcttcataatag

 9564 GGGACGAACAGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAA  9623
 9624 AGCACATTTGGAAGACAAGTACAGAT                                   9649

     gtaagtcgtgtatattaatgctgta.....ttgcaattttcttcttcctttgtag

 9650 ACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCC  9709
 9710 TTCTGCATGATTCTATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCA  9769
 9770 GTAACATTGAGCCAAGTGTCCGGAGCTGCTTCCAATTT                       9807

     gtaagttattcaccttctaggtaac.....ttctctctccctcctgtctttgcag

 9808 GCTAATAATAAGCCAGAGATCGAAGCGGCCCTCTTCCTAGACTGGATGAGACTGGAACCC  9867
 9868 CAGTCCATGGTGTGGCTGCCCGTCCTGCACAGAGTGGCTGCTGCAGAAACTGCCAAGCAT  9927
 9928 CAGGCCAAATGTAACATCTGCAAAGAGTGTCCAATCATTGGATTCAG               9974

     gtattaggaaccaaaaaaaaaatgt.....cgtgtttgtttttgctctttatcag
```

FIG. 20F

```
 9975 GTACAGGAGTCTAAAGCACTTTAATTATGACATCTGCCAAAGCTGCTTTTTTTCTGGTCG 10034
10035 AGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATATTGCACTCCG          10086

      gtaagtttgacgccagcctgacgtg.........gatctcaccatgatctccctttag

10087 ACTACATCAGGAGAAGATGTTCGAGACTTTGCCAAGGTACTAAAAAACAAATTTCGAACC 10146
10147 AAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTACCTGCCAGTGCAGACTGTCTTAGAG 10206
10207 GGGGACAACATGGAAAC                                            10223

      gtgagtagtagcaaaagcagaacac.........caccacctcatttttgttttgcag

10224 TCCCGTTACTCTGATCAACTTCTGGCCAGTAGATTCTGC                      10262

      gtgagtacttttttgctgaagggt.........actaatcacattttctgccttatag

10263 GCCTGCCTCGTCCCCTCAGCTTTCACACGATGATACTCATTCACGCATTGAACATTATGC 10322
10323 TAGCAG                                                      10328

      gtatgagactagttgtatgccaggc.........aatgagctttacgtttttatcag

10329 GCTAGCAGAAATGGAAAACAGCAATGGATCTTATCTAAATGATAGCATCTCTCCTAATGA 10388
10389 GAGCAT                                                      10394

      gtaagtatcccatctcttttacaa.........caaaacctttgattttatttccag

10395 AGATGATGAACATTTGTTAATCCAGCATTACTGCCAAAGTTTGAACCAGGACTCCCCCCT 10454
10455 GAGCCAGCCTCGTAGTCCTGCCCAGATCTTGATTTCCTTAGAGAGTGAGGAAAGAGGGGA 10514
10515 GCTAGAGAGAATCCTAGCAGATCTTGAGGAAGAAAACAG                      10553

      gtgagttttctttctagctttgtca.........ttttttacttttttgatgccaatag

10554 GAATCTGCAAGCAGAATATGACCGTCTAAAGCAGCAGCACGAACATAAAGGCCTGTCCCC 10613
10614 ACTGCCGTCCCCTCCTGAAATGATGCCCACCTCTCCCCAGAGTCCCCGGGATGCTGAGCT 10673
10674 CATTGCTGAGGCCAAGCTACTGCGTCAACACAAAGGCCGCCTGGAAGCCAGGATGCAAAT 10733
10734 CCTGGAAGACCACAATAAACAGCTGGAGTCACAGTTACACAGGCTAAGGCAGCTGCTGGA 10793
10794 GCAA                                                        10797

      gtgaggagagagatgggatttttac.........ttctgttttcttttggatgacttag

10798 CCCCAGGCAGAGGCCAAAGTGAATGGCACAACGGTGTCCTCTCCTTCTACCTCTCTACAG 10857
10858 AGGTCCGACAGCAGTCAGCCTATGCTGCTCCGAGTGGTTGGCAGTCAAACTTCGGACTCC 10917
10918 ATGG                                                        10921

      gtaagtgtcctagctactctcagat.........attatttgtttttgcttttattaag

10922 GTGAGGAAGATCTTCTCAGTCCTCCCCAGGACACAAGCACAGGGTTAGAGGAGGTGATGG 10981
10982 AGCAACTCAACAACTCCTTCCCTAGTTCAAGAG                          11014

      gtaagctccaatacctagaagggac.........cctcttcctctctctattattaaag

11015 GAAGAAATACCCCTGGAAAGCCAATGAGAGAG                          11046

      gttagtgagattcaggctcacggcc.........gtctttctttctctttgttttccag

11047 GACACAATGTAG                                             11058
```

FIG. 20G

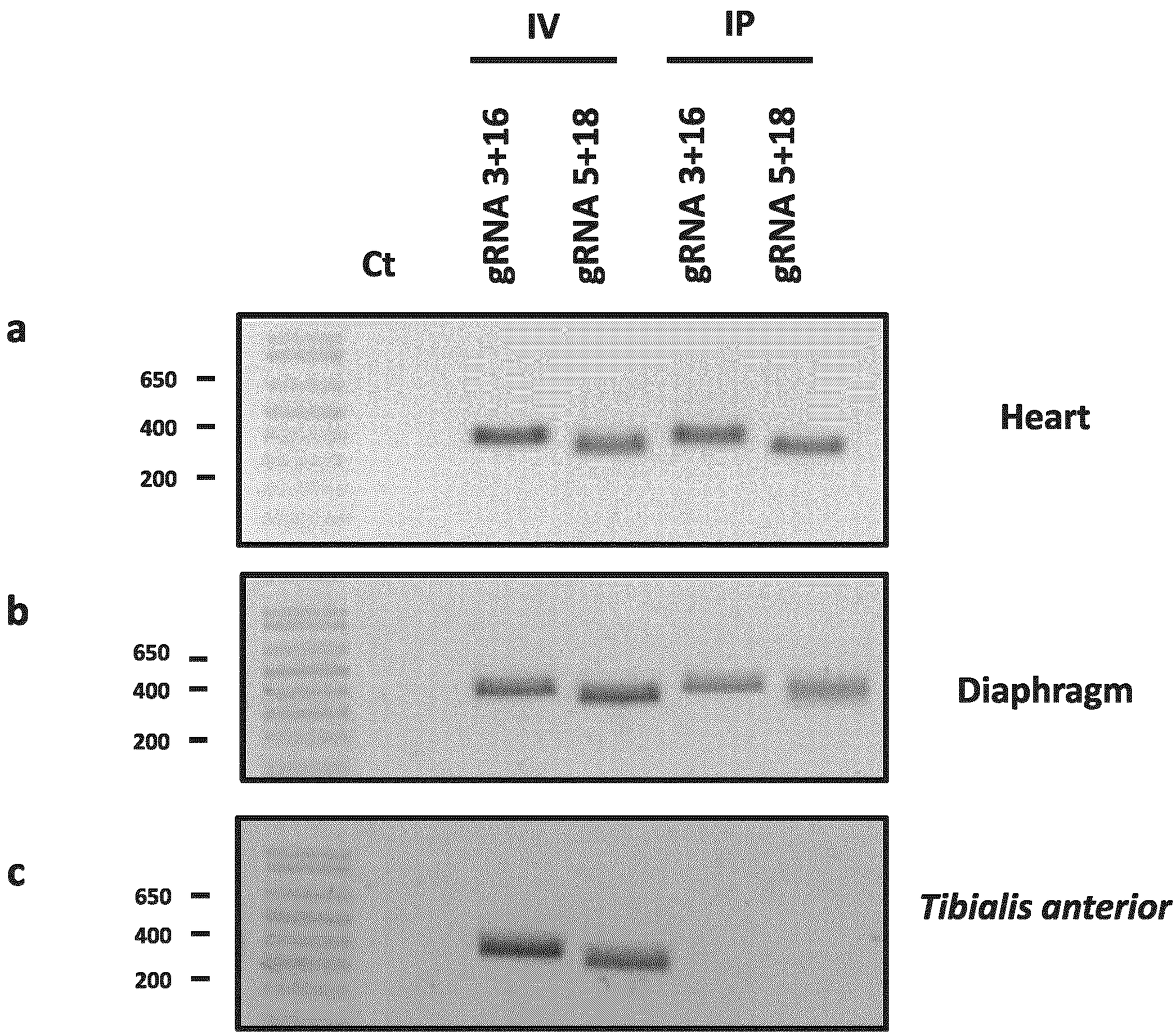

IV
IP
Ct
BRN
BRN
BRN
BRN
A 3+16
A 5+18
A 3+16
A 5+18
a
650
400
200
Heart
b
650
400
200
Diaphragm
c
650
400
200
Tibialis anterior

FIG. 21

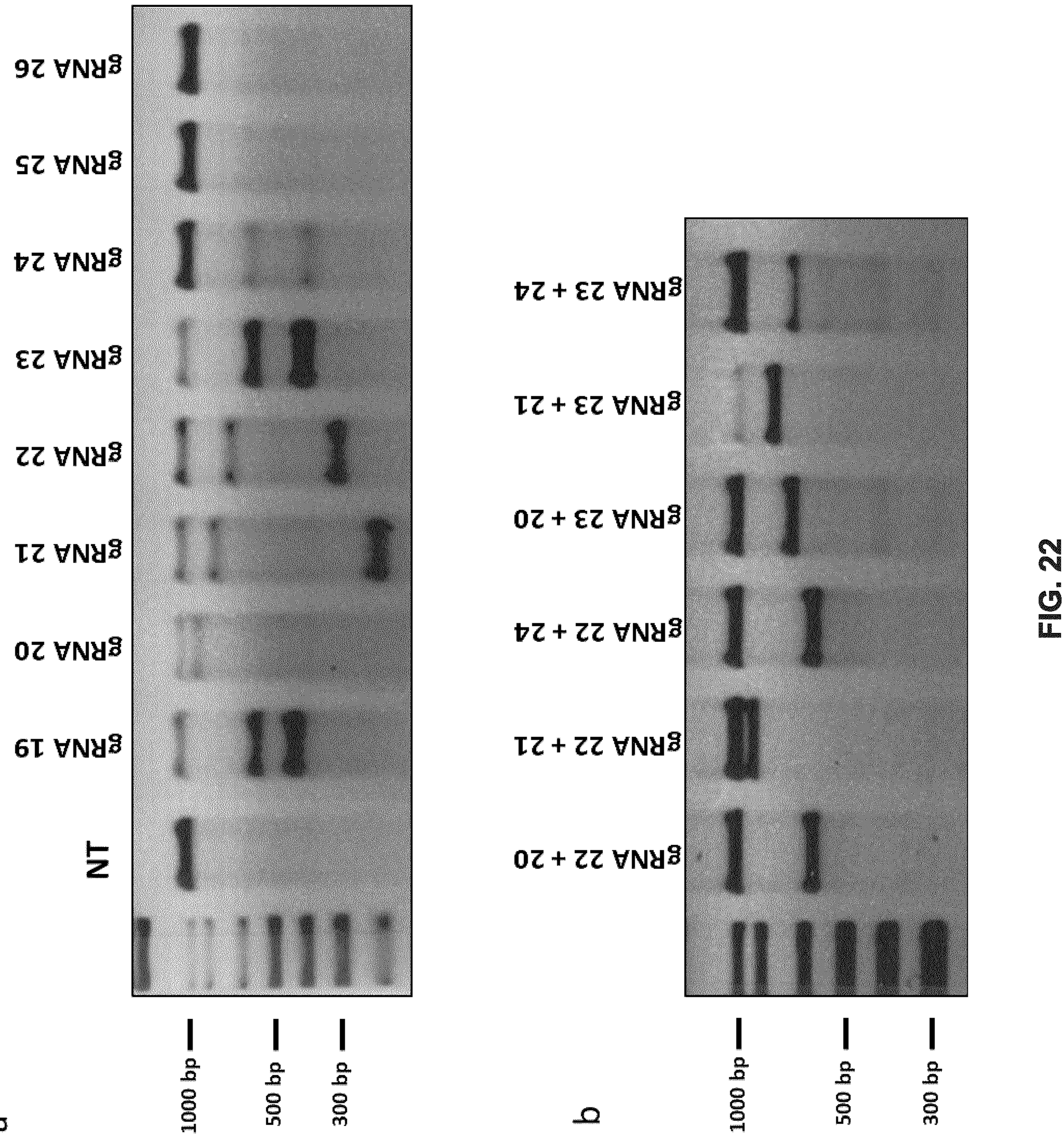

a
1000 bp
500 bp
300 bp
NT
gRNA 19
gRNA 20
gRNA 21
gRNA 22
gRNA 23
gRNA 24
gRNA 25
gRNA 26
b
1000 bp
500 bp
300 bp
gRNA 22 + 20
gRNA 22 + 21
gRNA 22 + 24
gRNA 23 + 20
gRNA 23 + 21
gRNA 23 + 24

FIG. 22

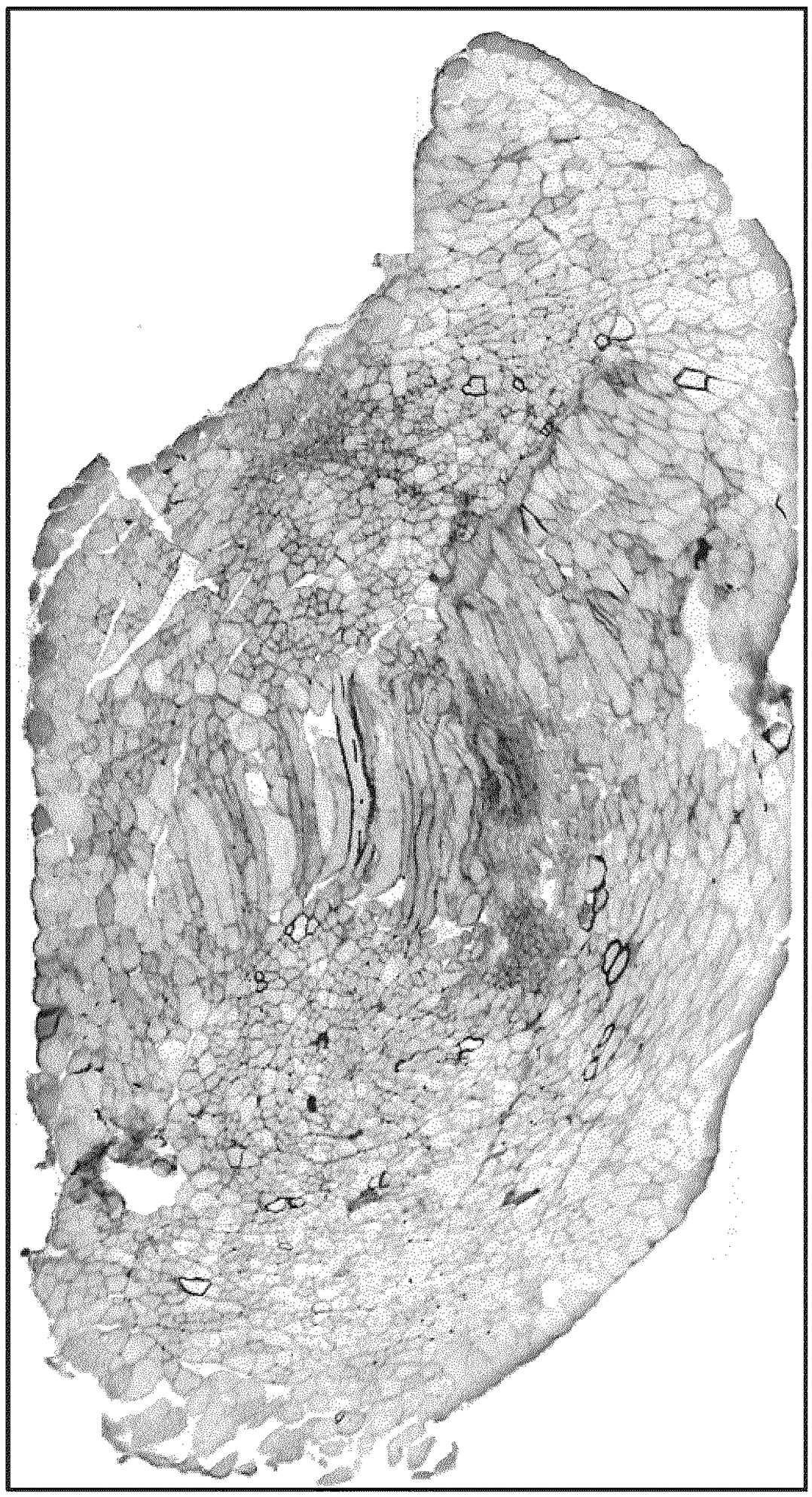


FIG. 23a



FIG. 23b



FIG. 23c

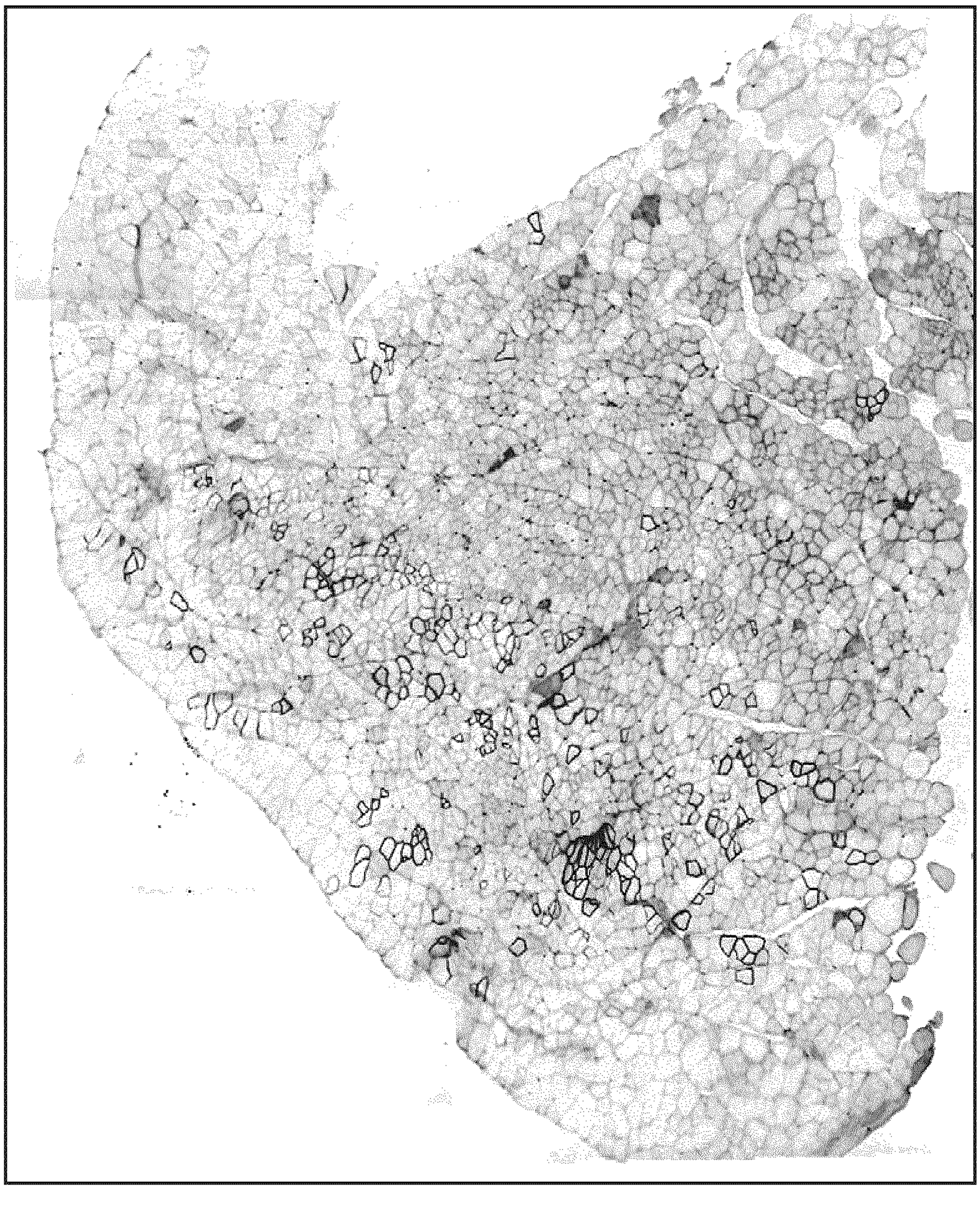


FIG. 23d



FIG. 23e

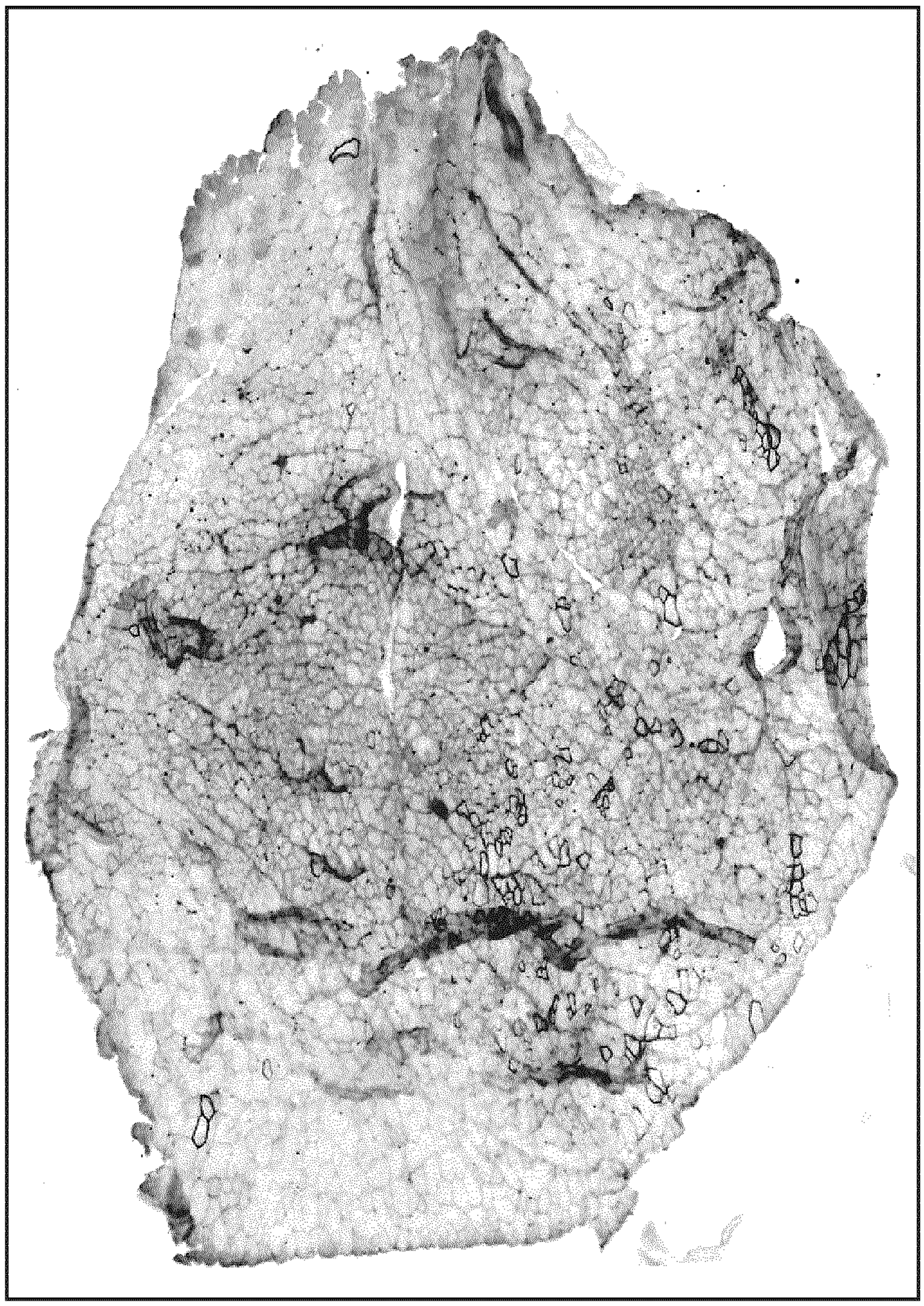


FIG. 23f



FIG. 23g

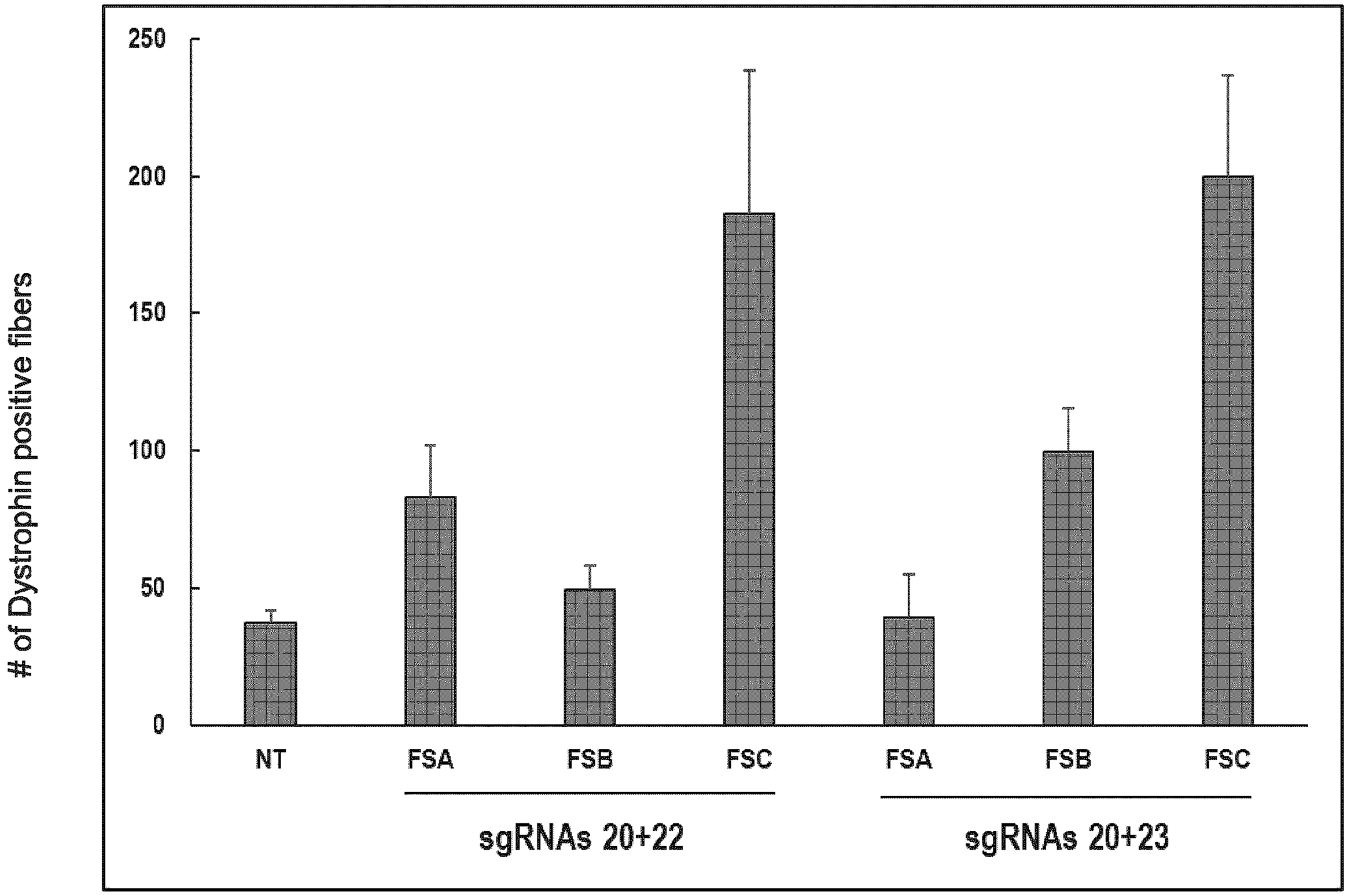
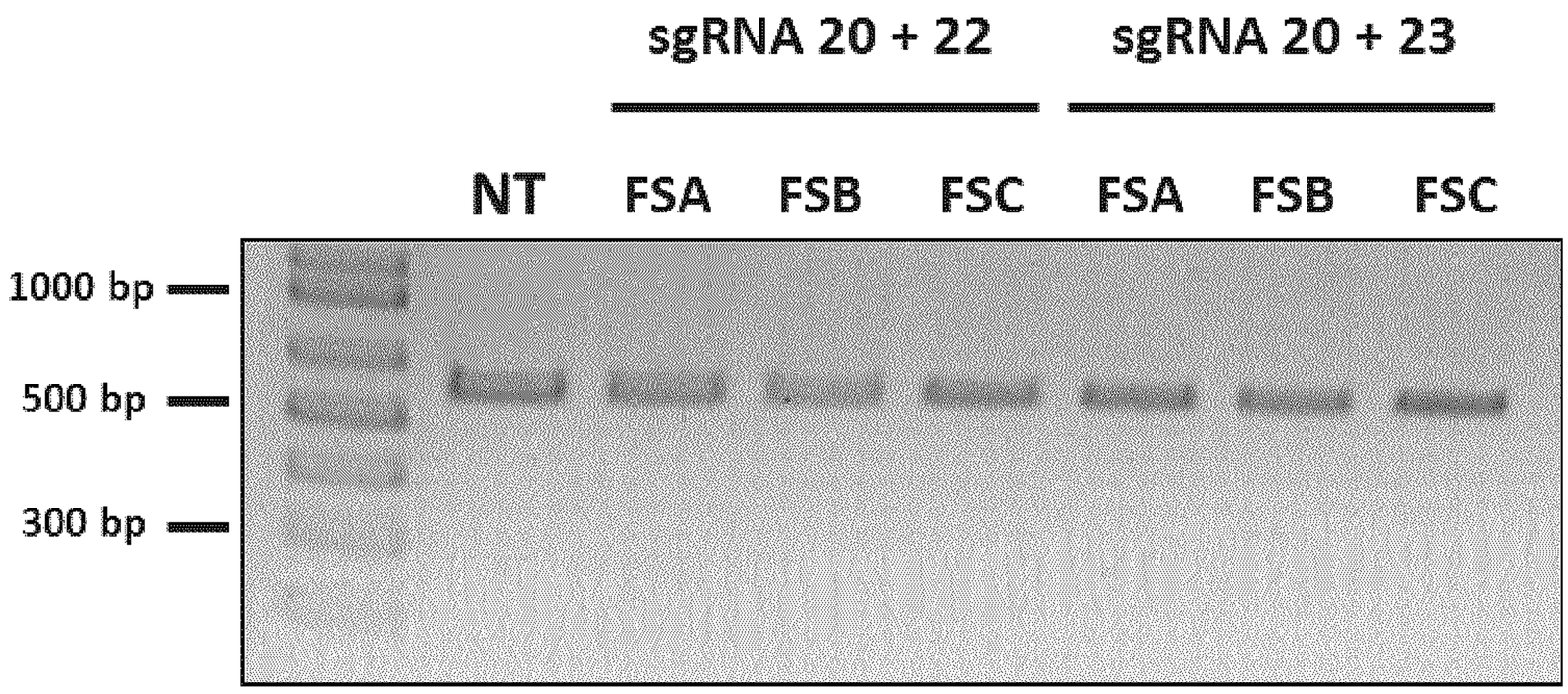

# of Dystrophin positive fibers
250
200
150
100
50
0
NT
FSA
FSB
FSC
FSA
FSB
FSC
sgRNAs 20+22
sgRNAs 20+23

FIG. 24

a


sgRNA 20 + 22
sgRNA 20 + 23
NT
FSA
FSB
FSC
FSA
FSB
FSC
1000 bp
500 bp
300 bp

b

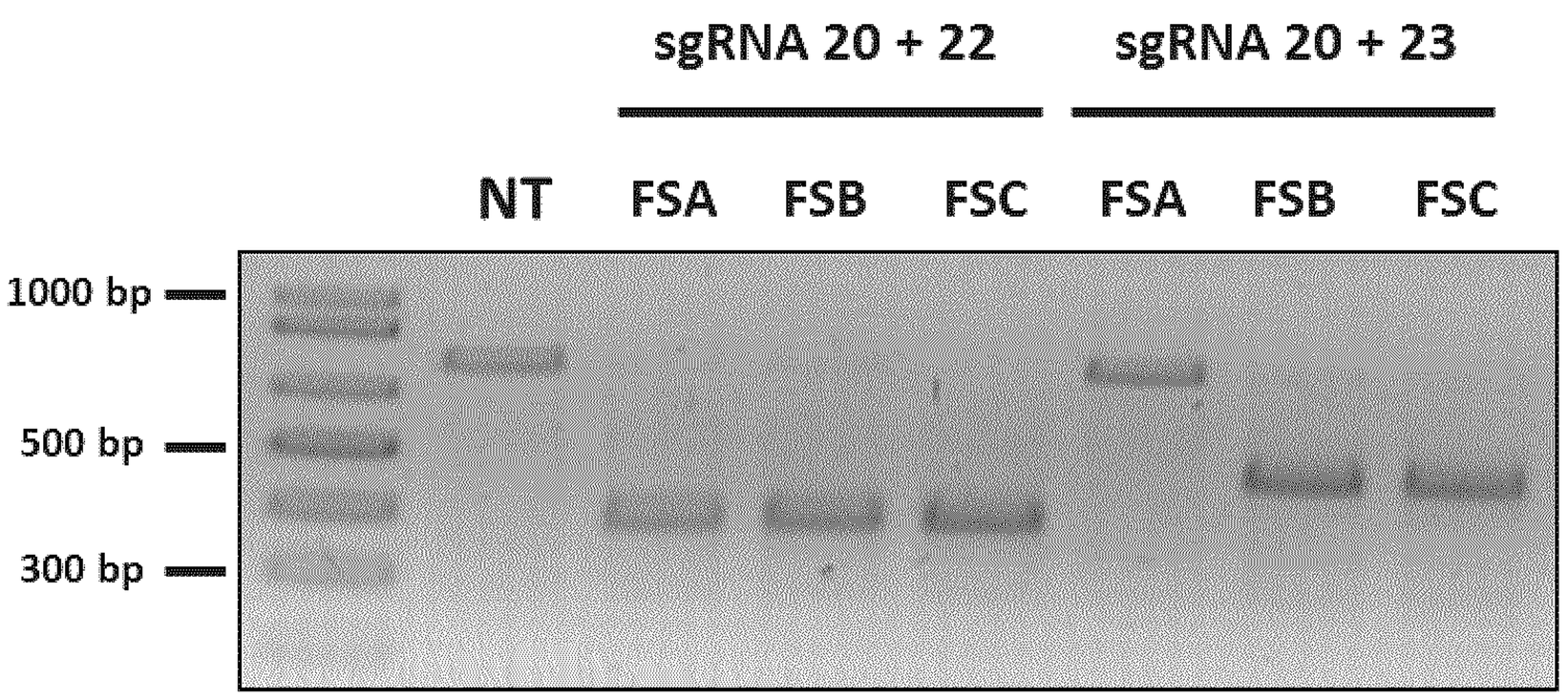

sgRNA 20 + 22
sgRNA 20 + 23
NT
FSA
FSB
FSC
FSA
FSB
FSC
1000 bp
500 bp
300 bp

FIG. 25

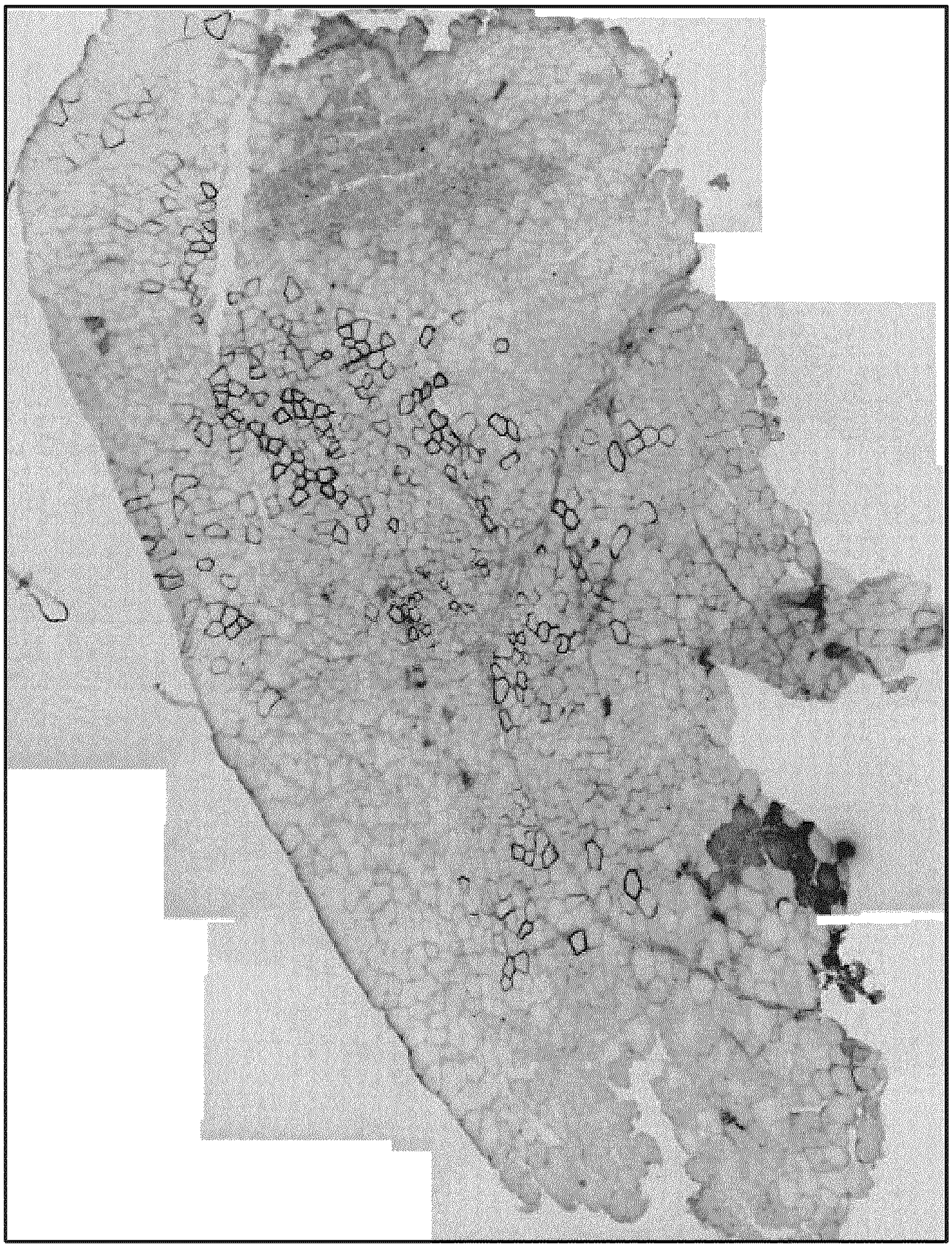


FIG. 26a

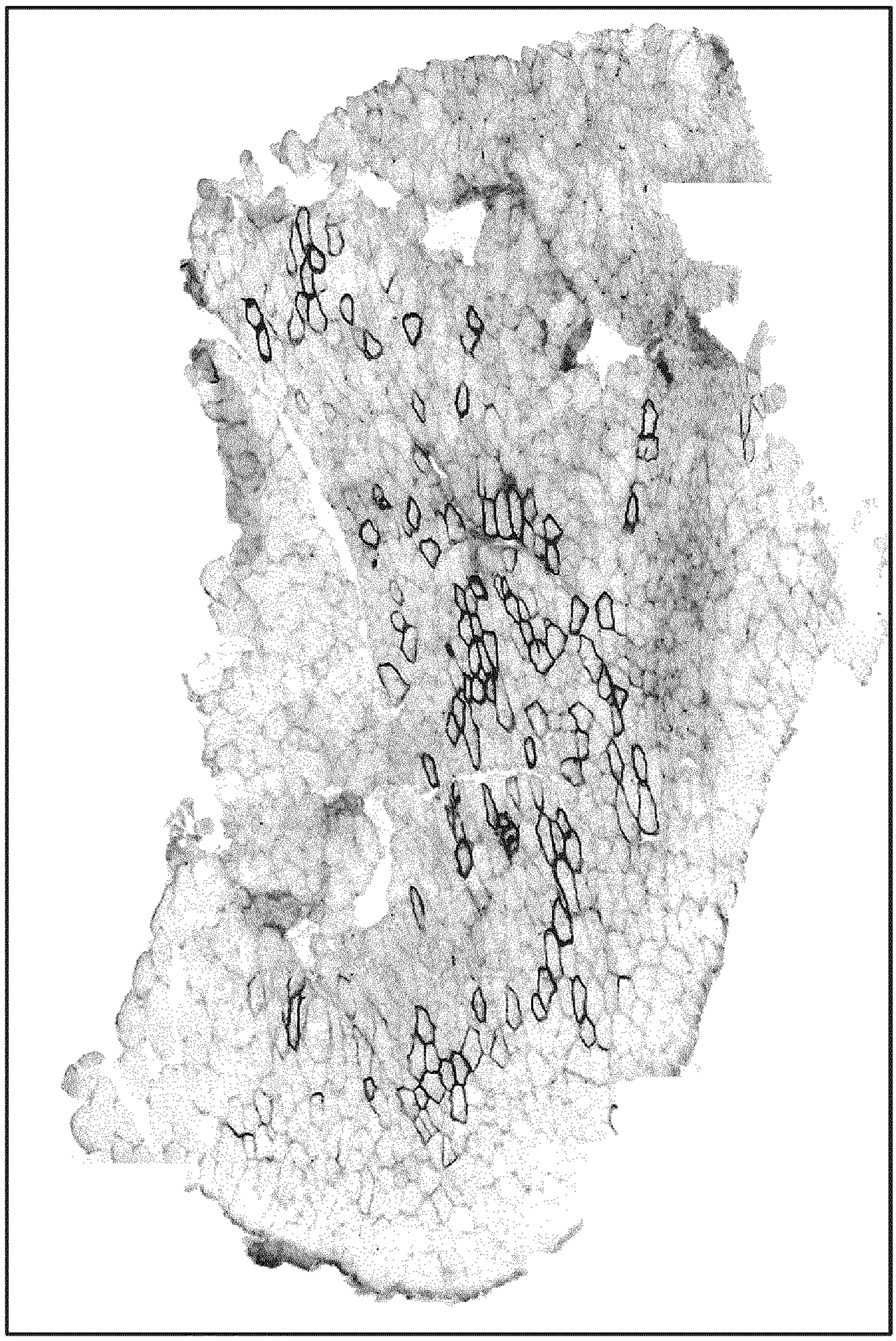


FIG. 26b

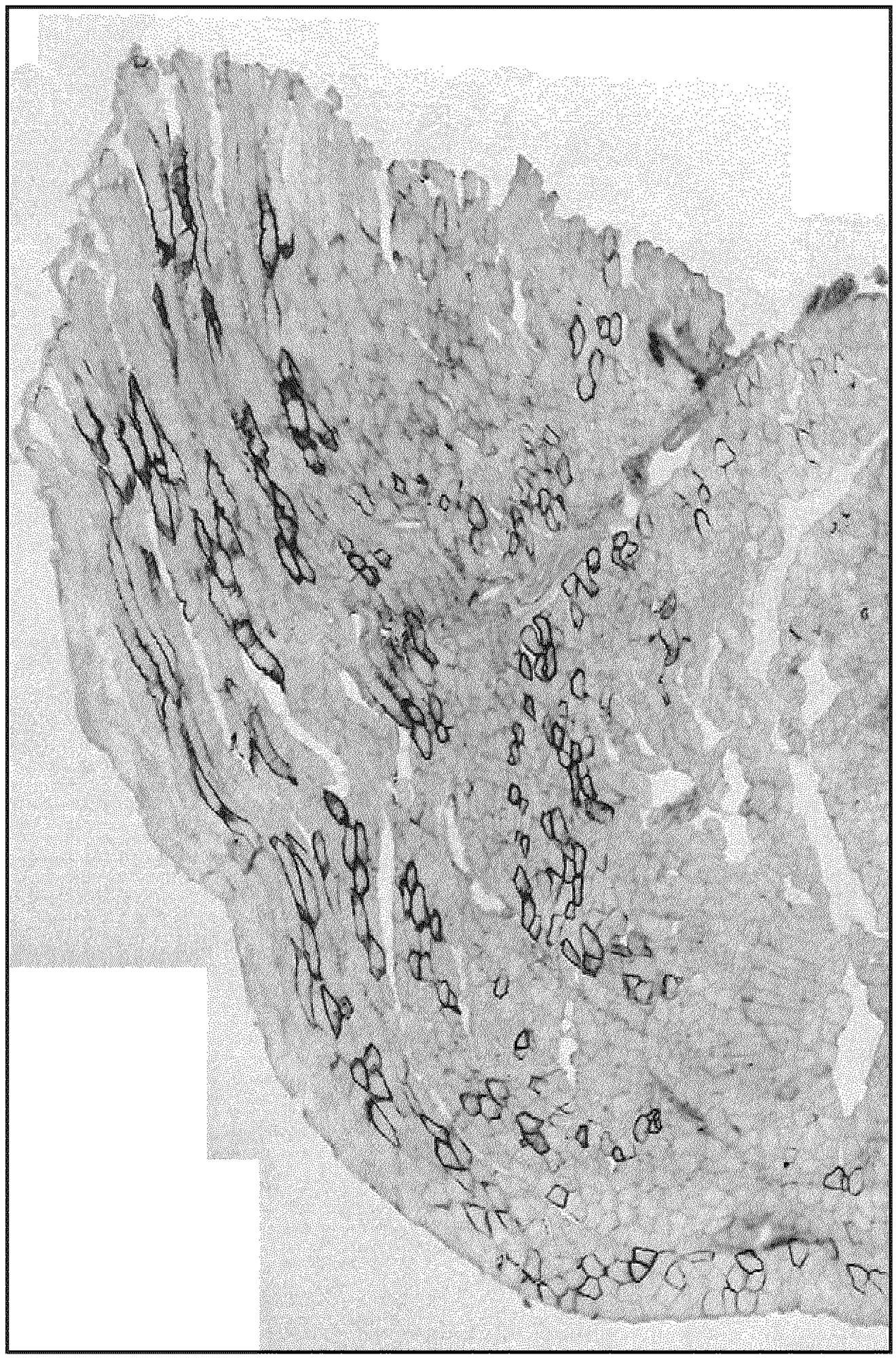


FIG. 26c

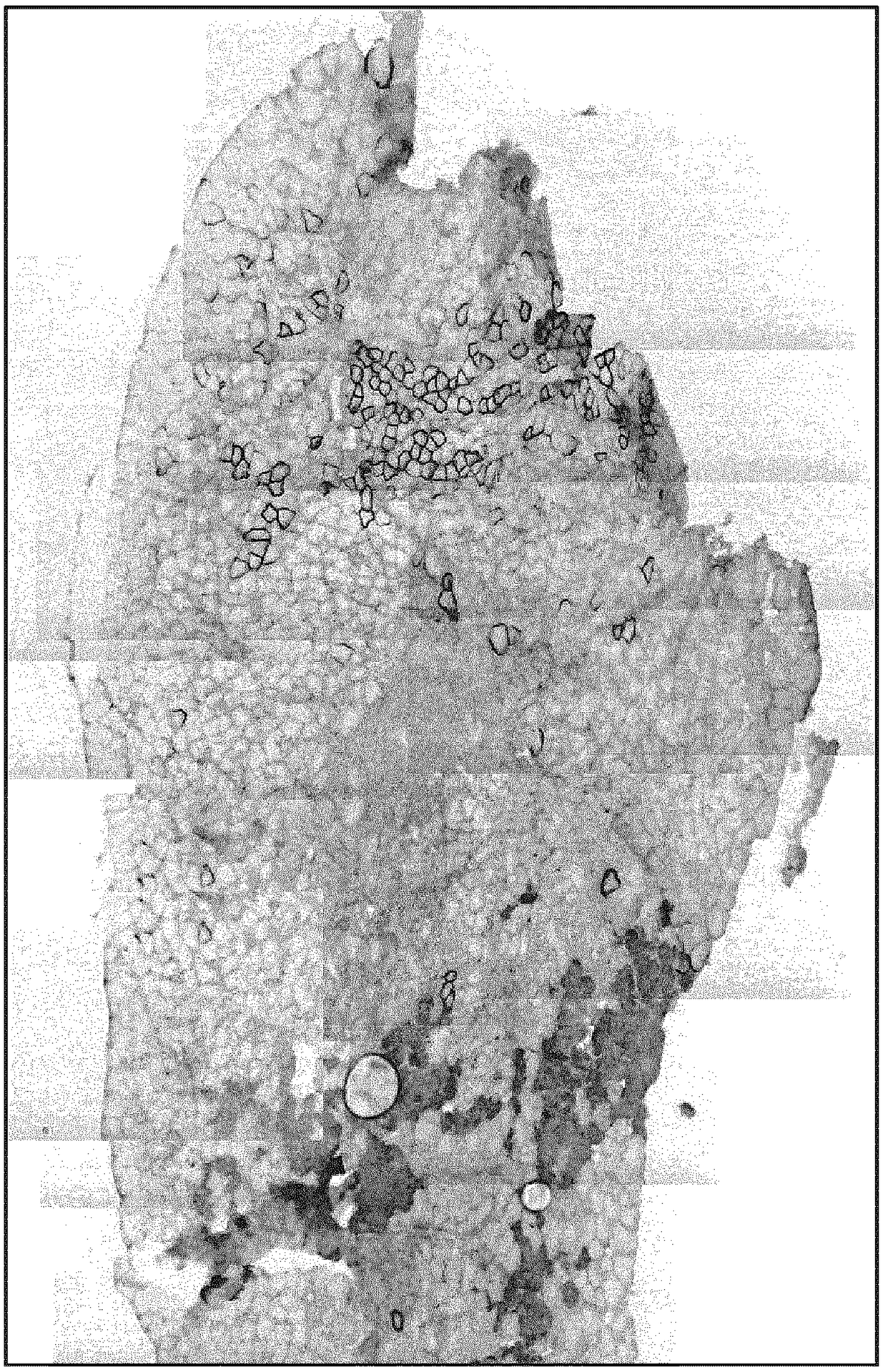


FIG. 26d

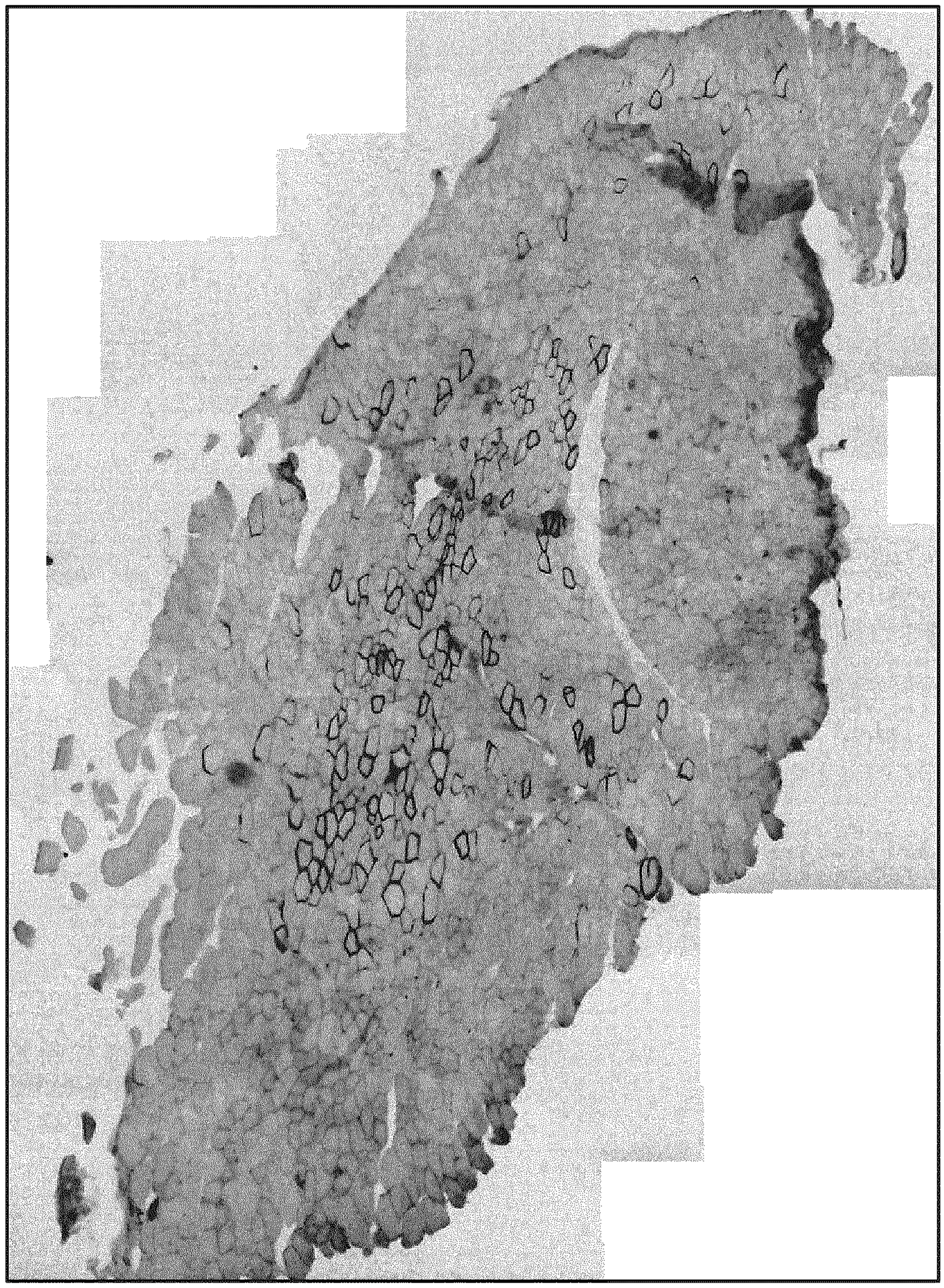


FIG. 26e

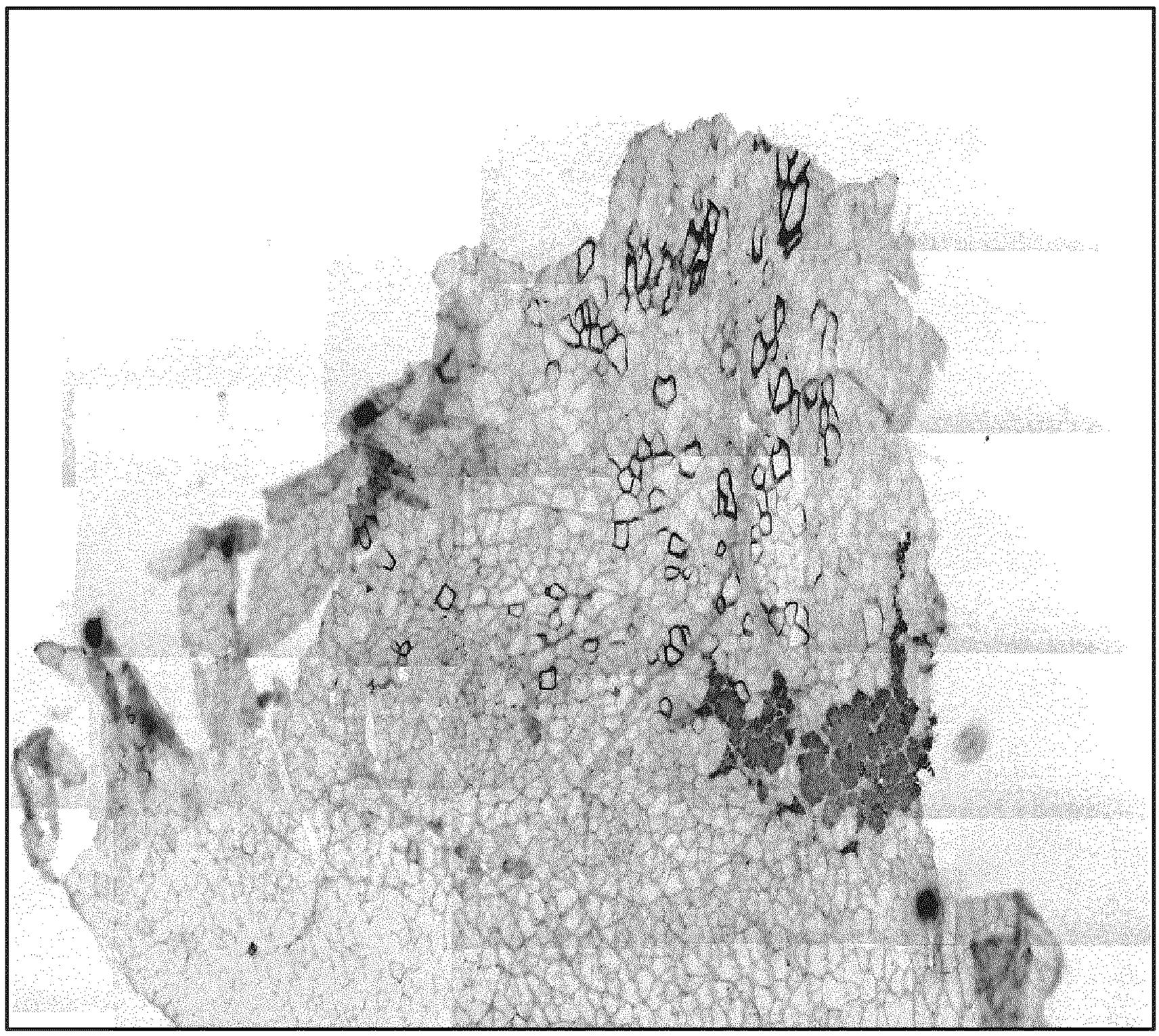


FIG. 26f

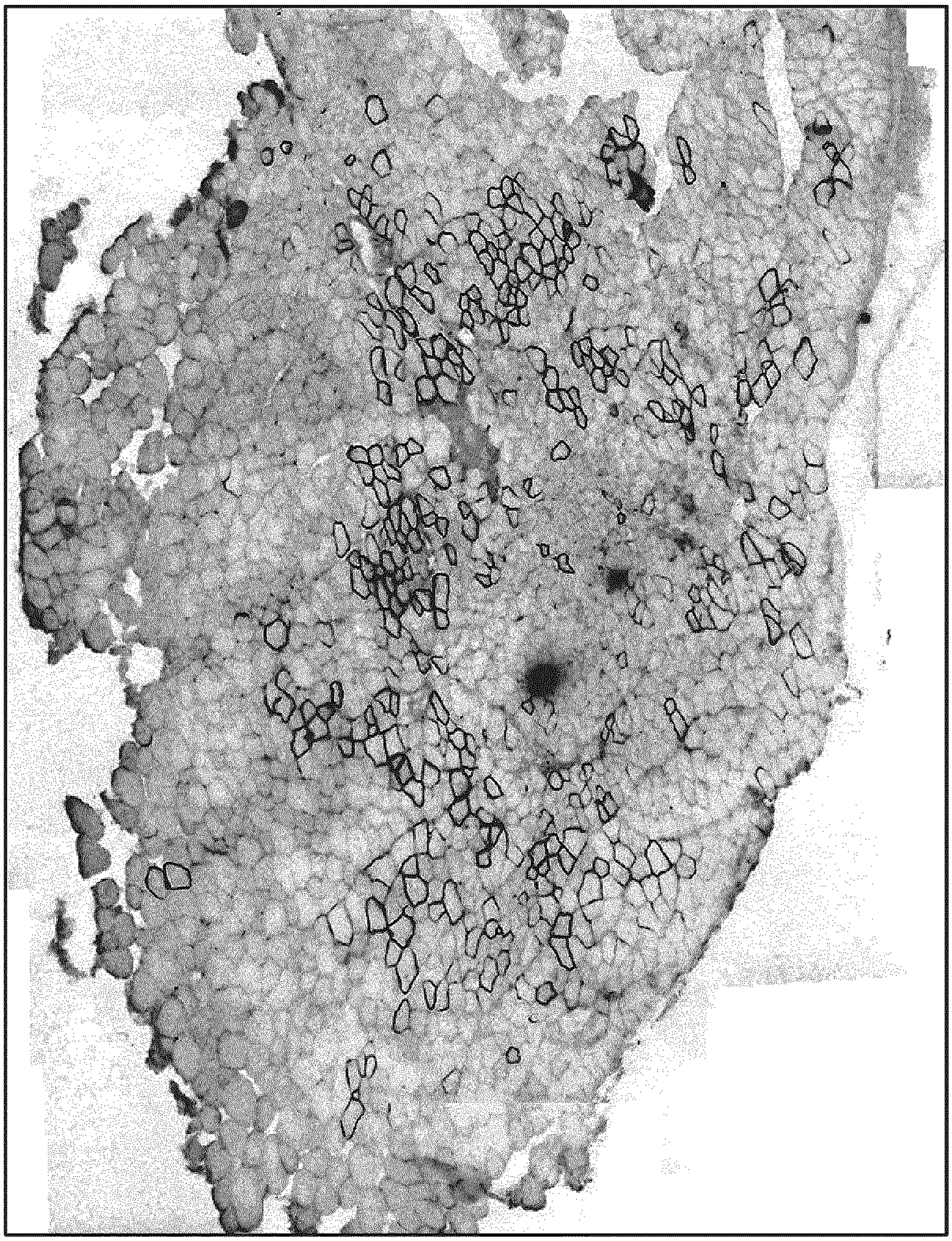


FIG. 26g

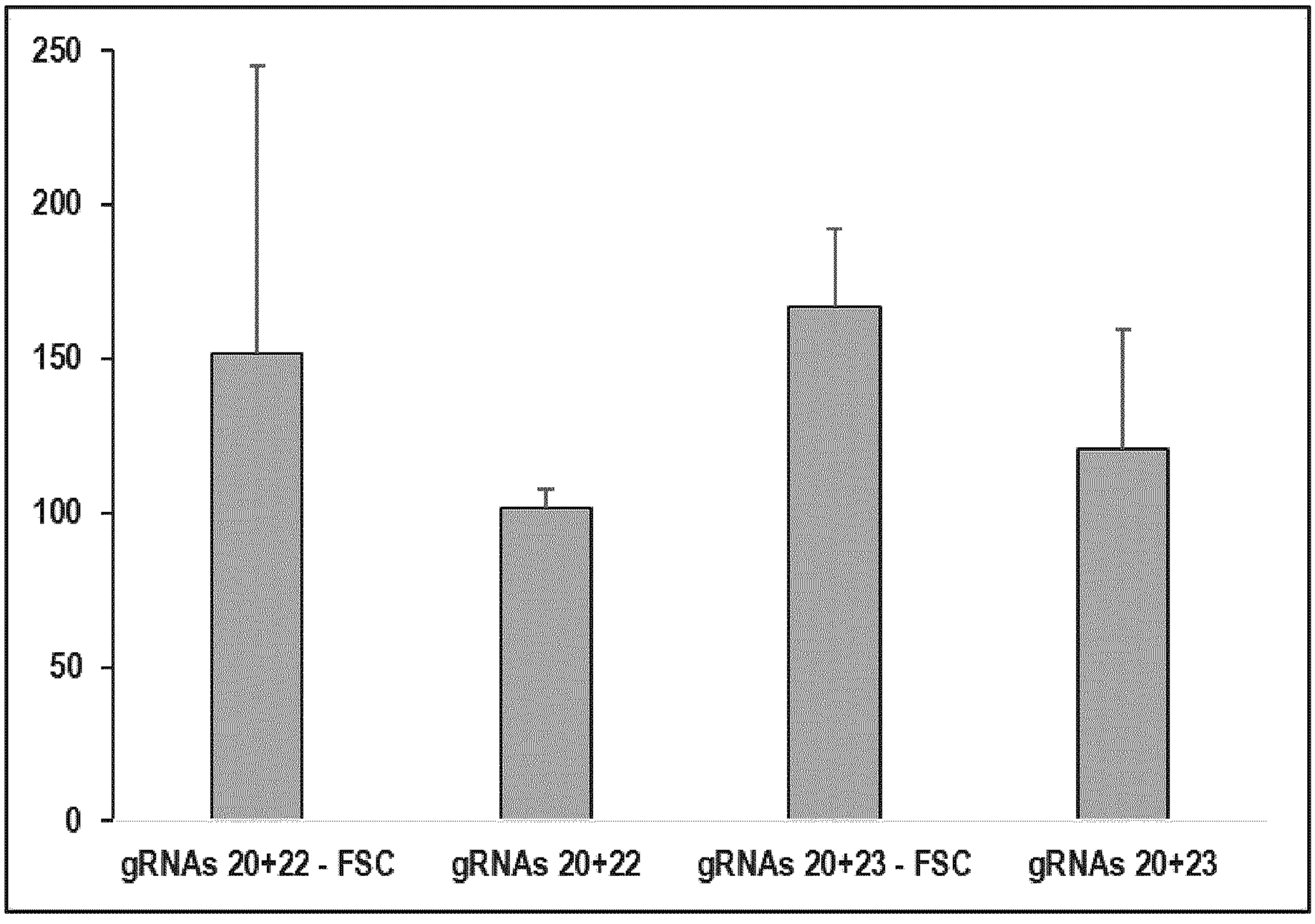
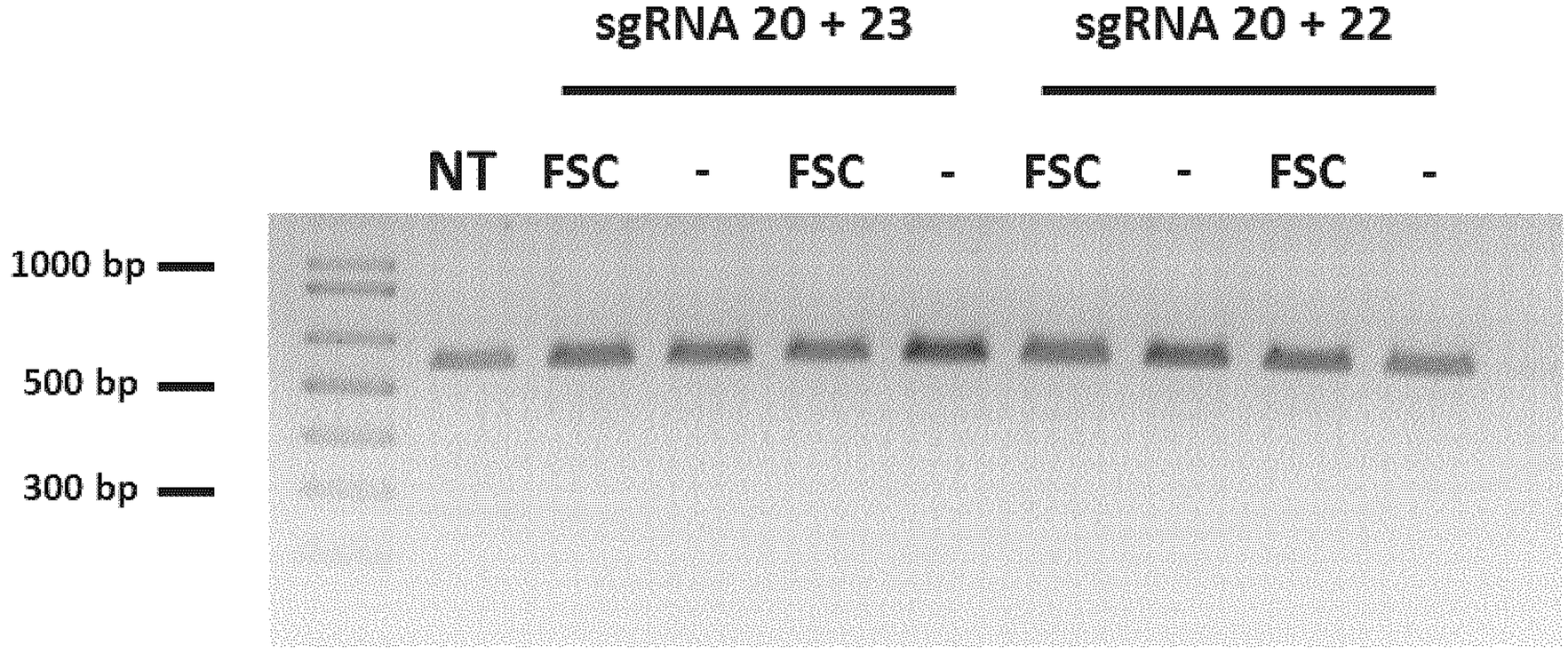

250
200
150
100
50
0
gRNAs 20+22 - FSC
gRNAs 20+22
gRNAs 20+23 - FSC
gRNAs 20+23

FIG. 27

a


sgRNA 20 + 23
sgRNA 20 + 22
NT FSC - FSC - FSC - FSC -
1000 bp
500 bp
300 bp

b

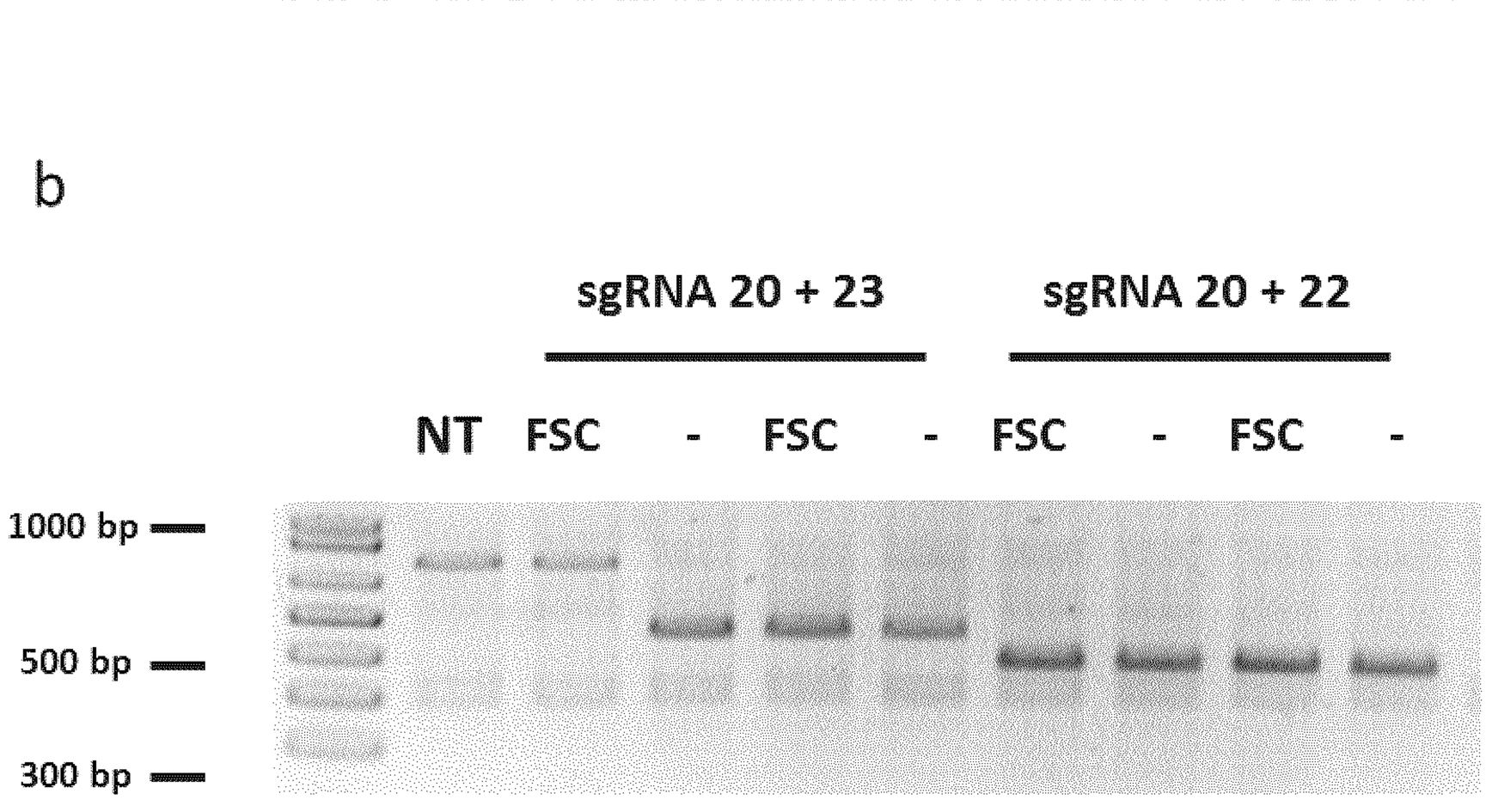

sgRNA 20 + 23
sgRNA 20 + 22
NT FSC - FSC - FSC - FSC -
1000 bp
500 bp
300 bp

FIG. 28

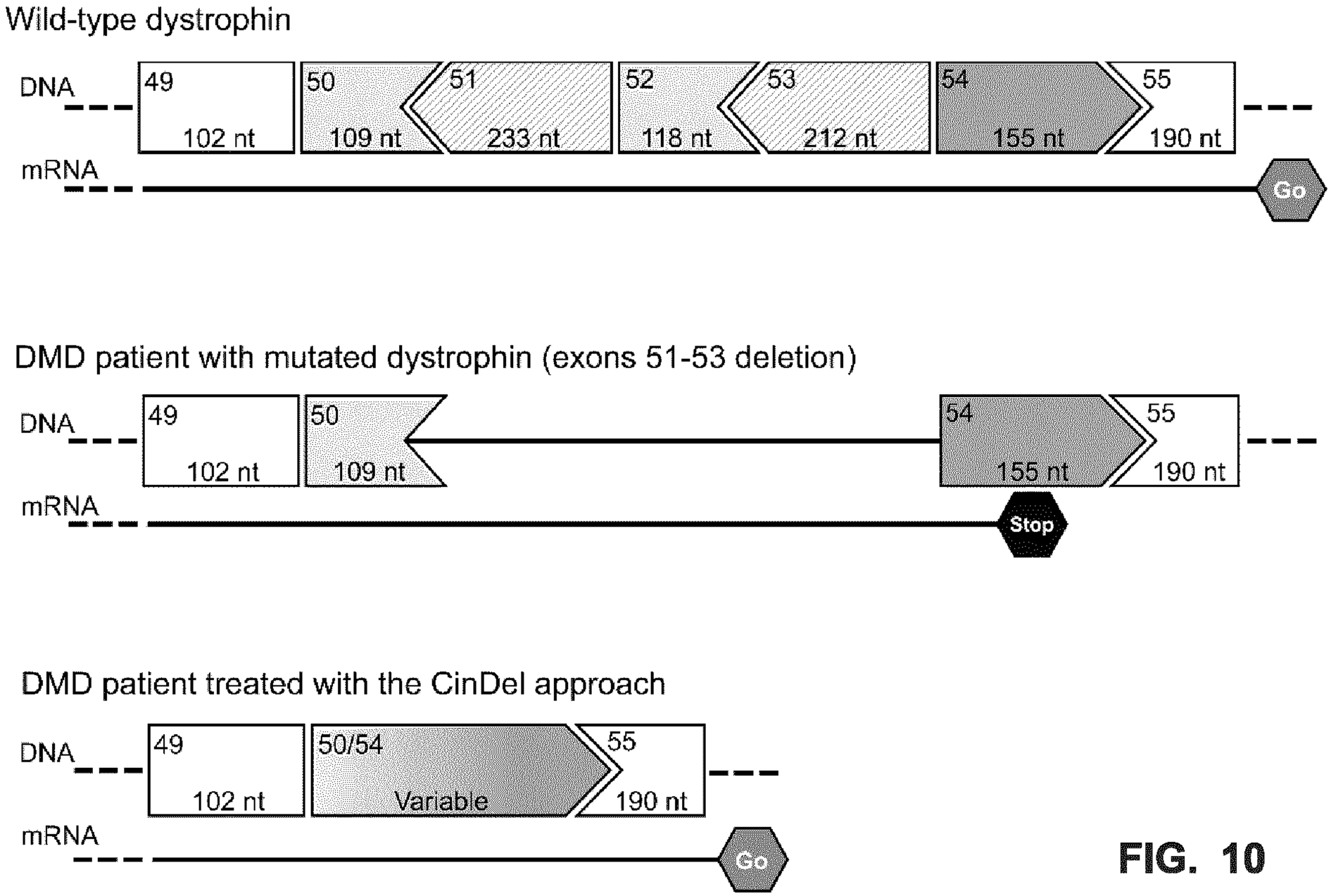

Wild-type dystrophin
DNA
49
102 nt
50
109 nt
51
233 nt
52
118 nt
53
212 nt
54
155 nt
55
190 nt
mRNA
Go
DMD patient with mutated dystrophin (exons 51-53 deletion)
DNA
49
102 nt
50
109 nt
54
155 nt
55
190 nt
mRNA
Stop
DMD patient treated with the CinDel approach
DNA
49
102 nt
50/54
Variable
55
190 nt
mRNA
Go

FIG. 10