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

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3137248 A1 2020/10/22

(21) 3 137 248

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2020/04/14
(87) Date publication PCT/PCT Publication Date: 2020/10/22
(85) Entrée phase nationale/National Entry: 2021/10/13
(86) N° demande PCT/PCT Application No.: US 2020/028148
(87) N° publication PCT/PCT Publication No.: 2020/214609
(30) Priorité/Priority: 2019/04/14 (US62/833,760)

(51) Cl.Int./Int.Cl. *C12N 15/52* (2006.01),
A61K 48/00 (2006.01), *C12N 15/11* (2006.01),
C12N 15/85 (2006.01), *C12N 9/22* (2006.01)

(71) Demandeur/Applicant:
DUKE UNIVERSITY, US

(72) Inventeurs/Inventors:
GERSBACH, CHARLES A., US;
BULAKLAK, KAREN, US;
ROBINSON-HAMM, JACQUELINE N., US

(74) Agent: SMART & BIGGAR LLP

(54) Titre : DELETION MEDIEE PAR UN VECTEUR AAV D'UN POINT CHAUD DE MUTATION IMPORTANT POUR LE TRAITEMENT DE LA DYSTROPHIE MUSCULAIRE DE DUCHENNE

(54) Title: AAV VECTOR-MEDIATED DELETION OF LARGE MUTATIONAL HOTSPOT FOR TREATMENT OF DUCHENNE MUSCULAR DYSTROPHY

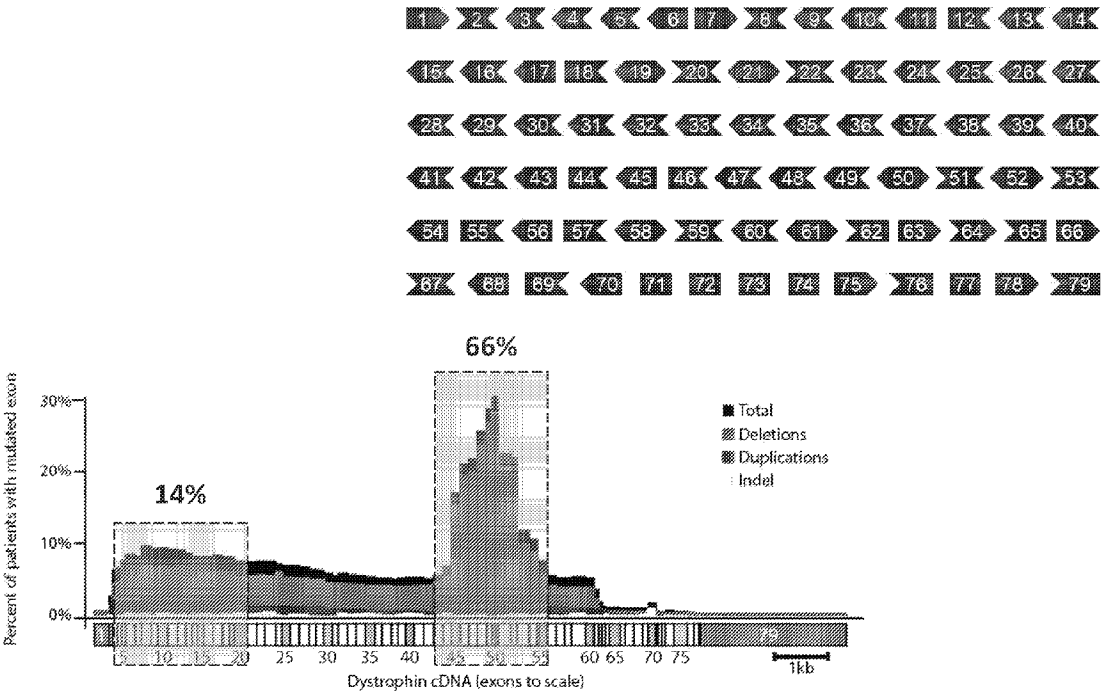


FIG. 1

(57) Abrégé/Abstract:
Disclosed herein are therapeutic targets for the correction of the human dystrophin gene by gene editing and methods of use.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 October 2020 (22.10.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/214609 A1

(51) International Patent Classification:
C12N 15/52 (2006.01) *C12N 15/85* (2006.01)
A61K 48/00 (2006.01) *C12N 9/22* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2020/028148

(22) International Filing Date:
14 April 2020 (14.04.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/833,760 14 April 2019 (14.04.2019) US

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

(72) Inventors: **GERSBACH, Charles A.**; 3220 Wilderness Road, Durham, NC 27712 (US). **BULAKLAK, Karen**; c/o Duke University, 2812 Erwin Road, Suite 306, Durham, NC 27705 (US). **ROBINSON-HAMM, Jacqueline N.**; c/o Duke University, 2812 Erwin Road, Suite 306, Durham, NC 27705 (US).

(74) Agent: **COX, Julia M.** et al.; Michael Best & Friedrich LLP, 100 East Wisconsin Avenue, Suite 3300, Milwaukee, WI 53202-4108 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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).

(54) Title: AAV VECTOR-MEDIATED DELETION OF LARGE MUTATIONAL HOTSPOT FOR TREATMENT OF DUCHENNE MUSCULAR DYSTROPHY

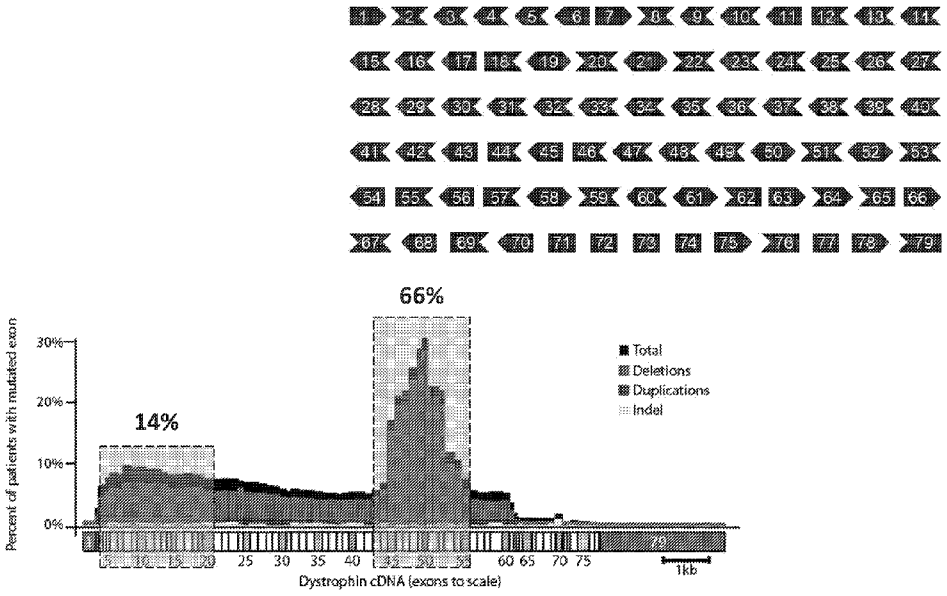


FIG. 1

(57) Abstract: Disclosed herein are therapeutic targets for the correction of the human dystrophin gene by gene editing and methods of use.

[Continued on next page]

WO 2020/214609 A1

WO 2020/214609 A1

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

AAV VECTOR-MEDIATED DELETION OF LARGE MUTATIONAL HOTSPOT FOR TREATMENT OF DUCHENNE MUSCULAR DYSTROPHY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 62/833,760, filed April 14, 2019, which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant R01AR069085 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD

[0003] The present disclosure relates to the field of gene expression alteration, genome engineering, and genomic alteration of genes using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) 9-based systems and viral delivery systems. The present disclosure also relates to the field of genome engineering and genomic alteration of genes in muscle, such as skeletal muscle and cardiac muscle.

INTRODUCTION

[0004] CRISPR/Cas9-based gene editing systems can be used to introduce site-specific double strand breaks at targeted genomic loci. This DNA cleavage stimulates the natural DNA-repair machinery, leading to one of two possible repair pathways. In the absence of a donor template, the break will be repaired by non-homologous end joining (NHEJ), an error-prone repair pathway that leads to small insertions or deletions of DNA. This method can be used to intentionally disrupt, delete, or alter the reading frame of targeted gene sequences. However, if a donor template is provided along with the nucleases, then the cellular machinery will repair the break by homologous recombination, which is enhanced several orders of magnitude in the presence of DNA cleavage. This method can be used to introduce specific changes in the DNA sequence at target sites. Engineered nucleases have been used for gene editing in a variety of human stem cells and cell lines, and for gene editing in the mouse liver. However, the major hurdle for implementation of these technologies is delivery to particular tissues *in vivo* in a way that is effective, efficient, and facilitates successful genome modification.

[0005] Hereditary genetic diseases have devastating effects on children in the United States. These diseases currently have no cure and can only be managed by attempts to alleviate the symptoms. For decades, the field of gene therapy has promised a cure to these diseases. However technical hurdles regarding the safe and efficient delivery of therapeutic genes to cells and patients have limited this approach. Duchenne muscular dystrophy (DMD) is a fatal genetic disease, clinically characterized by muscle wasting, loss of ambulation, and death typically in the third decade of life due to the loss of functional dystrophin. DMD is the result of inherited or spontaneous mutations in the dystrophin gene. Most mutations causing DMD are a result of deletions of exon(s), pushing the translational reading frame out of frame. The majority of DMD mutations are deletions (~ 68%) of one or more of its 79 exons that shift the reading frame and terminate expression of the full-length transcript. Deletions mostly occur in two "hotspots" of the gene, which encompass exons 2 through 20 (~ 1/3 of all deletions) and exons 45 through 55 (~2/3 of all deletions). Becker muscular dystrophy (BMD) patients with naturally occurring in-frame deletions of the entire 45 to 55 region of the dystrophin gene exhibit delayed disease onset and minimal skeletal muscle pathology.

[0006] Dystrophin is a key component of a protein complex that is responsible for regulating muscle cell integrity and function. DMD patients typically lose the ability to physically support themselves during childhood, become progressively weaker during the teenage years, and die in their twenties. Current experimental gene therapy strategies for DMD require repeated administration of transient gene delivery vehicles or rely on permanent integration of foreign genetic material into the genomic DNA. Both of these methods have serious safety concerns. Furthermore, these strategies have been limited by an inability to deliver the large and complex dystrophin gene sequence. There remains a need for more precise and efficient gene editing tools for correcting and treating patients with mutations in the dystrophin gene.

SUMMARY

[0007] In an aspect, the disclosure relates to a CRISPR-Cas system. The CRISPR-Cas system may include one or more vectors encoding a composition, the composition comprising: (a) a first guide RNA (gRNA) molecule targeting intron 44 of dystrophin; (b) a second gRNA molecule targeting intron 55 of dystrophin; and (c) a Cas9 protein; and (d) one or more Cas9 gRNA scaffolds. In some embodiments, the system comprises a single vector. In some embodiments, the system comprises two or more vectors, wherein the two or more vectors comprises a first vector and a second vector. In some embodiments, (a) the first vector encodes the first gRNA molecule and the second gRNA molecule; and (b) the

second vector encodes the Cas9 protein. In some embodiments, (a) the first vector encodes the first gRNA molecule; and (b) the second vector encodes the second gRNA molecule. In some embodiments, the first vector further encodes the Cas9 protein. In some embodiments, the second vector further encodes the Cas9 protein. In some embodiments, the expression of the Cas9 protein is driven by a constitutive promoter or a muscle-specific promoter. In some embodiments, the muscle-specific promoter comprises a MHCK7 promoter, a CK8 promoter, or a Spc512 promoter. In some embodiments, the single vector encodes the first gRNA molecule, the second gRNA molecule, and the Cas9 protein. In some embodiments, the vector comprises at least one bidirectional promoter. In some embodiments, the bidirectional promoter comprises: a first promoter driving expression of the first gRNA molecule and/or the second gRNA molecule; and a second promoter driving expression of the Cas9 protein. In some embodiments, the first gRNA targets the polynucleotide of SEQ ID NO:2 or a 5' truncation thereof. In some embodiments, the second gRNA targets the polynucleotide of SEQ ID NO:3 or a 5' truncation thereof. In some embodiments, the Cas9 protein is SpCas9, SaCas9, or St1Cas9 protein. In some embodiments, the Cas9 gRNA scaffold is a SaCas9 gRNA scaffold. In some embodiments, the SaCas9 gRNA scaffold comprises or is encoded by the polynucleotide of SEQ ID NO:4. In some embodiments, the Cas9 protein is a SaCas9 protein encoded by the polynucleotide of SEQ ID NO:11. In some embodiments, the vector comprises at least one polynucleotide selected from SEQ ID NOs: 1-13 and 24. In some embodiments, the vector comprises the polynucleotide sequence of SEQ ID NO: 24. In some embodiments, the vector comprises a polynucleotide sequence that is selected from SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 29, and SEQ ID NO: 30. In some embodiments, the vector is a viral vector. In some embodiments, the vector is an Adeno-associated virus (AAV) vector. In some embodiments, the AAV vector is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV-10, AAV-11, AAV-12, AAV-13, or AAVrh.74. In some embodiments, the vector comprises a ubiquitous promoter or a tissue-specific promoter operably linked to the polynucleotide sequence encoding the first gRNA molecule, the second gRNA molecule, and/or the Cas9 protein. In some embodiments, the tissue-specific promoter is a muscle specific promoter.

[0008] In a further aspect, the disclosure relates to a cell comprising the herein described system.

[0009] Another aspect of the disclosure provides a kit comprising the herein described system.

[00010] Another aspect of the disclosure provides a method of correcting a mutant dystrophin gene in a cell. The method may include administering to a cell the herein described system.

[00011] Another aspect of the disclosure provides a method of genome editing a mutant dystrophin gene in a subject. The method may include administering to the subject a herein described system or cell. The system or cell may be administered to the subject intramuscularly, intravenously, or a combination thereof.

[00012] Another aspect of the disclosure provides a method of treating a subject having a mutant dystrophin gene. The method may include administering to the subject the herein described system or cell. The system or cell may be administered to the subject intramuscularly, intravenously, or a combination thereof.

[00013] The disclosure provides for other aspects and embodiments that will be apparent in light of the following detailed description and accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[00014] **FIG. 1** shows the two deletion prone hotspots in dystrophin. The dystrophin gene (which may be referred to as DMD) is the largest known gene in humans (2.3 Mbp). Approximately, 68% of mutations are large exon deletions that lead to frameshift errors.

[00015] **FIG. 2** shows details relating to the exon 45 through exon 55 mutational hotspot. Approximately 45% of all DMD mutations, and many commonly deleted single exons, are located in this region. Patients with exon 45 to 55 in-frame deletion display milder dystrophic phenotype. AONs (antisense oligonucleotides) have been used to induce exon skipping in this region.

[00016] **FIG. 3** shows excision of exons 45 through 55 of dystrophin. This system is being tested in a humanized mouse carrying the human gene with a deletion of exon 52.

[00017] **FIG. 4** shows injection of a system to excise exons 45 through 55 of dystrophin in neonatal mice. Neonatal mice were systemically injected at 2 days postnatal (P2). Muscles were harvested 8 weeks post-treatment. PCR bands show the intended deletion.

[00018] **FIG. 5** shows dystrophin expression in systemically treated mice. 10x magnification, dual vector P2 injected, 8 weeks post-treatment.

[00019] **FIG. 6** shows the traditional two vector system as compared to the one vector system. Advantages to the one vector system may include: having all necessary editing components on a single vector, ability to increase effective dose, streamlining of other vector production (single therapeutic agent), use/incorporation of muscle-specific promoters (CK8, Spc512, MHCK7), and ability to target combinations of exons and large deletions (by changing guide sequences).

[00020] **FIG. 7** shows a vector design comparison. The all-in-one vector components (total packaged DNA < 4.8 kb include: SaCas9 (~3.2 kb); mini polyadenylation signal (60 bp) or bGH polyadenylation signal (232 bp); constitutive EFS promoter (252 bp) or muscle specific promoter).

[00021] **FIG. 8** shows the all-in-one vector for deletion of exons 45-55 and *in vitro* analyses in HEK293s.

[00022] **FIG. 9A** is a schematic diagram of the dystrophin gene from immortalized myoblasts isolated from a DMD patient, showing the deletion of exons 48-50. **FIG. 9B** shows results from deletion PCR of genomic DNA and cDNA from treated DMD patients, indicating that exon 45-55 was effectively deleted with vectors as detailed herein. **FIG. 9C** is a Western blot of cell lysates, showing that untreated myoblasts produced no dystrophin protein, while transfected myoblasts expressed a smaller dystrophin protein compared to the positive control, consistent with hotspot deletion.

[00023] **FIG. 10** are images of cardiac muscle cells from neonatal hDMDΔ52/mdx mice injected with either AAV-CRISPR targeting a control locus (top panel) or targeting exon 45-55 (bottom panel). Cells were harvested 8 weeks post injection. Cells were stained with DAPI or for dystrophin. 10x magnification, scale bar = 200 μm.

[00024] **FIG. 11** is a schematic diagram of the versions of all-in-one vector 5.

[00025] **FIG. 12** are images of TA muscle cells 8 weeks after injection with the vectors as indicated, at 10x magnification.

[00026] **FIG. 13** are graphs showing SaCas9 and gRNA *in vivo* expression resulting from treatment with the indicated all-in-one vectors, as determined by qRT-PCR using TA samples 8 weeks post-injection. N=3-4.

[00027] **FIG. 14** are graphs showing the stability of all-in-one (AIO) vectors *in vivo*. The left graph are results from qPCR using TA samples 8 weeks post-injection. The right graphs are results from IFN-gamma ELISpot assay against SaCas9. N=3-4 for both.

DETAILED DESCRIPTION

[00028] As described herein, certain methods and engineered gRNAs have been discovered to be useful with CRISPR/CRISPR-associated (Cas) 9-based gene editing systems for altering the expression, genome engineering, and correcting or reducing the effects of mutations in the dystrophin gene involved in genetic diseases, such as DMD. The disclosed gRNAs were generated to target sites that are more amenable to clinical translation. For example, the gene encoding *S. pyogenes* Cas9 (SpCas9) is too large to be delivered by adeno-associated virus (AAV), a vector used for the systemic gene delivery to muscle when all other necessary regulatory sequences are included. Instead, the disclosed gRNAs were selected and screened for use with *S. aureus* Cas9 (SaCas9), which is about 1 kb smaller than SpCas9. The disclosed gRNAs, which target human dystrophin gene sequences, can be used with the CRISPR/Cas9-based system to target exons 45 to 55 of the human dystrophin gene, causing genomic deletions of this region in order to restore expression of functional dystrophin in cells from DMD patients.

[00029] Also described herein are genetic constructs, compositions, and methods for delivering CRISPR/Cas9-based gene editing system and multiple gRNAs to target the dystrophin gene. The presently disclosed subject matter also provides for methods for delivering the genetic constructs (e.g., vectors) or compositions comprising thereof to skeletal muscle and cardiac muscle. The vector can be an AAV, including modified AAV vectors. The presently disclosed subject matter describes a way to deliver active forms of this class of therapeutics to skeletal muscle or cardiac muscle that is effective, efficient, and facilitates successful genome modification, as well as provide a means to rewrite the human genome for therapeutic applications and target model species for basic science applications. The methods may relate to the use of a single AAV vector for the delivery of all of the editing components necessary for the excision of exons 45 through 55 of dystrophin.

[00030] Section headings as used in this section and the entire disclosure herein are merely for organizational purposes and are not intended to be limiting.

1. Definitions

[00031] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention. All

publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[00032] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms "a," "and" and "the" include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments "comprising," "consisting of" and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[00033] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[00034] As used herein, the term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, "about" can mean within 3 or more than 3 standard deviations, per the practice in the art. Alternatively, "about" can mean a range of up to 20%, preferably up to 10%, more preferably up to 5%, and more preferably still up to 1% of a given value. In certain aspects, the term "about" refers to a range of values that fall within 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value). Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value.

[00035] "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 currently known to cause disease and consequently the virus causes a very mild immune response.

[00036] "Binding region" as used herein refers to the region within a nuclease target region that is recognized and bound by the nuclease.

[00037] "Cardiac muscle" or "heart muscle" as used interchangeably herein means a type of involuntary striated muscle found in the walls and histological foundation of the heart, the myocardium. Cardiac muscle is made of cardiomyocytes or myocardiocytes. Myocardiocytes show striations similar to those on skeletal muscle cells but contain only one, unique nucleus, unlike the multinucleated skeletal cells. In certain embodiments, "cardiac muscle condition" refers to a condition related to the cardiac muscle, such as cardiomyopathy, heart failure, arrhythmia, and inflammatory heart disease.

[00038] "Coding sequence" or "encoding nucleic acid" as used herein means the nucleic acids (RNA or DNA molecule) that comprise a nucleotide sequence which encodes a protein. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The coding sequence may be codon optimized.

[00039] "Complement" or "complementary" as used herein means a nucleic acid can mean Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[00040] "Correcting", "genome editing," and "restoring" as used herein refers to changing a mutant gene that encodes a truncated protein or no protein at all, such that a full-length functional or partially full-length functional protein expression is obtained. Correcting or restoring a mutant gene may include replacing the region of the gene that has the mutation or replacing the entire mutant gene with a copy of the gene that does not have the mutation with a repair mechanism such as homology-directed repair (HDR). Correcting or restoring a mutant gene may also include repairing a frameshift mutation that causes a premature stop codon, an aberrant splice acceptor site, or an aberrant splice donor site, by generating a double stranded break in the gene that is then repaired using non-homologous end joining (NHEJ). NHEJ may add or delete at least one base pair during repair which may restore the proper reading frame and eliminate the premature stop codon. Correcting or restoring a mutant gene may also include disrupting an aberrant splice acceptor site or splice donor sequence. Correcting or restoring a mutant gene may also include deleting a non-essential gene segment by the simultaneous action of two nucleases on the same DNA strand in

order to restore the proper reading frame by removing the DNA between the two nuclease target sites and repairing the DNA break by NHEJ.

[00041] The term "directional promoter" refers to two or more promoters that are capable of driving transcription of two separate sequences in both directions. In one embodiment, one promoter drives transcription from 5' to 3' and the other promoter drives transcription from 3' to 5'. In one embodiment, bidirectional promoters are double-strand transcription control elements that can drive expression of at least two separate sequences, for example, coding or non-coding sequences, in opposite directions. Such promoter sequences may be composed of two individual promoter sequences acting in opposite directions, such as one nucleotide sequence linked to the other (complementary) nucleotide sequence, including packaging constructs comprising the two promoters in opposite directions, for example, by hybrid, chimeric or fused sequences comprising the two individual promoter sequences, or at least core sequences thereof, or else by only one transcription regulating sequence that can initiate the transcription in both directions. The two individual promoter sequences, in some embodiments, may be juxtaposed or a linker sequence can be located between the first and second sequences. A promoter sequence may be reversed to be combined with another promoter sequence in the opposite orientation. Genes located on both sides of a bidirectional promoter can be operably linked to a single transcription control sequence or region that drives the transcription in both directions. In other embodiments, the bidirectional promoters are not juxtaposed. For example, one promoter may drive transcription on the 5' end of a nucleotide fragment, and another promoter may drive transcription from the 3' end of the same fragment. In another embodiment, a first gene can be operably linked to the bidirectional promoter with or without further regulatory elements, such as a reporter or terminator elements, and a second gene can be operably linked to the bidirectional promoter in the opposite direction and by the complementary promoter sequence, again with or without further regulatory elements.

[00042] "Donor DNA", "donor template," and "repair template" as used interchangeably herein refers to a double-stranded DNA fragment or molecule that includes at least a portion of the gene of interest. The donor DNA may encode a full-functional protein or a partially-functional protein.

[00043] "Duchenne Muscular Dystrophy" or "DMD" as used interchangeably herein refers to a recessive, fatal, X-linked disorder that results in muscle degeneration and eventual death. DMD is a common hereditary monogenic disease and occurs in 1 in 3500 males. DMD is the result of inherited or spontaneous mutations that cause nonsense or frame shift mutations in the dystrophin gene. The majority of dystrophin mutations that cause DMD are

deletions of exons that disrupt the reading frame and cause premature translation termination in the dystrophin gene. DMD patients typically lose the ability to physically support themselves during childhood, become progressively weaker during the teenage years, and die in their twenties.

[00044] "Dystrophin" as used herein refers to a rod-shaped cytoplasmic protein which is a part of a protein complex that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. Dystrophin provides structural stability to the dystroglycan complex of the cell membrane that is responsible for regulating muscle cell integrity and function. The dystrophin gene or "DMD gene" as used interchangeably herein is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons code for the protein which is over 3500 amino acids.

[00045] "Exons 45 through 55" of dystrophin as used herein refers to an area where roughly 45% of all dystrophin mutations are located. Exon 45–55 deletions are associated with very mild Becker phenotypes and have even been found in asymptomatic individuals. Exon 45–55 multiexon skipping would be beneficial for roughly 50% of all DMD patients.

[00046] "Frameshift" or "frameshift mutation" as used interchangeably herein refers to a type of gene mutation wherein the addition or deletion of one or more nucleotides causes a shift in the reading frame of the codons in the mRNA. The shift in reading frame may lead to the alteration in the amino acid sequence at protein translation, such as a missense mutation or a premature stop codon.

[00047] "Functional" and "full-functional" as used herein describes protein that has biological activity. A "functional gene" refers to a gene transcribed to mRNA, which is translated to a functional protein.

[00048] "Fusion protein" as used herein refers to a chimeric protein created through the joining of two or more genes that originally coded for separate proteins. The translation of the fusion gene results in a single polypeptide with functional properties derived from each of the original proteins.

[00049] "Genetic construct" as used herein refers to the DNA or RNA molecules that comprise a nucleotide sequence that encodes a protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered. As used herein, the term "expressible

form" refers to gene constructs that contain the necessary regulatory elements operably linked to a coding sequence that encodes a protein such that when present in the cell of the individual, the coding sequence will be expressed.

[00050] "Genetic disease" as used herein refers to a disease, partially or completely, directly or indirectly, caused by one or more abnormalities in the genome, especially a condition that is present from birth. The abnormality may be a mutation, an insertion or a deletion. The abnormality may affect the coding sequence of the gene or its regulatory sequence. The genetic disease may be, but not limited to DMD, Becker Muscular Dystrophy (BMD), hemophilia, cystic fibrosis, Huntington's chorea, familial hypercholesterolemia (LDL receptor defect), hepatoblastoma, Wilson's disease, congenital hepatic porphyria, inherited disorders of hepatic metabolism, Lesch Nyhan syndrome, sickle cell anemia, thalassaemias, xeroderma pigmentosum, Fanconi's anemia, retinitis pigmentosa, ataxia telangiectasia, Bloom's syndrome, retinoblastoma, and Tay-Sachs disease.

[00051] "Homology-directed repair" or "HDR" as used interchangeably herein refers to a mechanism in cells to repair double strand DNA lesions when a homologous piece of DNA is present in the nucleus, mostly in G2 and S phase of the cell cycle. HDR uses a donor DNA template to guide repair and may be used to create specific sequence changes to the genome, including the targeted addition of whole genes. If a donor template is provided along with the CRISPR/Cas9-based gene editing system, then the cellular machinery will repair the break by homologous recombination, which is enhanced several orders of magnitude in the presence of DNA cleavage. When the homologous DNA piece is absent, non-homologous end joining may take place instead.

[00052] "Genome editing" as used herein refers to changing a gene. Genome editing may include correcting or restoring a mutant gene. Genome editing may include knocking out a gene, such as a mutant gene or a normal gene. Genome editing may be used to treat disease or enhance muscle repair by changing the gene of interest.

[00053] "Identical" or "identity" as used herein in the context of two or more nucleic acids or polypeptide sequences means that the sequences have a specified percentage of residues that are the same over a specified region. The percentage may be calculated by optimally aligning the two sequences, comparing the two sequences over the specified region, determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the specified region, and multiplying the result by 100 to yield the percentage of sequence identity. In cases where the two sequences are

of different lengths or the alignment produces one or more staggered ends and the specified region of comparison includes only a single sequence, the residues of single sequence are included in the denominator but not the numerator of the calculation. When comparing DNA and RNA, thymine (T) and uracil (U) may be considered equivalent. Identity may be performed manually or by using a computer sequence algorithm such as BLAST or BLAST 2.0.

[00054] "Mutant gene" or "mutated gene" as used interchangeably herein refers to a gene that has undergone a detectable mutation. A mutant gene has undergone a change, such as the loss, gain, or exchange of genetic material, which affects the normal transmission and expression of the gene. A "disrupted gene" as used herein refers to a mutant gene that has a mutation that causes a premature stop codon. The disrupted gene product is truncated relative to a full-length undisrupted gene product.

[00055] "Non-homologous end joining (NHEJ) pathway" as used herein refers to a pathway that repairs double-strand breaks in DNA by directly ligating the break ends without the need for a homologous template. The template-independent re-ligation of DNA ends by NHEJ is a stochastic, error-prone repair process that introduces random micro-insertions and micro-deletions (indels) at the DNA breakpoint. This method may be used to intentionally disrupt, delete, or alter the reading frame of targeted gene sequences. NHEJ typically uses short homologous DNA sequences called microhomologies to guide repair. These microhomologies are often present in single-stranded overhangs on the end of double-strand breaks. When the overhangs are perfectly compatible, NHEJ usually repairs the break accurately, yet imprecise repair leading to loss of nucleotides may also occur, but is much more common when the overhangs are not compatible.

[00056] "Normal gene" as used herein refers to a gene that has not undergone a change, such as a loss, gain, or exchange of genetic material. The normal gene undergoes normal gene transmission and gene expression. For example, a normal gene may be a wild-type gene.

[00057] "Nuclease mediated NHEJ" as used herein refers to NHEJ that is initiated after a nuclease, such as a Cas9 molecule, cuts double stranded DNA.

[00058] "Nucleic acid" or "oligonucleotide" or "polynucleotide" as used herein means at least two nucleotides covalently linked together. The depiction of a single strand also defines the sequence of the complementary strand. Thus, a nucleic acid also encompasses the complementary strand of a depicted single strand. Many variants of a nucleic acid may

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

[00059] Nucleic acids may be single stranded or double stranded or may contain portions of both double stranded and single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine and isoguanine. Nucleic acids may be obtained by chemical synthesis methods or by recombinant methods.

[00060] "Operably linked" as used herein means that expression of a gene is under the control of a promoter with which it is spatially connected. A promoter may be positioned 5' (upstream) or 3' (downstream) of a gene under its control. The distance between the promoter and a gene may be approximately the same as the distance between that promoter and the gene it controls in the gene from which the promoter is derived. As is known in the art, variation in this distance may be accommodated without loss of promoter function.

[00061] "Partially-functional" as used herein describes a protein that is encoded by a mutant gene and has less biological activity than a functional protein but more than a non-functional protein.

[00062] "Premature stop codon" or "out-of-frame stop codon" as used interchangeably herein refers to nonsense mutation in a sequence of DNA, which results in a stop codon at location not normally found in the wild-type gene. A premature stop codon may cause a protein to be truncated or shorter compared to the full-length version of the protein.

[00063] "Promoter" as used herein means a synthetic or naturally-derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively (constitutive promoter), or differentially with respect to cell, the tissue or organ in which

expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter, human U6 (hU6) promoter, and CMV IE promoter. Examples of muscle-specific promoters may include a MHCK7 promoter, a CK8 promoter, and a Spc512 promoter.

[00064] "Skeletal muscle" as used herein refers to a type of striated muscle, which is under the control of the somatic nervous system and attached to bones by bundles of collagen fibers known as tendons. Skeletal muscle is made up of individual components known as myocytes, or "muscle cells", sometimes colloquially called "muscle fibers." Myocytes are formed from the fusion of developmental myoblasts (a type of embryonic progenitor cell that gives rise to a muscle cell) in a process known as myogenesis. These long, cylindrical, multinucleated cells are also called myofibers.

[00065] "Skeletal muscle condition" as used herein refers to a condition related to the skeletal muscle, such as muscular dystrophies, aging, muscle degeneration, wound healing, and muscle weakness or atrophy.

[00066] "Subject" and "patient" as used herein interchangeably refers to any vertebrate, including, but not limited to, a mammal (e.g., cow, pig, camel, llama, horse, goat, rabbit, sheep, hamsters, guinea pig, cat, dog, rat, and mouse, a non-human primate (for example, a monkey, such as a cynomolgous or rhesus monkey, chimpanzee, etc.) and a human). In some embodiments, the subject may be a human or a non-human. The subject or patient may be undergoing other forms of treatment.

[00067] "Target gene" as used herein refers to any nucleotide sequence encoding a known or putative gene product. The target gene may be a mutated gene involved in a genetic disease. In certain embodiments, the target gene is a human dystrophin gene. In certain embodiments, the target gene is a mutant human dystrophin gene.

[00068] "Target region" as used herein refers to the region of the target gene to which the CRISPR/Cas9-based gene editing system is designed to bind and cleave.

[00069] "Transgene" as used herein refers to a gene or genetic material containing a gene sequence that has been isolated from one organism and is introduced into a different organism. This non-native segment of DNA may retain the ability to produce RNA or protein

in the transgenic organism, or it may alter the normal function of the transgenic organism's genetic code. The introduction of a transgene has the potential to change the phenotype of an organism.

[00070] "Variant" used herein with respect to a nucleic acid means (i) a portion or fragment of a referenced nucleotide sequence; (ii) the complement of a referenced nucleotide sequence or portion thereof; (iii) a nucleic acid that is substantially identical to a referenced nucleic acid or the complement thereof; or (iv) a nucleic acid that hybridizes under stringent conditions to the referenced nucleic acid, complement thereof, or a sequences substantially identical thereto.

[00071] "Variant" with respect to a peptide or polypeptide that differs in amino acid sequence by the insertion, deletion, or conservative substitution of amino acids, but retain at least one biological activity. Variant may also mean a protein with an amino acid sequence that is substantially identical to a referenced protein with an amino acid sequence that retains at least one biological activity. A conservative substitution of an amino acid, i.e., replacing an amino acid with a different amino acid of similar properties (e.g., hydrophilicity, degree and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes may be identified, in part, by considering the hydropathic index of amino acids, as understood in the art. Kyte et al., J. Mol. Biol. 157:105-132 (1982). The hydropathic index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydropathic indexes may be substituted and still retain protein function. In one aspect, amino acids having hydropathic indexes of ±2 are substituted. The hydrophilicity of amino acids may also be used to reveal substitutions that would result in proteins retaining biological function. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide. Substitutions may be performed with amino acids having hydrophilicity values within ±2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties.

[00072] "Vector" as used herein means a nucleic acid sequence containing an origin of replication. A vector may be a viral vector, bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self-replicating extrachromosomal vector, and preferably, is a DNA plasmid.

[00073] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

2. Genetic Constructs for Genome Editing of Dystrophin Gene

[00074] Provided herein are genetic constructs for genome editing, genomic alteration, and/or altering gene expression of a dystrophin gene. The dystrophin gene may be a human dystrophin gene. The genetic constructs include at least one gRNA that targets a dystrophin gene sequence(s). The at least one gRNA may target human and/or rhesus monkey dystrophin gene sequences and may be SaCas9-compatible targets. The disclosed gRNAs can be included in a CRISPR/Cas9-based gene editing system, including systems that use SaCas9, to target exons 45 through 55 of the human dystrophin gene. The disclosed gRNAs, which may be included in a CRISPR/Cas9-based gene editing system, can cause genomic deletions of the region of exons 45 through 55 of the human dystrophin gene in order to restore expression of functional dystrophin in cells from DMD patients.

a. Dystrophin Gene

[00075] Dystrophin is a rod-shaped cytoplasmic protein that is a part of a protein complex that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. Dystrophin provides structural stability to the dystroglycan complex of the cell membrane. The dystrophin gene is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being approximately 14 kb. 79 exons code for the protein, which is over 3500 amino acids. Normal skeleton muscle tissue contains only small amounts of dystrophin, but its absence of abnormal expression leads to the development of severe and incurable symptoms. Some mutations in the dystrophin gene lead to the production of defective dystrophin and severe dystrophic phenotype in affected patients. Some mutations in the dystrophin gene lead to partially functional dystrophin protein and a much milder dystrophic phenotype in affected patients.

[00076] DMD is the result of inherited or spontaneous mutations that cause nonsense or frame shift mutations in the dystrophin gene. Naturally occurring mutations and their consequences are relatively well understood for DMD. In-frame deletions that occur in the exon 45-55 regions (**FIG. 1**, **FIG. 2**) contained within the rod domain can produce highly functional dystrophin proteins, and many carriers are asymptomatic or display mild symptoms. Furthermore, more than 60% of patients may theoretically be treated by targeting this region as a whole (exons 45 through 55) or specific exons in this region of the dystrophin gene (for example, targeting exon 51 only). Efforts have been made to restore the disrupted dystrophin reading frame in DMD patients by skipping non-essential exon(s) (for example, exon 51 skipping) during mRNA splicing to produce internally deleted but functional dystrophin proteins. The deletion of internal dystrophin exon(s) (for example, deletion of exon 51) retains the proper reading frame but cause the less severe Becker muscular dystrophy (BMD). The BMD genotype is similar to DMD in that deletions are present in the dystrophin gene. However, the deletions in BMD leave the reading frame intact. Thus an internally truncated but partially functional dystrophin protein is created. BMD has a wide array of phenotypes, but often if deletions are between exons 45-55 of dystrophin, the phenotype is much milder compared to DMD. Thus changing a DMD genotype to a BMD genotype is a common strategy to correct dystrophin. There are many strategies to correct dystrophin, many of which rely on restoring the reading frame of the endogenous dystrophin. This shifts the disease genotype from DMD to Becker muscular dystrophy. Many BMD patients have intragenic deletions that maintain the translational reading frame, leading to a shorter but largely functional dystrophin protein.

[00077] In certain embodiments, modification of exons 45-55 (such as deletion or excision of exons 45 through 55 by, for example, NHEJ) to restore reading frame ameliorates the phenotype DMD in subjects, including DMD subjects with deletion mutations. Exons 45 through 55 of a dystrophin gene refers to the 45th exon, 46th exon, 47th exon, 48th exon, 49th exon, 50th exon, 51st exon, 52nd exon, 53rd exon, 54th exon, and the 55th exon of the dystrophin gene. Mutations in the 45th through 55th exon region are ideally suited for permanent correction by NHEJ-based genome editing.

[00078] The presently disclosed genetic constructs can generate deletions in the dystrophin gene. The dystrophin gene may be a human dystrophin gene. In certain embodiments, the vector is configured to form two double stand breaks (a first double strand break and a second double strand break) in two introns (a first intron and a second intron) flanking a target position of the dystrophin gene, thereby deleting a segment of the dystrophin gene comprising the dystrophin target position. A "dystrophin target position" can

be a dystrophin exonic target position or a dystrophin intra-exonic target position, as described herein. Deletion of the dystrophin exonic target position can optimize the dystrophin sequence of a subject suffering from Duchenne muscular dystrophy. For example, it can increase the function or activity of the encoded dystrophin protein, and/or result in an improvement in the disease state of the subject. In certain embodiments, excision of the dystrophin exonic target position restores reading frame. The dystrophin exonic target position can comprise one or more exons of the dystrophin gene. In certain embodiments, the dystrophin target position comprises exon 51 of the dystrophin gene (e.g., human dystrophin gene).

[00079] A presently disclosed genetic construct can mediate highly efficient gene editing at the exon 45 through exon 55 region of a dystrophin gene. A presently disclosed genetic construct can restore dystrophin protein expression in cells from DMD patients.

[00080] Elimination of exons 45 through 55 from the dystrophin transcript by exon skipping can be used to treat approximately 50% of all DMD patients. This class of dystrophin mutations is suited for permanent correction by NHEJ-based genome editing and HDR. The genetic constructs described herein have been developed for targeted modification of exon 45 through exon 55 in the human dystrophin gene. A presently disclosed genetic construct may be transfected into human DMD cells and mediate efficient gene modification and conversion to the correct reading frame. Protein restoration may be concomitant with frame restoration and detected in a bulk population of CRISPR/Cas9-based gene editing system-treated cells.

b. CRISPR System

[00081] A presently disclosed genetic construct may encode a CRISPR/Cas9-based gene editing system that is specific for a dystrophin gene. "Clustered Regularly Interspaced Short Palindromic Repeats" and "CRISPRs", as used interchangeably herein, refers to loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea. The CRISPR system is a microbial nuclease system involved in defense against invading phages and plasmids that provides a form of acquired immunity. The CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage. Short segments of foreign DNA, called spacers, are incorporated into the genome between CRISPR repeats, and serve as a 'memory' of past exposures. Cas9 forms a complex with the 3' end of the sgRNA (also referred interchangeably herein as "gRNA"), and the protein-

RNA pair recognizes its genomic target by complementary base pairing between the 5' end of the sgRNA sequence and a predefined 20 bp DNA sequence, known as the protospacer. This complex is directed to homologous loci of pathogen DNA via regions encoded within the crRNA, i.e., the protospacers, and protospacer-adjacent motifs (PAMs) within the pathogen genome. The non-coding CRISPR array is transcribed and cleaved within direct repeats into short crRNAs containing individual spacer sequences, which direct Cas nucleases to the target site (protospacer). By simply exchanging the 20 bp recognition sequence of the expressed sgRNA, the Cas9 nuclease can be directed to new genomic targets. CRISPR spacers are used to recognize and silence exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

[00082] Three classes of CRISPR systems (Types I, II, and III effector systems) are known. The Type II effector system carries out targeted DNA double-strand break in four sequential steps, using a single effector enzyme, Cas9, to cleave dsDNA. Compared to the Type I and Type III effector systems, which require multiple distinct effectors acting as a complex, the Type II effector system may function in alternative contexts such as eukaryotic cells. The Type II effector system consists of a long pre-crRNA, which is transcribed from the spacer-containing CRISPR locus, the Cas9 protein, and a tracrRNA, which is involved in pre-crRNA processing. The tracrRNAs hybridize to the repeat regions separating the spacers of the pre-crRNA, thus initiating dsRNA cleavage by endogenous RNase III. This cleavage is followed by a second cleavage event within each spacer by Cas9, producing mature crRNAs that remain associated with the tracrRNA and Cas9, forming a Cas9:crRNA-tracrRNA complex.

[00083] The Cas9:crRNA-tracrRNA complex unwinds the DNA duplex and searches for sequences matching the crRNA to cleave. Target recognition occurs upon detection of complementarity between a "protospacer" sequence in the target DNA and the remaining spacer sequence in the crRNA. Cas9 mediates cleavage of target DNA if a correct protospacer-adjacent motif (PAM) is also present at the 3' end of the protospacer. For protospacer targeting, the sequence must be immediately followed by the protospacer-adjacent motif (PAM), a short sequence recognized by the Cas9 nuclease that is required for DNA cleavage. Different Type II systems have differing PAM requirements. The *S. pyogenes* CRISPR system may have the PAM sequence for this Cas9 (SpCas9) as 5'-NRG-3', where R is either A or G, and characterized the specificity of this system in human cells. A unique capability of the CRISPR/Cas9-based gene editing system is the straightforward ability to simultaneously target multiple distinct genomic loci by co-expressing a single Cas9 protein with two or more sgRNAs. For example, the *Streptococcus 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 (Hsu et al., Nature Biotechnology (2013) doi:10.1038/nbt.2647). 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 (Esvelt et al. Nature Methods (2013) doi:10.1038/nmeth.2681).

[00084] A Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) (SEQ ID NO: 25) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) (SEQ ID NO: 26) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRT (R = A or G) (SEQ ID NO: 27) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G) (SEQ ID NO: 28) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In the aforementioned embodiments, N can be any nucleotide residue, e.g., any of A, G, C, or T. Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

i) CRISPR/Cas9-based Gene Editing System

[00085] An engineered form of the Type II effector system of *Streptococcus pyogenes* was shown to function in human cells for genome engineering. In this system, the Cas9 protein was directed to genomic target sites by a synthetically reconstituted "guide RNA" ("gRNA", also used interchangeably herein as a chimeric single guide RNA ("sgRNA")), which is a crRNA- tracrRNA fusion that obviates the need for RNase III and crRNA processing in general. Provided herein are CRISPR/Cas9-based engineered systems for use in genome editing and treating genetic diseases. The CRISPR/Cas9-based engineered systems can be designed to target any gene, including genes involved in a genetic disease, aging, tissue regeneration, or wound healing. The CRISPR/Cas9-based gene editing systems can include a Cas9 protein or Cas9 fusion protein and at least one gRNA. In certain embodiments, the system comprises two gRNA molecules. The Cas9 fusion protein may, for example, include a domain that has a different activity that what is endogenous to Cas9, such as a transactivation domain.

[00086] The target gene (e.g., a dystrophin gene, e.g., human dystrophin gene) can be involved in differentiation of a cell or any other process in which activation of a gene can be desired, or can have a mutation such as a frameshift mutation or a nonsense mutation. If the target gene has a mutation that causes a premature stop codon, an aberrant splice acceptor site or an aberrant splice donor site, the CRISPR/Cas9-based gene editing system can be designed to recognize and bind a nucleotide sequence upstream or downstream from the premature stop codon, the aberrant splice acceptor site or the aberrant splice donor site. The CRISPR-Cas9-based system can also be used to disrupt normal gene splicing by targeting splice acceptors and donors to induce skipping of premature stop codons or restore a disrupted reading frame. The CRISPR/Cas9-based gene editing system may or may not mediate off-target changes to protein-coding regions of the genome.

(1) Cas9 molecules and Cas9 fusion proteins

[00087] The CRISPR/Cas9-based gene editing system can include a Cas9 protein or a Cas9 fusion protein. Cas9 protein is an endonuclease that cleaves nucleic acid and is encoded by the CRISPR loci and is involved in the Type II CRISPR system. The Cas9 protein can be from any bacterial or archaea species, including, but not limited to, *Streptococcus pyogenes*, *Staphylococcus aureus* (*S. aureus*), *Acidovorax avenae, Actinobacillus pleuropneumoniae, Actinobacillus succinogenes, Actinobacillus suis, Actinomyces sp., cycliphilus denitrificans, Aminomonas paucivorans, Bacillus cereus, Bacillus smithii, Bacillus thuringiensis, Bacteroides sp., Blastopirellula marina, Bradyrhizobium sp., Brevibacillus laterosporus, Campylobacter coli, Campylobacter jejuni, Campylobacter lari, Candidatus Puniceispirillum, Clostridium cellulolyticum, Clostridium perfringens, Corynebacterium accolens, Corynebacterium diphtheria, Corynebacterium matruchotii, Dinoroseobacter shibae, Eubacterium dolichum, gamma proteobacterium, Gluconacetobacter diazotrophicus, Haemophilus parainfluenzae, Haemophilus sputorum, Helicobacter canadensis, Helicobacter cinaedi, Helicobacter mustelae, Ilyobacter polytropus, Kingella kingae, Lactobacillus crispatus, Listeria ivanovii, Listeria monocytogenes, Listeriaceae bacterium, Methylocystis sp., Methylosinus trichosporium, Mobiluncus mulieris, Neisseria bacilliformis, Neisseria cinerea, Neisseria flavescens, Neisseria lactamica, Neisseria sp., Neisseria wadsworthii, Nitrosomonas sp., Parvibaculum lavamentivorans, Pasteurella multocida, Phascolarctobacterium succinatutens, Ralstonia syzygii, Rhodopseudomonas palustris, Rhodovulum sp., Simonsiella muelleri, Sphingomonas sp., Sporolactobacillus vineae, Staphylococcus lugdunensis, Streptococcus sp., Subdoligranulum sp., Tistrella mobilis, Treponema sp.,* or *Verminephrobacter eiseniae*. In certain embodiments, the Cas9 molecule is a *Streptococcus pyogenes* Cas9 molecule (also

referred herein as "SpCas9"). In certain embodiments, the Cas9 molecule is a *Staphylococcus aureus* Cas9 molecule (also referred herein as "SaCas9").

[00088] A Cas9 molecule or a Cas9 fusion protein can interact with one or more gRNA molecule and, in concert with the gRNA molecule(s), localizes to a site which comprises a target domain, and in certain embodiments, a PAM sequence. The ability of a Cas9 molecule or a Cas9 fusion protein to recognize a PAM sequence can be determined, for example, using a transformation assay as known in the art.

[00089] In certain embodiments, the ability of a Cas9 molecule or a Cas9 fusion protein to interact with and cleave a target nucleic acid is PAM sequence dependent. A PAM sequence is a sequence in the target nucleic acid. In certain embodiments, cleavage of the target nucleic acid occurs upstream from the PAM sequence. Cas9 molecules from different bacterial species can recognize different sequence motifs (e.g., PAM sequences). In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) (SEQ ID NO: 25) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) (SEQ ID NO: 26) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRT (R = A or G) (SEQ ID NO: 27) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G; V = A or C or G) (SEQ ID NO: 28) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In the aforementioned embodiments, N can be any nucleotide residue, e.g., any of A, G, C, or T. Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

[00090] In certain embodiments, the vector encodes at least one Cas9 molecule that recognizes a Protospacer Adjacent Motif (PAM) of either NNGRRT (SEQ ID NO: 27) or NNGRRV (SEQ ID NO: 28). In certain embodiments, the at least one Cas9 molecule is an *S. aureus* Cas9 molecule. In certain embodiments, the at least one Cas9 molecule is a mutant *S. aureus* Cas9 molecule.

[00091] Additionally or alternatively, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art.

[00092] Exemplary codon optimized nucleic acid sequences encoding a Cas9 molecule of *S. aureus*, and optionally containing nuclear localization sequences (NLSs), are set forth in SEQ ID NOs: 31-37. Another exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. aureus* comprises the nucleotides 1293-4451 of SEQ ID NO: 38.

[00093] In some embodiments, the nucleotide sequence encoding a *S. aureus* Cas9 molecule includes the polynucleotide sequence of SEQ ID NO: 37. An amino acid sequence of an *S. aureus* Cas9 molecule is set forth in SEQ ID NO: 39. An amino acid sequence of an *S. aureus* Cas9 molecule is set forth in SEQ ID NO: 40.

[00094] Alternatively or additionally, the CRISPR/Cas9-based gene editing system can include a fusion protein. The fusion protein can comprise two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein and the second polypeptide domain has an activity such as transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, nucleic acid association activity, methylase activity, or demethylase activity. The fusion protein can include a Cas9 protein or a mutated Cas9 protein, fused to a second polypeptide domain that has an activity such as transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, nucleic acid association activity, methylase activity, or demethylase activity.

(a) Transcription Activation Activity

[00095] The second polypeptide domain can have transcription activation activity, i.e., a transactivation domain. For example, gene expression of endogenous mammalian genes, such as human genes, can be achieved by targeting a fusion protein of iCas9 and a transactivation domain to mammalian promoters via combinations of gRNAs. The transactivation domain can include a p300 protein, VP16 protein, multiple VP16 proteins, such as a VP48 domain or VP64 domain, or p65 domain of NF kappa B transcription activator activity. For example, the fusion protein may be dCas9-VP64 or dCas9-p300.

(b) Transcription Repression Activity

[00096] The second polypeptide domain can have transcription repression activity. The second polypeptide domain can have a Kruppel associated box activity, such as a KRAB domain, ERF repressor domain activity, Mxil repressor domain activity, SID4X repressor domain activity, Mad-SID repressor domain activity or TATA box binding protein activity. For example, the fusion protein may be dCas9-KRAB.

(c) Transcription Release Factor Activity

[00097] The second polypeptide domain can have transcription release factor activity. The second polypeptide domain can have eukaryotic release factor 1 (ERF1) activity or eukaryotic release factor 3 (ERF3) activity.

(d) Histone Modification Activity

[00098] The second polypeptide domain can have histone modification activity. The second polypeptide domain can have histone deacetylase, histone acetyltransferase, histone demethylase, or histone methyltransferase activity. The histone acetyltransferase may be p300 or CREB-binding protein (CBP) protein, or fragments thereof. For example, the fusion protein may be dCas9-p300.

(e) Nuclease Activity

[00099] The second polypeptide domain can have nuclease activity that is different from the nuclease activity of the Cas9 protein. A nuclease, or a protein having nuclease activity, is an enzyme capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids. Nucleases are usually further divided into endonucleases and exonucleases, although some of the enzymes may fall in both categories. Well known nucleases are deoxyribonuclease and ribonuclease.

(f) Nucleic Acid Association Activity

[000100] The second polypeptide domain can have nucleic acid association activity or nucleic acid binding protein-DNA-binding domain (DBD) is an independently folded protein domain that contains at least one motif that recognizes double- or single-stranded DNA. A DBD can recognize a specific DNA sequence (a recognition sequence) or have a general affinity to DNA. nucleic acid association region selected from the group consisting of helix-turn-helix region, leucine zipper region, winged helix region, winged helix-turn-helix region, helix-loop-helix region, immunoglobulin fold, B3 domain, Zinc finger, HMG-box, Wor3 domain, TAL effector DNA-binding domain.

(g) Methylase Activity

[000101] The second polypeptide domain can have methylase activity, which involves transferring a methyl group to DNA, RNA, protein, small molecule, cytosine or adenine. The second polypeptide domain may include a DNA methyltransferase.

(h) Demethylase Activity

[000102] The second polypeptide domain can have demethylase activity. The second polypeptide domain can include an enzyme that remove methyl (CH3-) groups from nucleic acids, proteins (in particular histones), and other molecules. Alternatively, the second polypeptide can covert the methyl group to hydroxymethylcytosine in a mechanism for demethylating DNA. The second polypeptide can catalyze this reaction. For example, the second polypeptide that catalyzes this reaction can be Tet1.

(2) gRNA Targeting the Dystrophin Gene

[000103] The CRISPR/Cas9-based gene editing system includes at least one gRNA molecule, for example, two gRNA molecules. The gRNA provides the targeting of a CRISPR/Cas9-based gene editing system. The gRNA is a fusion of two noncoding RNAs: a crRNA and a tracrRNA. The sgRNA may target any desired DNA sequence by exchanging the sequence encoding a 20 bp protospacer which confers targeting specificity through complementary base pairing with the desired DNA target. The gRNA mimics the naturally occurring crRNA:tracrRNA duplex involved in the Type II Effector system. This duplex, which may include, for example, a 42-nucleotide crRNA and a 75-nucleotide tracrRNA, acts as a guide for the Cas9 to cleave the target nucleic acid. The "target region", "target sequence," or "protospacer" may be used interchangeably herein and refers to the region of the target gene (e.g., a dystrophin gene) to which the CRISPR/Cas9-based gene editing system targets. The CRISPR/Cas9-based gene editing system may include at least one gRNA, wherein each gRNA targets a different DNA sequence. The target DNA sequences may be overlapping. The target sequence or protospacer is followed by a PAM sequence at the 3' end of the protospacer. Different Type II systems have differing PAM requirements. For example, the *Streptococcus pyogenes* Type II system uses an "NGG" sequence, where "N" can be any nucleotide. In some embodiments, the PAM sequence may be "NGG", where "N" can be any nucleotide. In some embodiments, the PAM sequence may be NNGRRT (SEQ ID NO: 27) or NNGRRV (SEQ ID NO: 28).

[000104] The number of gRNA molecules encoded by a presently disclosed genetic construct (e.g., an AAV vector) can be at least 1 gRNA, at least 2 different gRNAs, at least 3 different gRNAs, at least 4 different gRNAs, at least 5 different gRNAs, at least 6 different gRNAs, at least 7 different gRNAs, at least 8 different gRNAs, at least 9 different gRNAs, at least 10 different gRNAs, at least 11 different gRNAs, at least 12 different gRNAs, at least 13 different gRNAs, at least 14 different gRNAs, at least 15 different gRNAs, at least 16 different gRNAs, at least 17 different gRNAs, at least 18 different gRNAs, at least 18

different gRNAs, at least 20 different gRNAs, at least 25 different gRNAs, at least 30 different gRNAs, at least 35 different gRNAs, at least 40 different gRNAs, at least 45 different gRNAs, or at least 50 different gRNAs. The number of gRNA molecules encoded by a presently disclosed genetic construct can be less than 50 gRNAs, less than 45 different gRNAs, less than 40 different gRNAs, less than 35 different gRNAs, less than 30 different gRNAs, less than 25 different gRNAs, less than 20 different gRNAs, less than 19 different gRNAs, less than 18 different gRNAs, less than 17 different gRNAs, less than 16 different gRNAs, less than 15 different gRNAs, less than 14 different gRNAs, less than 13 different gRNAs less than 12 different gRNAs, less than 11 different gRNAs, less than 10 different gRNAs, less than 9 different gRNAs, less than 8 different gRNAs, less than 7 different gRNAs, less than 6 different gRNAs, less than 5 different gRNAs, less than 4 different gRNAs, or less than 3 different gRNAs. The number of gRNAs encoded by a presently disclosed genetic construct can be between at least 1 gRNA to at least 50 different gRNAs, at least 1 gRNA to at least 45 different gRNAs, at least 1 gRNA to at least 40 different gRNAs, at least 1 gRNA to at least 35 different gRNAs, at least 1 gRNA to at least 30 different gRNAs, at least 1 gRNA to at least 25 different gRNAs, at least 1 gRNA to at least 20 different gRNAs, at least 1 gRNA to at least 16 different gRNAs, at least 1 gRNA to at least 12 different gRNAs, at least 1 gRNA to at least 8 different gRNAs, at least 1 gRNA to at least 4 different gRNAs, at least 4 gRNAs to at least 50 different gRNAs, at least 4 different gRNAs to at least 45 different gRNAs, at least 4 different gRNAs to at least 40 different gRNAs, at least 4 different gRNAs to at least 35 different gRNAs, at least 4 different gRNAs to at least 30 different gRNAs, at least 4 different gRNAs to at least 25 different gRNAs, at least 4 different gRNAs to at least 20 different gRNAs, at least 4 different gRNAs to at least 16 different gRNAs, at least 4 different gRNAs to at least 12 different gRNAs, at least 4 different gRNAs to at least 8 different gRNAs, at least 8 different gRNAs to at least 50 different gRNAs, at least 8 different gRNAs to at least 45 different gRNAs, at least 8 different gRNAs to at least 40 different gRNAs, at least 8 different gRNAs to at least 35 different gRNAs, 8 different gRNAs to at least 30 different gRNAs, at least 8 different gRNAs to at least 25 different gRNAs, 8 different gRNAs to at least 20 different gRNAs, at least 8 different gRNAs to at least 16 different gRNAs, or 8 different gRNAs to at least 12 different gRNAs. In certain embodiments, the genetic construct (e.g., an AAV vector) encodes one gRNA molecule, i.e., a first gRNA molecule, and optionally a Cas9 molecule. In certain embodiments, a first genetic construct (e.g., a first AAV vector) encodes one gRNA molecule, i.e., a first gRNA molecule, and optionally a Cas9 molecule, and a second genetic construct (e.g., a second AAV vector) encodes one gRNA molecule, i.e., a second gRNA molecule, and optionally a Cas9 molecule.

[000105] The gRNA molecule comprises a targeting domain (also referred to as a targeting sequence), which is a complementary polynucleotide sequence of the target DNA sequence followed by a PAM sequence. The gRNA may comprise a "G" at the 5' end of the targeting domain or complementary polynucleotide sequence. The targeting domain of a gRNA molecule may comprise at least a 10 base pair, at least a 11 base pair, at least a 12 base pair, at least a 13 base pair, at least a 14 base pair, at least a 15 base pair, at least a 16 base pair, at least a 17 base pair, at least a 18 base pair, at least a 19 base pair, at least a 20 base pair, at least a 21 base pair, at least a 22 base pair, at least a 23 base pair, at least a 24 base pair, at least a 25 base pair, at least a 30 base pair, or at least a 35 base pair complementary polynucleotide sequence of the target DNA sequence followed by a PAM sequence. The targeting domain of a gRNA molecule may comprise less than a 40 base pair, less than a 35 base pair, less than a 30 base pair, less than a 25 base pair, less than a 20 base pair, less than a 19 base pair, less than a 18 base pair, less than a 17 base pair, less than a 16 base pair, less than a 15 base pair, less than a 14 base pair, less than a 13 base pair, less than a 12 base pair, less than a 11 base pair, or less than a 10 base pair complementary polynucleotide sequence of the target DNA sequence followed by a PAM sequence. In certain embodiments, the targeting domain of a gRNA molecule has 19-25 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 20 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 21 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 22 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 23 nucleotides in length.

[000106] The gRNA may target a region of the dystrophin gene (DMD). In certain embodiments, the gRNA can target at least one of exons, introns, the promoter region, the enhancer region, the transcribed region of the dystrophin gene. In certain embodiments, the gRNA molecule targets intron 44 of the human dystrophin gene. In certain embodiments, the gRNA molecule targets intron 55 of the human dystrophin gene. In some embodiments, a first gRNA and a second gRNA each target an intron of a human dystrophin gene such that exons 45 through 55 are deleted. A gRNA may bind and target a polynucleotide sequence corresponding to SEQ ID NO: 2 or a fragment thereof or a complement thereof. A gRNA may be encoded by a polynucleotide sequence comprising SEQ ID NO: 2 or a fragment thereof or a complement thereof. The targeting sequence of the gRNA may comprise the polynucleotide of SEQ ID NO: 2 or a fragment thereof, such as a 5' truncation thereof, or a complement thereof. Truncations may be, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides shorter than SEQ ID NO: 2. In some embodiments, the gRNA may bind and target the polynucleotide of SEQ ID NO: 2. In some embodiments, the

gRNA may bind and target a 5' truncation of the polynucleotide of SEQ ID NO: 2. A gRNA may bind and target a polynucleotide sequence corresponding to SEQ ID NO: 3 or a fragment thereof or a complement thereof. A gRNA may be encoded by a polynucleotide sequence comprising SEQ ID NO: 3 or a fragment thereof or a complement thereof. The targeting sequence of the gRNA may comprise the polynucleotide of SEQ ID NO: 3 or a fragment thereof, such as a 5' truncation thereof, or a complement thereof. Truncations may be, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides shorter than SEQ ID NO: 3. In some embodiments, the gRNA may bind and target the polynucleotide of SEQ ID NO: 3. In some embodiments, the gRNA may bind and target a 5' truncation of the polynucleotide of SEQ ID NO: 3. In some embodiments, a gRNA that binds and targets or is encoded by a polynucleotide sequence comprising or corresponding to SEQ ID NO: 2 or truncation thereof is paired with a gRNA that binds and targets or is encoded by a polynucleotide sequence comprising or corresponding to SEQ ID NO: 3 or truncation thereof.

[000107] Single or multiplexed gRNAs can be designed to restore the dystrophin reading frame by targeting the mutational hotspot in exons 45-55 of dystrophin. Following treatment with a presently disclosed vector, dystrophin expression can be restored in Duchenne patient muscle cells *in vitro*. Human dystrophin was detected *in vivo* following transplantation of genetically corrected patient cells into immunodeficient mice. Significantly, the unique multiplex gene editing capabilities of the CRISPR/Cas9-based gene editing system enable efficiently generating large deletions of this mutational hotspot region that can correct up to 62% of patient mutations by universal or patient- specific gene editing approaches. In some embodiments, candidate gRNAs are evaluated and chosen based on off-target activity, on-target activity as measured by surveyor, and distance from the exon.

(3) gRNA Scaffold

[000108] The CRISPR/Cas9-based gene editing system includes at least one gRNA scaffold. The gRNA scaffold facilitates Cas9 binding to the gRNA and endonuclease activity. The gRNA scaffold is a polynucleotide sequence that follows the gRNA targeting sequence. Together, the gRNA targeting sequence and gRNA scaffold form one polynucleotide. In some embodiments, the gRNA scaffold comprises the polynucleotide sequence of SEQ ID NO: 4, or a complement thereof. In some embodiments, the gRNA scaffold is encoded by the polynucleotide sequence of SEQ ID NO: 4, or a complement thereof. In some embodiments, the gRNA comprises a polynucleotide that targets a sequence of SEQ ID NO: 2 or SEQ ID NO: 3 or a truncation thereof, and a polynucleotide corresponding to or encoded by the gRNA scaffold of SEQ ID NO: 4.

3. DNA Targeting Compositions

[000109] Further disclosed herein are DNA targeting compositions that comprise such genetic constructs. The DNA targeting compositions include at least one gRNA molecule (for example, two gRNA molecules) that targets a dystrophin gene (for example, human dystrophin gene), as described above. The at least one gRNA molecule can bind and recognize a target region. The target regions can be chosen immediately upstream of possible out-of-frame stop codons such that insertions or deletions during the repair process restore the dystrophin reading frame by frame conversion. Target regions can also be splice acceptor sites or splice donor sites, such that insertions or deletions during the repair process disrupt splicing and restore the dystrophin reading frame by splice site disruption and exon exclusion. Target regions can also be aberrant stop codons such that insertions or deletions during the repair process restore the dystrophin reading frame by eliminating or disrupting the stop codon.

[000110] In certain embodiments, the presently disclosed DNA targeting composition includes a first gRNA and a second gRNA. The first gRNA molecule and the second gRNA molecule may bind or target a polynucleotide of SEQ ID NO:2 and SEQ ID NO:3, respectively, or a truncation or a complement thereof. The first gRNA molecule and the second gRNA molecule may comprise a polynucleotide corresponding to SEQ ID NO:2 and SEQ ID NO:3, respectively, or a truncation or a complement thereof.

[000111] The deletion efficiency of the presently disclosed vectors can be related to the deletion size, i.e., the size of the segment deleted by the vectors. In certain embodiments, the length or size of specific deletions is determined by the distance between the PAM sequences in the gene being targeted (e.g., a dystrophin gene). In certain embodiments, a specific deletion of a segment of the dystrophin gene, which is defined in terms of its length and a sequence it comprises (e.g., exon 51), is the result of breaks made adjacent to specific PAM sequences within the target gene (e.g., a dystrophin gene).

[000112] In certain embodiments, the deletion size is about 50 to about 2,000 base pairs (bp), e.g., about 50 to about 1999 bp, about 50 to about 1900 bp, about 50 to about 1800 bp, about 50 to about 1700 bp, about 50 to about 1650 bp, about 50 to about 1600 bp, about 50 to about 1500 bp, about 50 to about 1400 bp, about 50 to about 1300 bp, about 50 to about 1200 bp, about 50 to about 1150 bp, about 50 to about 1100 bp, about 50 to about 1000 bp, about 50 to about 900 bp, about 50 to about 850 bp, about 50 to about 800 bp, about 50 to about 750 bp, about 50 to about 700 bp, about 50 to about 600 bp, about 50 to about 500 bp, about 50 to about 400 bp, about 50 to about 350 bp, about 50 to about 300 bp, about 50

to about 250 bp, about 50 to about 200 bp, about 50 to about 150 bp, about 50 to about 100 bp, about 100 to about 1999 bp, about 100 to about 1900 bp, about 100 to about 1800 bp, about 100 to about 1700 bp, about 100 to about 1650 bp, about 100 to about 1600 bp, about 100 to about 1500 bp, about 100 to about 1400 bp, about 100 to about 1300 bp, about 100 to about 1200 bp, about 100 to about 1150 bp, about 100 to about 1100 bp, about 100 to about 1000 bp, about 100 to about 900 bp, about 100 to about 850 bp, about 100 to about 800 bp, about 100 to about 750 bp, about 100 to about 700 bp, about 100 to about 600 bp, about 100 to about 1000 bp, about 100 to about 400 bp, about 100 to about 350 bp, about 100 to about 300 bp, about 100 to about 250 bp, about 100 to about 200 bp, about 100 to about 150 bp, about 200 to about 1999 bp, about 200 to about 1900 bp, about 200 to about 1800 bp, about 200 to about 1700 bp, about 200 to about 1650 bp, about 200 to about 1600 bp, about 200 to about 1500 bp, about 200 to about 1400 bp, about 200 to about 1300 bp, about 200 to about 1200 bp, about 200 to about 1150 bp, about 200 to about 1100 bp, about 200 to about 1000 bp, about 200 to about 900 bp, about 200 to about 850 bp, about 200 to about 800 bp, about 200 to about 750 bp, about 200 to about 700 bp, about 200 to about 600 bp, about 200 to about 2000 bp, about 200 to about 400 bp, about 200 to about 350 bp, about 200 to about 300 bp, about 200 to about 250 bp, about 300 to about 1999 bp, about 300 to about 1900 bp, about 300 to about 1800 bp, about 300 to about 1700 bp, about 300 to about 1650 bp, about 300 to about 1600 bp, about 300 to about 1500 bp, about 300 to about 1400 bp, about 300 to about 1300 bp, about 300 to about 1200 bp, about 300 to about 1150 bp, about 300 to about 1100 bp, about 300 to about 1000 bp, about 300 to about 900 bp, about 300 to about 850 bp, about 300 to about 800 bp, about 300 to about 750 bp, about 300 to about 700 bp, about 300 to about 600 bp, about 300 to about 3000 bp, about 300 to about 400 bp, or about 300 to about 350 bp. In certain embodiments, the deletion size can be about 118 base pairs, about 233 base pairs, about 326 base pairs, about 766 base pairs, about 805 base pairs, or about 1611 base pairs.

4. Compositions for Genome Editing in Muscle

[000113] Disclosed herein is a genetic construct or a composition thereof for genome editing a target gene in a subject, such as, for example, a target gene in skeletal muscle and/or cardiac muscle of a subject. The genetic construct may be a vector. The vector may be a modified AAV vector. The composition may include a polynucleotide sequence encoding a CRISPR/Cas9-based gene editing system. The composition may deliver active forms of CRISPR/Cas9-based gene editing systems to skeletal muscle or cardiac muscle. The presently disclosed genetic constructs can be used in correcting or reducing the effects of mutations in the dystrophin gene involved in genetic diseases and/or other skeletal or

cardiac muscle conditions, such as, for example, DMD. The composition may further comprise a donor DNA or a transgene. These compositions may be used in genome editing, genome engineering, and correcting or reducing the effects of mutations in genes involved in genetic diseases and/or other skeletal and/or cardiac muscle conditions.

a. CRISPR/Cas9-based gene editing system for targeting dystrophin

[000114] A CRISPR/Cas9-based gene editing system specific for dystrophin gene is disclosed herein. The CRISPR/Cas9-based gene editing system may include Cas9 and at least one gRNA to target the dystrophin gene. The CRISPR/Cas9-based gene editing system may bind and recognize a target region. The target regions may be chosen immediately upstream of possible out-of-frame stop codons such that insertions or deletions during the repair process restore the dystrophin reading frame by frame conversion. Target regions may also be splice acceptor sites or splice donor sites, such that insertions or deletions during the repair process disrupt splicing and restore the dystrophin reading frame by splice site disruption and exon exclusion. Target regions may also be aberrant stop codons such that insertions or deletions during the repair process restore the dystrophin reading frame by eliminating or disrupting the stop codon. Target regions may include an intron of the dystrophin gene. Target regions may include an exon of the dystrophin gene.

b. Adeno-Associated Virus Vectors

[000115] The composition may also include a viral delivery system. In certain embodiments, the vector is an adeno-associated virus (AAV) vector. The AAV vector is a small virus belonging to the genus Dependovirus of the Parvoviridae family that infects humans and some other primate species. AAV vectors may be used to deliver CRISPR/Cas9-based gene editing systems using various construct configurations. For example, AAV vectors may deliver Cas9 and gRNA expression cassettes on separate vectors or on the same vector. Alternatively, if the small Cas9 proteins, derived from species such as *Staphylococcus aureus* or *Neisseria meningitidis*, are used then both the Cas9 and up to two gRNA expression cassettes may be combined in a single AAV vector within the 4.7 kb packaging limit.

[000116] In certain embodiments, the AAV vector is a modified AAV vector. The modified AAV vector may have enhanced cardiac and skeletal muscle tissue tropism. The modified AAV vector may be capable of delivering and expressing the CRISPR/Cas9-based gene editing system 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 nucleases to skeletal and cardiac muscle in vivo. The modified AAV vector may be based on one or more of several capsid types, including AAV1, 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 (Seto et al. Current Gene Therapy (2012) 12:139-151). The modified AAV vector may be AAV2i8G9 (Shen et al. J. Biol. Chem. (2013) 288:28814-28823). The AAV vector may be AAVrh74.

5. Methods

a. Methods of Genome Editing in Muscle

[000117] Disclosed herein are methods of genome editing in subject. The genome editing may be in a skeletal muscle and/or cardiac muscle of a subject. The method may comprise administering to the skeletal muscle and/or cardiac muscle of the subject the system or composition for genome editing, as described above. The genome editing may include correcting a mutant gene or inserting a transgene. Correcting the mutant gene may include deleting, rearranging, or replacing the mutant gene. Correcting the mutant gene may include nuclease-mediated NHEJ or HDR.

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

[000118] Disclosed herein are methods of correcting a mutant gene (e.g., a mutant dystrophin gene, e.g., a mutant human dystrophin gene) in a cell and treating a subject suffering from a genetic disease, such as DMD. The method can include administering to a cell or a subject a presently disclosed system or genetic construct (e.g., a vector) or a composition comprising thereof as described above. The method can comprise administering to the skeletal muscle and/or cardiac muscle of the subject the presently disclosed system or genetic construct (e.g., a vector) or a composition comprising the same for genome editing in skeletal muscle and/or cardiac muscle, as described above. Use of the presently disclosed system or genetic construct (e.g., a vector) or a composition comprising the same to deliver the CRISPR/Cas9-based gene editing system to the skeletal muscle or cardiac muscle may restore the expression of a fully-functional or partially-functional protein with a repair template or donor DNA, which can replace the entire gene or the region containing the mutation. The CRISPR/Cas9-based gene editing system may be used to introduce site-specific double strand breaks at targeted genomic loci. Site-specific double-strand breaks are created when the CRISPR/Cas9-based gene editing system binds

to a target DNA sequences, thereby permitting cleavage of the target DNA. This DNA cleavage may stimulate the natural DNA-repair machinery, leading to one of two possible repair pathways: homology-directed repair (HDR) or the non-homologous end joining (NHEJ) pathway.

[000119] Provided herein is genome editing with a CRISPR/Cas9-based gene editing system without a repair template, which can efficiently correct the reading frame and restore the expression of a functional protein involved in a genetic disease. The disclosed CRISPR/Cas9-based gene editing systems may involve using homology-directed repair or nuclease-mediated non-homologous end joining (NHEJ)-based correction approaches, which enable efficient correction in proliferation-limited primary cell lines that may not be amenable to homologous recombination or selection-based gene correction. This strategy integrates the rapid and robust assembly of active CRISPR/Cas9-based gene editing systems with an efficient gene editing method for the treatment of genetic diseases caused by mutations in nonessential coding regions that cause frameshifts, premature stop codons, aberrant splice donor sites or aberrant splice acceptor sites.

i) Nuclease mediated non-homologous end joining

[000120] Restoration of protein expression from an endogenous mutated gene may be through template-free NHEJ-mediated DNA repair. In contrast to a transient method targeting the target gene RNA, the correction of the target gene reading frame in the genome by a transiently expressed CRISPR/Cas9-based gene editing system may lead to permanently restored target gene expression by each modified cell and all of its progeny. In certain embodiments, NHEJ is a nuclease mediated NHEJ, which in certain embodiments, refers to NHEJ that is initiated a Cas9 molecule, cuts double stranded DNA. The method comprises administering a presently disclosed genetic construct (e.g., a vector) or a composition comprising thereof to the skeletal muscle or cardiac muscle of the subject for genome editing in skeletal muscle or cardiac muscle.

[000121] Nuclease mediated NHEJ gene correction may correct the mutated target gene and offers several potential advantages over the HDR pathway. For example, NHEJ does not require a donor template, which may cause nonspecific insertional mutagenesis. In contrast to HDR, NHEJ operates efficiently in all stages of the cell cycle and therefore may be effectively exploited in both cycling and post-mitotic cells, such as muscle fibers. This provides a robust, permanent gene restoration alternative to oligonucleotide-based exon skipping or pharmacologic forced read-through of stop codons and could theoretically require as few as one drug treatment. NHEJ-based gene correction using a CRISPR/Cas9-

based gene editing system, as well as other engineered nucleases including meganucleases and zinc finger nucleases, may be combined with other existing ex vivo and in vivo platforms for cell- and gene-based therapies, in addition to the plasmid electroporation approach described here. For example, delivery of a CRISPR/Cas9-based gene editing system by mRNA-based gene transfer or as purified cell permeable proteins could enable a DNA-free genome editing approach that would circumvent any possibility of insertional mutagenesis.

ii) Homology-Directed Repair

[000122] Restoration of protein expression from an endogenous mutated gene may involve homology-directed repair. The method as described above further includes administrating a donor template to the cell. The donor template may include a nucleotide sequence encoding a full-functional protein or a partially-functional protein. For example, the donor template may include a miniaturized dystrophin construct, termed minidystrophin ("minidys"), a full-functional dystrophin construct for restoring a mutant dystrophin gene, or a fragment of the dystrophin gene that after homology-directed repair leads to restoration of the mutant dystrophin gene.

iii) Methods of Correcting a Mutant Gene and Treating a Subject Using CRISPR/Cas9

[000123] The present disclosure is also directed to genome editing with the CRISPR/Cas9-based gene editing system to restore the expression of a full-functional or partially-functional protein with a repair template or donor DNA, which can replace the entire gene or the region containing the mutation. The CRISPR/Cas9-based gene editing system may be used to introduce site-specific double strand breaks at targeted genomic loci. Site-specific double-strand breaks are created when the CRISPR/Cas9-based gene editing system binds to a target DNA sequences using the gRNA, thereby permitting cleavage of the target DNA. The CRISPR/Cas9-based gene editing system has the advantage of advanced genome editing due to their high rate of successful and efficient genetic modification. This DNA cleavage may stimulate the natural DNA-repair machinery, leading to one of two possible repair pathways: homology-directed repair (HDR) or the non-homologous end joining (NHEJ) pathway.

[000124] The present disclosure is directed to genome editing with CRISPR/Cas9-based gene editing system without a repair template, which can efficiently correct the reading frame and restore the expression of a functional protein involved in a genetic disease. The disclosed CRISPR/Cas9-based gene editing system and methods may involve using

homology-directed repair or nuclease-mediated non-homologous end joining (NHEJ)-based correction approaches, which enable efficient correction in proliferation-limited primary cell lines that may not be amenable to homologous recombination or selection-based gene correction. This strategy integrates the rapid and robust assembly of active CRISPR/Cas9-based gene editing system with an efficient gene editing method for the treatment of genetic diseases caused by mutations in nonessential coding regions that cause frameshifts, premature stop codons, aberrant splice donor sites or aberrant splice acceptor sites.

[000125] The present disclosure provides methods of correcting a mutant gene in a cell and treating a subject suffering from a genetic disease, such as DMD. The method may include administering to a cell or subject a CRISPR/Cas9-based gene editing system, a polynucleotide or vector encoding said CRISPR/Cas9-based gene editing system, or composition of said CRISPR/Cas9-based gene editing system as described above. The method may include administering a CRISPR/Cas9-based gene editing system, such as administering a Cas9 protein or Cas9 fusion protein containing a second domain having nuclease activity, a nucleotide sequence encoding said Cas9 protein or Cas9 fusion protein, and/or at least one gRNA, wherein the gRNAs target different DNA sequences. The target DNA sequences may be overlapping. The number of gRNA administered to the cell may be at least 1 gRNA, at least 2 different gRNA, at least 3 different gRNA at least 4 different gRNA, at least 5 different gRNA, at least 6 different gRNA, at least 7 different gRNA, at least 8 different gRNA, at least 9 different gRNA, at least 10 different gRNA, at least 15 different gRNA, at least 20 different gRNA, at least 30 different gRNA, or at least 50 different gRNA, as described above. The method may involve homology-directed repair or non-homologous end joining.

c. Methods of Treating Disease

[000126] The present disclosure is directed to a method of treating a subject in need thereof. The method comprises administering to a tissue of a subject the presently disclosed system or genetic construct (e.g., a vector) or a composition comprising thereof, as described above. In certain embodiments, the method may comprise administering to the skeletal muscle or cardiac muscle of the subject the presently disclosed system or genetic construct (e.g., a vector) or composition comprising thereof, as described above. In certain embodiments, the method may comprise administering to a vein of the subject the presently disclosed system or genetic construct (e.g., a vector) or composition comprising thereof, as described above. In certain embodiments, the subject is suffering from a skeletal muscle or cardiac muscle condition causing degeneration or weakness or a genetic disease. For

example, the subject may be suffering from Duchenne muscular dystrophy, as described above.

i) **Duchenne muscular dystrophy**

[000127] The method, as described above, may be used for correcting the dystrophin gene and recovering full-functional or partially-functional protein expression of said mutated dystrophin gene. In some aspects and embodiments the disclosure provides a method for reducing the effects (e.g., clinical symptoms/indications) of DMD in a patient. In some aspects and embodiments the disclosure provides a method for treating DMD in a patient. In some aspects and embodiments the disclosure provides a method for preventing DMD in a patient. In some aspects and embodiments the disclosure provides a method for preventing further progression of DMD in a patient.

6. Constructs and Plasmids

[000128] The compositions, as described above, may comprise one or more genetic constructs that encode the CRISPR/Cas9-based gene editing system, as disclosed herein. The genetic construct, such as a plasmid, may comprise a nucleic acid that encodes the CRISPR/Cas9-based gene editing system, such as the Cas9 protein and/or Cas9 fusion proteins and/or at least one of the gRNAs. The compositions, as described above, may comprise genetic constructs that encodes the modified AAV vector and a nucleic acid sequence that encodes the CRISPR/Cas9-based gene editing system, as disclosed herein. The genetic construct, such as a plasmid, may comprise a nucleic acid that encodes the CRISPR/Cas9-based gene editing system. The compositions, as described above, may comprise genetic constructs that encodes the modified lentiviral vector, as disclosed herein.

[000129] The genetic construct, such as a recombinant plasmid or recombinant viral particle, may comprise a nucleic acid that encodes the Cas9-fusion protein and at least one gRNA. In some embodiments, the genetic construct may comprise a nucleic acid that encodes the Cas9-fusion protein and at least two different gRNAs. In some embodiments, the genetic construct may comprise a nucleic acid that encodes the Cas9-fusion protein and more than two different gRNAs. In some embodiments, the genetic construct may comprise a promoter that operably linked to the nucleotide sequence encoding the at least one gRNA molecule and/or a Cas9 molecule. In some embodiments, the promoter is operably linked to the nucleotide sequence encoding a first gRNA molecule, a second gRNA molecule, and/or a Cas9 molecule. The genetic construct may be present in the cell as a functioning

extrachromosomal molecule. The genetic construct may be a linear minichromosome including centromere, telomeres or plasmids or cosmids.

[000130] The genetic construct may also be part of a genome of a recombinant viral vector, including recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The genetic construct may be part of the genetic material in attenuated live microorganisms or recombinant microbial vectors which live in cells. The genetic constructs may comprise regulatory elements for gene expression of the coding sequences of the nucleic acid. The regulatory elements may be a promoter, an enhancer, an initiation codon, a stop codon, or a polyadenylation signal.

[000131] In certain embodiments, the genetic construct is a vector. The vector can be an Adeno-associated virus (AAV) vector, which encode at least one Cas9 molecule and at least one gRNA molecule; the vector is capable of expressing the at least one Cas9 molecule and the at least gRNA molecule, in the cell of a mammal. The vector can be a plasmid. The vectors can be used for in vivo gene therapy. The vector may be recombinant. The vector may comprise heterologous nucleic acid encoding the fusion protein, such as the Cas9-fusion protein or CRISPR/Cas9-based gene editing system. The vector may be a plasmid. The vector may be useful for transfecting cells with nucleic acid encoding the Cas9-fusion protein or CRISPR/Cas9-based gene editing system, which the transformed host cell is cultured and maintained under conditions wherein expression of the Cas9-fusion protein or the CRISPR/Cas9-based gene editing system takes place.

[000132] Coding sequences may be optimized for stability and high levels of expression. In some instances, codons are selected to reduce secondary structure formation of the RNA such as that formed due to intramolecular bonding.

[000133] The vector may comprise heterologous nucleic acid encoding the CRISPR/Cas9-based gene editing system and may further comprise an initiation codon, which may be upstream of the CRISPR/Cas9-based gene editing system coding sequence, and a stop codon, which may be downstream of the CRISPR/Cas9-based gene editing system coding sequence. The initiation and termination codon may be in frame with the CRISPR/Cas9-based gene editing system coding sequence. The vector may also comprise a promoter that is operably linked to the CRISPR/Cas9-based gene editing system coding sequence. The promoter that is operably linked to the CRISPR/Cas9-based gene editing system coding sequence may be a promoter from simian virus 40 (SV40), a mouse mammary tumor virus (MMTV) promoter, a human immunodeficiency virus (HIV) promoter such as the bovine immunodeficiency virus (BIV) long terminal repeat (LTR) promoter, a Moloney virus

promoter, an avian leukosis virus (ALV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter, Epstein Barr virus (EBV) promoter, a U6 promoter, such as the human U6 promoter, or a Rous sarcoma virus (RSV) promoter. The promoter may also be a promoter from a human gene such as human ubiquitin C (hUbC), human actin, human myosin, human hemoglobin, human muscle creatine, or human metalothionein. The promoter may also be a tissue specific promoter, such as a muscle or skin specific promoter, natural or synthetic. Examples of such promoters are described in US Patent Application Publication Nos. US20040175727 and US20040192593, the contents of which are incorporated herein in their entirety. Examples of muscle-specific promoters include a Spc5-12 promoter (described in US Patent Application Publication No. US 20040192593, which is incorporated by reference herein in its entirety; Hakim et al. Mol. Ther. Methods Clin. Dev. (2014) 1:14002; and Lai et al. Hum Mol Genet. (2014) 23(12): 3189–3199), a MHCK7 promoter (described in Salva et al., Mol. Ther. (2007) 15:320-329), a CK8 promoter (described in Park et al. PLoS ONE (2015) 10(4): e0124914), and a CK8e promoter (described in Muir et al., Mol. Ther. Methods Clin. Dev. (2014) 1:14025). In some embodiments, the expression of the gRNA and/or Cas9 protein is driven by tRNAs.

[000134] Each of the polynucleotide sequences encoding the gRNA molecule and/or Cas9 molecule may each be operably linked to a promoter. The promoters that are operably linked to the gRNA molecule and/or Cas9 molecule may be the same promoter. The promoters that are operably linked to the gRNA molecule and/or Cas9 molecule may be different promoters. The promoter may be a constitutive promoter, an inducible promoter, a repressible promoter, or a regulatable promoter. The promoter may be a tissue specific promoter. The tissue specific promoter may be a muscle specific promoter. Examples of muscle-specific promoters may include a MHCK7 promoter, a CK8 promoter, and a Spc512 promoter. The promoter may be a CK8 promoter, a Spc512 promoter, a MHCK7 promoter, for example.

[000135] The vector may also comprise a polyadenylation signal, which may be downstream of the CRISPR/Cas9-based gene editing system. The polyadenylation signal may be a SV40 polyadenylation signal, LTR polyadenylation signal, bovine growth hormone (bGH) polyadenylation signal, human growth hormone (hGH) polyadenylation signal, or human β-globin polyadenylation signal. The SV40 polyadenylation signal may be a polyadenylation signal from a pCEP4 vector (Invitrogen, San Diego, CA).

[000136] The vector may also comprise an enhancer upstream of the CRISPR/Cas9-based gene editing system, i.e., the Cas9 protein or Cas9 fusion protein coding sequence or sgRNAs, or the CRISPR/Cas9-based gene editing system. The enhancer may be

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

[000137] The vector may be expression vectors or systems to produce protein by routine techniques and readily available starting materials including Sambrook et al., Molecular Cloning and Laboratory Manual, Second Ed., Cold Spring Harbor (1989), which is incorporated fully by reference. In some embodiments the vector may comprise the nucleic acid sequence encoding the CRISPR/Cas9-based gene editing system, including the nucleic acid sequence encoding the Cas9 protein or Cas9 fusion protein and the nucleic acid sequence encoding the at least one gRNA.

7. Pharmaceutical Compositions

[000138] The presently disclosed subject matter provides for compositions comprising the above-described genetic constructs. The pharmaceutical compositions as detailed herein can be formulated according to the mode of administration to be used. In cases where pharmaceutical compositions are injectable pharmaceutical compositions, they are sterile, pyrogen free and particulate free. An isotonic formulation is preferably used. Generally, additives for isotonicity may include sodium chloride, dextrose, mannitol, sorbitol and lactose. In some cases, isotonic solutions such as phosphate buffered saline are preferred. Stabilizers include gelatin and albumin. In some embodiments, a vasoconstriction agent is added to the formulation.

[000139] The composition may further comprise a pharmaceutically acceptable excipient. The pharmaceutically acceptable excipient may be functional molecules as vehicles, adjuvants, carriers, or diluents. The pharmaceutically acceptable excipient may be a transfection facilitating agent, which may include surface active agents, such as immune-stimulating complexes (ISCOMS), Freunds incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs, vesicles such as squalene and

squalene, hyaluronic acid, lipids, liposomes, calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents.

[000140] The transfection facilitating agent is a polyanion, polycation, including poly-L-glutamate (LGS), or lipid. The transfection facilitating agent is poly-L-glutamate, and more preferably, the poly-L-glutamate is present in the composition for genome editing in skeletal muscle or cardiac muscle at a concentration less than 6 mg/ml. The transfection facilitating agent may also include surface active agents such as immune-stimulating complexes (ISCOMS), Freunds incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs and vesicles such as squalene and squalene, and hyaluronic acid may also be used administered in conjunction with the genetic construct. In some embodiments, the DNA vector encoding the composition may also include a transfection facilitating agent such as lipids, liposomes, including lecithin liposomes or other liposomes known in the art, as a DNA-liposome mixture (see for example International Patent Publication No. W09324640), calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents. Preferably, the transfection facilitating agent is a polyanion, polycation, including poly-L-glutamate (LGS), or lipid.

8. Methods of Delivery

[000141] Provided herein is a method for delivering the presently disclosed genetic construct (e.g., a vector) or a composition thereof to a cell. The delivery of the compositions may be the transfection or electroporation of the composition as a nucleic acid molecule that is expressed in the cell and delivered to the surface of the cell. The nucleic acid molecules may be electroporated using BioRad Gene Pulser Xcell or Amaxa Nucleofector IIb devices. Several different buffers may be used, including BioRad electroporation solution, Sigma phosphate-buffered saline product #D8537 (PBS), Invitrogen OptiMEM I (OM), or Amaxa Nucleofector solution V (N.V.). Transfections may include a transfection reagent, such as Lipofectamine 2000.

[000142] Upon delivery of the presently disclosed genetic construct or composition to the tissue, and thereupon the vector into the cells of the mammal, the transfected cells will express the gRNA molecule(s) and the Cas9 molecule. The genetic construct or composition may be administered to a mammal to alter gene expression or to re-engineer or alter the genome. For example, the genetic construct or composition may be administered to a mammal to correct the dystrophin gene in a mammal. The mammal may be human, non-human primate, cow, pig, sheep, goat, antelope, bison, water buffalo, bovids, deer,

hedgehogs, elephants, llama, alpaca, mice, rats, or chicken, and preferably human, cow, pig, or chicken.

[000143] The genetic construct (e.g., a vector) encoding the gRNA molecule(s) and the Cas9 molecule can be delivered to the mammal by DNA injection (also referred to as DNA vaccination) with and without *in vivo* electroporation, liposome mediated, nanoparticle facilitated, and/or recombinant vectors. The recombinant vector can be delivered by any viral mode. The viral mode can be recombinant lentivirus, recombinant adenovirus, and/or recombinant adeno-associated virus.

[000144] A presently disclosed genetic construct (e.g., a vector) or a composition comprising thereof can be introduced into a cell to genetically correct a dystrophin gene (e.g., human dystrophin gene). In certain embodiments, a presently disclosed genetic construct (e.g., a vector) or a composition comprising thereof is introduced into a myoblast cell from a DMD patient. In certain embodiments, the genetic construct (e.g., a vector) or a composition comprising thereof is introduced into a fibroblast cell from a DMD patient, and the genetically corrected fibroblast cell can be treated with MyoD to induce differentiation into myoblasts, which can be implanted into subjects, such as the damaged muscles of a subject to verify that the corrected dystrophin protein is functional and/or to treat the subject. The modified cells can also be stem cells, such as induced pluripotent stem cells, bone marrow-derived progenitors, skeletal muscle progenitors, human skeletal myoblasts from DMD patients, CD 133+ cells, mesoangioblasts, and MyoD- or Pax7- transduced cells, or other myogenic progenitor cells. For example, the CRISPR/Cas9-based gene editing system may cause neuronal or myogenic differentiation of an induced pluripotent stem cell.

9. Routes of Administration

[000145] The presently disclosed genetic constructs (e.g., vectors) or a composition comprising thereof may be administered to a subject by different routes including orally, parenterally, sublingually, transdermally, rectally, transmucosally, topically, via inhalation, via buccal administration, intrapleurally, intravenous, intraarterial, intraperitoneal, subcutaneous, intramuscular, intranasal intrathecal, and intraarticular or combinations thereof. In certain embodiments, the presently disclosed genetic construct (e.g., a vector) or a composition is administered to a subject (e.g., a subject suffering from DMD) intramuscularly, intravenously or a combination thereof. For veterinary use, the presently disclosed genetic constructs (e.g., vectors) or compositions may be administered as a suitably acceptable formulation in accordance with normal veterinary practice. The veterinarian may readily determine the dosing regimen and route of administration that is most appropriate for a particular animal.

The compositions may be administered by traditional syringes, needleless injection devices, "microprojectile bombardment gone guns", or other physical methods such as electroporation ("EP"), "hydrodynamic method", or ultrasound.

[000146] The presently disclosed genetic construct (e.g., a vector) or a composition may be delivered to the mammal by several technologies including DNA injection (also referred to as DNA vaccination) with and without in vivo electroporation, liposome mediated, nanoparticle facilitated, recombinant vectors such as recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The composition may be injected into the skeletal muscle or cardiac muscle. For example, the composition may be injected into the tibialis anterior muscle or tail.

[000147] In some embodiments, the presently disclosed genetic construct (e.g., a vector) or a composition thereof is administered by 1) tail vein injections (systemic) into adult mice; 2) intramuscular injections, for example, local injection into a muscle such as the TA or gastrocnemius in adult mice; 3) intraperitoneal injections into P2 mice; or 4) facial vein injection (systemic) into P2 mice.

10. Cell types

[000148] Any of these delivery methods and/or routes of administration can be utilized with a myriad of cell types. Cell types may include, but are not limited to, immortalized myoblast cells, such as wild-type and DMD patient derived lines, for example Δ48-50 DMD, DMD 6594 (del48-50), DMD 8036 (del48-50), C25C14 and DMD-7796 cell lines, primal DMD dermal fibroblasts, induced pluripotent stem cells, bone marrow-derived progenitors, skeletal muscle progenitors, human skeletal myoblasts from DMD patients, CD 133+ cells, mesoangioblasts, cardiomyocytes, hepatocytes, chondrocytes, mesenchymal progenitor cells, hematopoietic stem cells, smooth muscle cells, and MyoD- or Pax7-transduced cells, or other myogenic progenitor cells. Immortalization of human myogenic cells can be used for clonal derivation of genetically corrected myogenic cells. Cells can be modified *ex vivo* to isolate and expand clonal populations of immortalized DMD myoblasts that include a genetically corrected dystrophin gene and are free of other nuclease-introduced mutations in protein coding regions of the genome. Alternatively, transient *in vivo* delivery of CRISPR/Cas9-based systems by non-viral or non-integrating viral gene transfer, or by direct delivery of purified proteins and gRNAs containing cell-penetrating motifs may enable highly specific correction *in situ* with minimal or no risk of exogenous DNA integration.

11. Kits

[000149] Provided herein is a kit, which may be used to correct a mutated dystrophin gene. The kit comprises at least a gRNA for correcting a mutated dystrophin gene and instructions for using the CRISPR/Cas9-based gene editing system. Also provided herein is a kit, which may be used for genome editing of a dystrophin gene in skeletal muscle or cardiac muscle. The kit may comprise genetic constructs (e.g., vectors) or a composition comprising thereof for genome editing in skeletal muscle or cardiac muscle, as described above, and instructions for using said composition.

[000150] Instructions included in kits may be affixed to packaging material or may be included as a package insert. While the instructions are typically written or printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (e.g., CD ROM), and the like. As used herein, the term "instructions" may include the address of an internet site that provides the instructions.

[000151] The genetic constructs (e.g., vectors) or a composition comprising thereof for correcting a mutated dystrophin or genome editing of a dystrophin gene in skeletal muscle or cardiac muscle may include a modified AAV vector that includes a gRNA molecule(s) and a Cas9 molecule, as described above, that specifically binds and cleaves a region of the dystrophin gene. The CRISPR/Cas9-based gene editing system, as described above, may be included in the kit to specifically bind and target a particular region in the mutated dystrophin gene. The kit may further include donor DNA, a different gRNA, or a transgene, as described above.

12. Examples

[000152] It will be readily apparent to those skilled in the art that other suitable modifications and adaptations of the methods of the present disclosure described herein are readily applicable and appreciable, and may be made using suitable equivalents without departing from the scope of the present disclosure or the aspects and embodiments disclosed herein. Having now described the present disclosure in detail, the same will be more clearly understood by reference to the following examples, which are merely intended only to illustrate some aspects and embodiments of the disclosure, and should not be viewed as limiting to the scope of the disclosure. The disclosures of all journal references,

U.S. patents, and publications referred to herein are hereby incorporated by reference in their entireties.

[000153] The present disclosure details multiple embodiments and aspects, illustrated by the following non-limiting examples.

Example 1

Dual vector system

[000154] Conventional CRISPR/Cas9 systems for the treatment of DMD typically include more than one vector (**FIG. 6**, **FIG. 7**). For example, one vector may encode a Cas9 protein, and a second vector may encode two gRNAs. As another example, one vector may encode a Cas9 protein and a first gRNA, and a second vector may encode a Cas9 protein and a second gRNA.

[000155] A schematic of an experiment that uses multiple vectors to excise exons 45-55 of dystrophin in mice is shown in **FIG. 3** with results shown in **FIG. 4**, **FIG. 5**, and **FIG. 10**. Neonatal mice were treated with the dual vector system via systemic/temporal vein injection. At 8 weeks post-treatment, tissue was harvested. As shown in **FIG. 4**, PCR and sequencing confirmed the deletion of the mutational hotspot exon 45-55. Additional results are shown in **FIG. 10** with either AAV-CRISPR targeting a control locus (**FIG. 10**, top panel) or targeting exon 45-55 (**FIG. 10**, bottom panel),showing that widespread dystrophin expression was observed in cardiac muscle after deletion of exon 45-55, but not in sham vector-treated mice.

Example 2

Validation of therapeutic approach for dual vector system

[000156] Additional validation of the CRISPR-based approach to restore functional dystrophin gene with the dual vectors of Example 1 was performed using immortalized myoblasts isolated from a DMD patient. The immortalized myoblasts contained a deletion of exons 48-50, creating an out-of-frame mutation (**FIG. 9A**). Patient myoblasts were transfected with the same AAV plasmids used in the HEK293 *in vitro* experiment in Example 1.

[000157] Deletion PCR of genomic DNA and cDNA revealed that exon 45-55 was effectively deleted, which was confirmed by Sanger sequencing (**FIG. 9B**). Western blot of cell lysates showed that untreated myoblasts produced no dystrophin protein, while

transfected myoblasts expressed a smaller dystrophin protein compared to the positive control, consistent with hotspot deletion (**FIG. 9C**). These results additionally provided *in vitro* validation that the dual vector constructs can be used to edit a human mutation and restore dystrophin expression.

Example 3

Components for All-in-One Vectors

[000158] A one-vector CRISPR/Cas9 system was developed for the treatment of DMD (**FIG. 6**, **FIG. 7**). Advantages to a one vector system may include having all necessary editing components on a single vector, ability to increase effective dose, streamlining of other vector production (single therapeutic agent), use/incorporation of muscle-specific promoters (for example, CK8, Spc512, MHCK7), and ability to target combinations of exons and large deletions (for example, by changing guide sequences). A schematic diagram of the all-in-one vectors developed is shown in **FIG. 8**. Sequences included in some or all of the herein described all-in-one vectors are shown in **TABLE 1**. **FIG. 12**, **FIG. 13**, and **FIG 14** show results from testing these constructs in the mdx mouse. The all-in-one vectors are further detailed in Examples 4-7.

TABLE 1

Component	Sequence
AAV ITR	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCC GGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAG CGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGT TCCT (SEQ ID NO:1)
JCR143: guide sequence RNA targeting human dystrophin intron 44 region	ACATTTCCTCTCTATACAAATG (SEQ ID NO:2)
JCR120: guide sequence RNA targeting human dystrophin intron 55 region	ATATAGTAATGAAATTATTGGCAC (SEQ ID NO:3)
SaCas9 guide RNA scaffold	TCTCGCCAACAAGTTGACGAGATAAACACGGCATTTTGCCTTGT TTTAGTAGATTCTGTTTCCAGAGTACTAAAAC (SEQ ID NO:4)
U6 promoter	GGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGA AATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAAC ATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTAC GTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAA TTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATG AAGGAATCATGGGAAATAGGCCCTC (SEQ ID NO:5)
H1 promoter	GAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCC AGTGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGC AGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAAT

	ATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGA AATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACCA C (SEQ ID NO:6)
EFS promoter	TCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCG CCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGA ACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGT GATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAG AACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGC AACGGGTTTGCCGCCAGAACACAGGTGTCGTGACCGCGG (SEQ ID NO:7)
CK8 promoter	CTAGACTAGCATGCTGCCCATGTAAGGAGGCAAGGCCTGGGG ACACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTG CCCCCCCCCCCCCCAACACCTGCTGCCTCTAAAAATAACCCTGC ATGCCATGTTCCCGGCGAAGGGCCAGCTGTCCCCCGCCAGCT AGACTCAGCACTTAGTTTAGGAACCAGTGAGCAAGTCAGCCCT TGGGGCAGCCCATACAAGGCCATGGGGCTGGGCAAGCTGCAC GCCTGGGTCCGGGGTGGGCACGGTGCCCGGGCAACGAGCTG AAAGCTCATCTGCTCTCAGGGGCCCCTCCCTGGGGACAGCCC CTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATATAACCCAG GGGCACAGGGGCTGCCCTCATTCTACCACCACCTCCACAGCAC AGACAGACACTCAGGAGCCAGCCAG (SEQ ID NO:8)
Spc512 promoter	GAGCTCCACCGCGGTGGCGGCCGTCCGCCTTCGGCACCATCC TCACGACACCCAAATATGGCGACGGGTGAGGAATGGTGGGGA GTTATTTTTAGAGCGGTGAGGAAGGTGGGCAGGCAGCAGGTGT TGGCGCTCTAAAAATAACTCCCGGGAGTTATTTTTAGAGCGGAG GAATGGTGGACACCCAAATATGGCGACGGTTCCTCACCCGTCG CCATATTTGGGTGTCCGCCCTCGGCCGGGGCCGCATTCCTGG GGGCCGGGCGGTGCTCCCGCCCGCCTCGATAAAAGGCTCCGG GGCCGGCGGCGGCCCACGAGCTACCCGGAGGAGCGGGAGGC GCCAAGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTC GATAT (SEQ ID NO:9)
MHCK7 promoter	GTTTAAACAAGCTTGCATGTCTAAGCTAGACCCTTCAGATTAAA AATAACTGAGGTAAGGGCCTGGGTAGGGGAGGTGGTGTGAGA CGCTCCTGTCTCTCCTCTATCTGCCCATCGGCCCTTTGGGGAG GAGGAATGTGCCCAAGGACTAAAAAAAGGCCATGGAGCCAGAG GGGCGAGGGCAACAGACCTTTCATGGGCAAACCTTGGGGCCC TGCTGTCTAGCATGCCCCACTACGGGTCTAGGCTGCCCATGTA AGGAGGCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATT AACCCAGACATGTGGCTGCCCCCCCCCCCCCCAACACCTGCTGC CTCTAAAAATAACCCTGTCCCTGGTGGATCCCCTGCATGCGAA GATCTTCGAACAAGGCTGTGGGGGACTGAGGGCAGGCTGTAA CAGGCTTGGGGGCCAGGGCTTATACGTGCCTGGGACTCCCAA AGTATTACTGTTCCATGTTCCCGGCGAAGGGCCAGCTGTCCCC CGCCAGCTAGACTCAGCACTTAGTTTAGGAACCAGTGAGCAAG TCAGCCCTTGGGGCAGCCCATACAAGGCCATGGGGCTGGGCA AGCTGCACGCCTGGGTCCGGGGTGGGCACGGTGCCCGGGCA ACGAGCTGAAAGCTCATCTGCTCTCAGGGGCCCCTCCCTGGG GACAGCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATA TAACCCAGGGGCACAGGGGCTGCCCTCATTCTACCACCACCTC CACAGCACAGACAGACACTCAGGAGCCAGCCAGCGGCGCGCC C (SEQ ID NO:10)
SaCas9	AAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCG TGGGCTACGGCATCATCGACTACGAGACACGGGACGTGATCGA TGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAA

	CGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGC GGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGT TCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCAT CAACCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCT GAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAA GAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACAC CGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAG CAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGA ACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAG ATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTG AAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCG ACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGA GGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAA AGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCC GAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTG TACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGG ACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCAT CGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAG ATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCT ACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAA GGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATT ATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGA CCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAA TCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCT AATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGG CCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAA CCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAG GTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTG GACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCC AGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCT GCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCC AAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACC GGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCG GCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCA CGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATC CCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGG ACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAA CAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGG CAACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAG ATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAA GGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTG GAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCA TCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCT GATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGAC GTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGC GGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACA AGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTT CATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTG ATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGC CCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCAC CCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAG TACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTA ACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACA CCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAA TGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTG

	CTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGC TGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAA GTACTACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAA AAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCA ACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAA CAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGA TTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCG TGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTG AATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCA GCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCT GATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAAC AACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCA CCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCC CAGGATCATTAAGACAATCGCCTCCAAGACCCAGAGCATTAAG AAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATC TAAGAAGCACCCTCAGATCATCAAAAAGGGC (SEQ ID NO:11)
Mini polyadenylation signal	TAGCAATAAAGGATCGTTTATTTTCATTGGAAGCGTGTGTTGGT TTTTTGATCAGGCGCG (SEQ ID NO:12)
bGH polyadenylation signal	CTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCA GCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAGAA TAGCAGGCATGCTGGGGA (SEQ ID NO:13)
SV40 intron	TCTAGAGGATCCGGTACTCGAGGAACTGAAAAACCAGAAAGTT AACTGGTAAGTTTAGTCTTTTTGTCTTTTATTTCAGGTCCCGGAT CCGGTGGTGGTGCAAATCAAAGAACTGCTCCTCAGTGGATGTT GCCTTTACTTCTAGGCCTGTACGGAAGTGTTAC (SEQ ID NO: 24)

Example 4

All-in-One Vector 1 (Versions 1 and 2)

[000159] Two versions of vector 1 were generated. Vector 1 contained exon 45-55 targeted gRNAs with all promoters (U6, H1, and SaCas9-driving) in forward direction and mini polyadenylation signal for SaCas9.

[000160] Version 1 of vector 1 contained an EFS constitutive promoter. The sequence for version 1 of vector 1 is in SEQ ID NO:14.

[000161] Version 2 of vector 1 contained a CK8 constitutive promoter. The sequence for version 2 of vector 1 is in SEQ ID NO:15.

Example 5

All-in-One Vector 2 (Versions 1-4)

[000162] Four versions of vector 2 were generated. Vector 2 contained exon 45-55 targeted gRNAs with U6 promoter in reverse direction facing away from SaCas9-driving promoter and mini polyadenylation signal for SaCas9.

[000163] Version 1 of vector 2 contained an EFS constitutive promoter. The sequence for version 1 of vector 2 is in SEQ ID NO:16.

[000164] Version 2 of vector 2 contained a CK8 constitutive promoter. The sequence for version 2 of vector 2 is as in SEQ ID NO:17.

[000165] Version 3 of vector 2 contained a Spc512 promoter. The sequence for version 3 of vector 2 is as in SEQ ID NO:18.

[000166] Version 4 of vector 2 contained a MHCK7 promoter. The sequence for version 4 of vector 2 is as in SEQ ID NO:19.

Example 6

All-in-One Vector 3 (Versions 1-4)

[000167] Four versions of vector 3 were generated. Vector 3 contained exon 45-55 targeted gRNAs with U6 promoter in reverse direction facing away from SaCas9-driving promoter and mini polyadenylation signal for SaCas9.

[000168] Version 1 of vector 3 contained an EFS constitutive promoter. The sequence for version 1 of vector 3 is as in SEQ ID NO:20.

[000169] Version 2 of vector 3 contained a CK8 promoter. The sequence for version 2 of vector 3 is as in SEQ ID NO:21.

[000170] Version 3 of vector 3 contained a Spc512 promoter. The sequence for version 3 of vector 3 is as in SEQ ID NO:22.

[000171] Version 4 of vector 3 contained a MHCK7 promoter. The sequence for version 4 of vector 3 is as in SEQ ID NO:23.

Example 7

All-in-One Vector 5 (Versions 1-4)

[000172] After screening a panel of all-in-one vector designs to determine the effect of guide placement, regulatory elements, and Pol-III promoters, a new set of all-in-one vectors was created with constitutive and muscle-specific promoters (**FIG. 11**). Versions of vector 5 of the all-in-one vector included an SV40 intron (see SEQ ID NO: 24) and placement of different elements.

[000173] Version 1 of vector 5 included a constitutive promoter. The sequence for version 1 of vector 5 is as in SEQ ID NO: 41.

[000174] Version 2 of vector 5 included a CK8 promoter. The sequence for version 2 of vector 5 is as in SEQ ID NO: 42.

[000175] Version 3 of vector 5 included a Spc-512 promoter. The sequence for version 3 of vector 5 is as in SEQ ID NO: 29.

[000176] Version 4 of vector 5 included a MHCK7 promoter. The sequence for version 4 of vector 5 is as in SEQ ID NO: 30.

***

[000177] The foregoing description of the specific aspects will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art, readily modify and/or adapt for various applications such specific aspects, without undue experimentation, without departing from the general concept of the present disclosure. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed aspects, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance.

[000178] The breadth and scope of the present disclosure should not be limited by any of the above-described exemplary aspects but should be defined only in accordance with the following claims and their equivalents.

[000179] All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

[000180] For reasons of completeness, various aspects of the disclosure are set out in the following numbered clauses:

[000181] Clause 1. A CRISPR-Cas system comprising one or more vectors encoding a composition, the composition comprising: (a) a first guide RNA (gRNA) molecule targeting intron 44 of dystrophin; (b) a second gRNA molecule targeting intron 55 of dystrophin; and (c) a Cas9 protein; and (d) one or more Cas9 gRNA scaffolds.

[000182] Clause 2. The system of clause 1, wherein the system comprises a single vector.

[000183] Clause 3. The system of clause 1, wherein the system comprises two or more vectors, wherein the two or more vectors comprises a first vector and a second vector.

[000184] Clause 4. The system of clause 3, wherein (a) the first vector encodes the first gRNA molecule and the second gRNA molecule; and (b) the second vector encodes the Cas9 protein.

[000185] Clause 5. The system of clause 3, wherein (a) the first vector encodes the first gRNA molecule; and (b) the second vector encodes the second gRNA molecule.

[000186] Clause 6. The system of clause 5, wherein the first vector further encodes the Cas9 protein.

[000187] Clause 7. The system of clause 5 or 6, wherein the second vector further encodes the Cas9 protein.

[000188] Clause 8. The system of any one of clauses 1-7, wherein the expression of the Cas9 protein is driven by a constitutive promoter or a muscle-specific promoter.

[000189] Clause 9. The system of clause 8, where the muscle-specific promoter comprises a MHCK7 promoter, a CK8 promoter, or a Spc512 promoter.

[000190] Clause 10. The system of clause 2, wherein the single vector encodes the first gRNA molecule, the second gRNA molecule, and the Cas9 protein.

[000191] Clause 11. The system of any one of clauses 1-10, wherein the vector comprises at least one bidirectional promoter.

[000192] Clause 12. The system of clause 11, wherein the bidirectional promoter comprises: a first promoter driving expression of the first gRNA molecule and/or the second gRNA molecule; and a second promoter driving expression of the Cas9 protein.

[000193] Clause 13. The system of any one of clauses 1-12, wherein the first gRNA targets the polynucleotide of SEQ ID NO:2 or a 5' truncation thereof.

[000194] Clause 14. The system of any one of clauses 1-13, wherein the second gRNA targets the polynucleotide of SEQ ID NO:3 or a 5' truncation thereof.

[000195] Clause 15. The system of any one of clauses 1-14, wherein the Cas9 protein is SpCas9, SaCas9, or St1Cas9 protein.

[000196] Clause 16. The system of any one of clauses 1-15, wherein the Cas9 gRNA scaffold is a SaCas9 gRNA scaffold.

[000197] Clause 17. The system of clause 16, wherein the SaCas9 gRNA scaffold comprises or is encoded by the polynucleotide of SEQ ID NO:4.

[000198] Clause 18. The system of any one of clauses 1-17, wherein the Cas9 protein is a SaCas9 protein encoded by the polynucleotide of SEQ ID NO:11.

[000199] Clause 19. The system of any one of clauses 1-18, wherein the vector comprises at least one polynucleotide selected from SEQ ID NOs: 1-13 and 24.

[000200] Clause 20. The system of any one of clauses 1-19, wherein the vector comprises the polynucleotide sequence of SEQ ID NO: 24.

[000201] Clause 21. The system of any one of clauses 1-20, wherein the vector comprises a polynucleotide sequence that is selected from SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 29, and SEQ ID NO: 30.

[000202] Clause 22. The system of any one of clauses 1-21, wherein the vector is a viral vector.

[000203] Clause 23. The system of any one of clauses 1-22, wherein the vector is an Adeno-associated virus (AAV) vector.

[000204] Clause 24. The system of clause 23, wherein the AAV vector is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV-10, AAV-11, AAV-12, AAV-13 or AAVrh.74.

[000205] Clause 25. The system of any one of clauses 1-24, wherein the vector comprises a ubiquitous promoter or a tissue-specific promoter operably linked to the polynucleotide sequence encoding the first gRNA molecule, the second gRNA molecule, and/or the Cas9 protein.

[000206] Clause 26. The system of clause 25, wherein the tissue-specific promoter is a muscle specific promoter.

[000207] Clause 27. A cell comprising the system of any one of clauses 1-26.

[000208] Clause 28. A kit comprising the system of any one of clauses 1-26.

[000209] Clause 29. A method of correcting a mutant dystrophin gene in a cell, the method comprising administering to a cell the system of any one of clauses 1-26.

[000210] Clause 30. A method of genome editing a mutant dystrophin gene in a subject, the method comprising administering to the subject the system of any one of clauses 1-26 or the cell of clause 27.

[000211] Clause 31. A method of treating a subject having a mutant dystrophin gene, the method comprising administering to the subject the system of any one of clauses 1-26 or the cell of clause 27.

[000212] Clause 32. The method of clause 30 or 31, wherein the system or the cell is administered to the subject intramuscularly, intravenously, or a combination thereof.

SEQUENCES

SEQ ID NO: 1, AAV ITR
cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttgg
tcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggg
gttcct (SEQ ID NO:1)

SEQ ID NO: 2, JCR143: DNA target sequence of gRNA targeting human dystrophin intron 44 region
acatttcctctctatacaaatg (SEQ ID NO:2)

SEQ ID NO: 3, JCR120: DNA target sequence of gRNA targeting human dystrophin intron 55 region
atatagtaatgaaattattggcac (SEQ ID NO:3)

SEQ ID NO: 4, SaCas9 guide RNA scaffold, scaffold of gRNAs
tctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattctgtttc
cagagtactaaaac (SEQ ID NO:4)

SEQ ID NO: 5, U6 promoter
ggtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacgg
taagcatatgatagtccattttaaaacataattttaaaactgcaaactacccaagaaattat
tactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattccaat
tatctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggccct
c (SEQ ID NO:5)

SEQ ID NO: 6, H1 promoter
gaacgctgacgtcatcaacccgctccaaggaatcgcgggcccagtgtcactaggcgggaaca
cccagcgcgcgtgcgccctggcaggaagatggctgtgagggacaggggagtggcgccctgca
atatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatgtctttggatttg
ggaatcttataagttctgtatgagaccac (SEQ ID NO:6)

SEQ ID NO: 7, EFS promoter
tcgagtggctccggtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaagtt
ggggggaggggtcggcaattgaaccggtgcctagagaaggtggcgcggggtaaactgggaaa
gtgatgtcgtgtactggctccgccttttcccgagggtgggggagaaccgtatataagtgca
gtagtcgccgtgaacgttcttttcgcaacgggtttgccgccagaacacaggtgtcgtgacc
gcgg (SEQ ID NO:7)

SEQ ID NO: 8, CK8 promoter
ctagactagcatgctgcccatgtaaggaggcaaggcctggggacacccgagatgcctggtta
taattaacccagacatgtggctgccccccccccccaacacctgctgcctctaaaaataacc
ctgcatgccatgttcccggcgaagggccagctgtcccccgccagctagactcagcacttagt
ttaggaaccagtgagcaagtcagcccttggggcagcccatacaaggccatggggctgggcaa
gctgcacgcctgggtccggggtgggcacggtgcccgggcaacgagctgaaagctcatctgct
ctcaggggcccctccctggggacagcccctcctggctagtcacaccctgtaggctcctctat
ataacccaggggcacaggggctgccctcattctaccaccacctccacagcacagacagacac
tcaggagccagccag (SEQ ID NO:8)

SEQ ID NO: 9, Spc512 promoter
gagctccaccgcggtggcggccgtccgccttcggcaccatcctcacgacacccaaatatggc
gacgggtgaggaatggtggggagttatttttagagcggtgaggaaggtgggcaggcagcagg
tgttggcgctctaaaaataactcccgggagttatttttagagcggaggaatggtggacaccc
aaatatggcgacggttcctcacccgtcgccatatttgggtgtccgccctcggccggggccgc
attcctgggggccgggcggtgctcccgcccgcctcgataaaaggctccgggggccggcggcgg
cccacgagctacccggaggagcgggaggcgccaagctctagaactagtggatcccccgggct
gcaggaattcgatat (SEQ ID NO:9)

SEQ ID NO: 10, MHCK7 promoter
gtttaaacaagcttgcatgtctaagctagacccttcagattaaaaataactgaggtaagggc
ctgggtaggggaggtggtgtgagacgctcctgtctctcctctatctgcccatcggccctttg
gggaggaggaatgtgcccaaggactaaaaaaaggccatggagccagaggggcgagggcaaca

gacctttcatgggcaaaccttggggccctgctgtctagcatgccccactacgggtctaggct
gcccatgtaaggaggcaaggcctggggacacccgagatgcctggttataattaacccagaca
tgtggctgcccccccccccccaacacctgctgcctctaaaaataaccctgtccctggtggat
cccctgcatgcgaagatcttcgaacaaggctgtgggggactgagggcaggctgtaacaggct
tgggggccagggcttatacgtgcctgggactcccaaagtattactgttccatgttcccggcg
aagggccagctgtcccccgccagctagactcagcacttagtttaggaaccagtgagcaagtc
agcccttggggcagcccatacaaggccatggggctgggcaagctgcacgcctgggtccgggg
tgggcacggtgcccgggcaacgagctgaaagctcatctgctctcaggggcccctccctgggg
acagcccctcctggctagtcacaccctgtaggctcctctatataacccaggggcacaggggc
tgccctcattctaccaccacctccacagcacagacagacactcaggagccagccagcggcgc
gccc (SEQ ID NO:10)

SEQ ID NO: 11, polynucleotide encoding SaCas9

aagcggaactacatcctgggcctggacatcggcatcaccagcgtgggctacggcatcatcga
ctacgagacacgggacgtgatcgatgccggcgtgcggctgttcaaagaggccaacgtggaaa
acaacgagggcaggcggagcaagagaggcgccagaaggctgaagcggcggaggcggcataga
atccagagagtgaagaagctgctgttcgactacaacctgctgaccgaccacagcgagctgag
cggcatcaacccctacgaggccagagtgaagggcctgagccagaagctgagcgaggaagagt
tctctgccgccctgctgcacctggccaagagaagaggcgtgcacaacgtgaacgaggtggaa
gaggacaccggcaacgagctgtccaccaaagagcagatcagccggaacagcaaggccctgga
agagaaatacgtggccgaactgcagctggaacggctgaagaaagacggcgaagtgcgggca
gcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagctgctgaaggtgcag
aaggcctaccaccagctggaccagagcttcatcgacacctacatcgacctgctggaaacccg
gcggacctactatgagggacctggcgagggcagccccttcggctggaaggacatcaaagaat
ggtacgagatgctgatgggccactgcacctacttccccgaggaactgcggagcgtgaagtac
gcctacaacgccgacctgtacaacgccctgaacgacctgaacaatctcgtgatcaccaggga
cgagaacgagaagctggaatattacgagaagttccagatcatcgagaacgtgttcaagcaga
agaagaagcccaccctgaagcagatcgccaaagaaatcctcgtgaacgaagaggatattaag
ggctacagagtgaccagcaccggcaagcccgagttcaccaacctgaaggtgtaccacgacat
caaggacattaccgcccggaaagagattattgagaacgccgagctgctggatcagattgcca
agatcctgaccatctaccagagcagcgaggacatccaggaagaactgaccaatctgaactcc
gagctgacccaggaagagatcgagcagatctctaatctgaagggctataccggcacccacaa
cctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccaga
tcgctatcttcaaccggctgaagctggtgcccaagaaggtggacctgtcccagcagaaagag
atccccaccaccctggtggacgacttcatcctgagccccgtcgtgaagagaagcttcatcca
gagcatcaaagtgatcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcg
agctggcccgcgagaagaactccaaggacgcccagaaaatgatcaacgagatgcagaagcgg
aaccggcagaccaacgagcggatcgaggaaatcatccggaccaccggcaaagagaacgccaa
gtacctgatcgagaagatcaagctgcacgacatgcaggaaggcaagtgcctgtacagcctgg
aagccatccctctggaagatctgctgaacaacccccttcaactatgaggtggaccacatcatc
cccagaagcgtgtccttcgacaacagcttcaacaacaaggtgctcgtgaagcaggaagaaaa
cagcaagaagggcaaccggaccccattccagtacctgagcagcagcgacagcaagatcagct
acgaaaccttcaagaagcacatcctgaatctggccaagggcaagggcagaatcagcaagacc
aagaaagagtatctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttcat
caaccggaacctggtggataccagatacgccaccagaggcctgatgaacctgctgcggagct
acttcagagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttcaccagcttt
ctgcggcggaagtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgagga
cgccctgatcattgccaacgccgatttcatcttcaaagagtggaagaaactggacaaggcca
aaaaagtgatggaaaaccagatgttcgaggaaaagcaggccgagagcatgcccgagatcgaa
accgagcaggagtacaaagagatcttcatcacccccaccagatcaagcacattaaggactt
caaggactacaagtacagccaccgggtggacaagaagcctaatagagagctgattaacgaca

ccctgtactccacccggaaggacgacaagggcaacaccctgatcgtgaacaatctgaacggc
ctgtacgacaaggacaatgacaagctgaaaaagctgatcaacaagagccccgaaaagctgct
gatgtaccaccacgacccccagacctaccagaaactgaagctgattatggaacagtacggcg
acgagaagaatcccctgtacaagtactacgaggaaaccgggaactacctgaccaagtactcc
aaaaaggacaacggccccgtgatcaagaagattaagtattacggcaacaaactgaacgccca
tctggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagctgtccctgaagc
cctacagattcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctg
gatgtgatcaaaaaagaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaa
gctgaagaagatcagcaaccaggccgagtttatcgcctccttctacaacaacgatctgatca
agatcaacggcgagctgtatagagtgatcggcgtgaacaacgacctgctgaaccggatcgaa
gtgaacatgatcgacatcacctaccgcgagtacctggaaaacatgaacgacaagaggccccc
caggatcattaagacaatcgcctccaagacccagagcattaagaagtacagcacagacattc
tgggcaacctgtatgaagtgaaatctaagaagcaccctcagatcatcaaaaagggc (SEQ
ID NO:11)

SEQ ID NO: 12, Mini polyadenylation signal
tagcaataaaggatcgtttattttcattggaagcgtgtgttggtttttgatcaggcgcg
(SEQ ID NO:12)

SEQ ID NO: 13, bGH polyadenylation signal
ctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccc
tcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatga
ggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcagg
acagcaaggggaggattgggaagagaatagcaggcatgctgggga (SEQ ID NO:13)

SEQ ID NO:14, Version 1 of vector 1
cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGAT
AATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAAT
TTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGA
AAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGcatttgtatagagagg
aaatgtgttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtca
acttgttggcgagatttttCTCGAGTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGC
CCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGG
GTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTGTCGTGAC
CGCGGCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGGAACT
ACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGAC
GTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAA
GAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCG
ACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAAGGGC
CTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGT
GCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGA
ACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAA
GTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGT
GCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGC
GGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAG
ATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGA
CCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAAT
ATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATC
GCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGA
GTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACG
CCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAA

CTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATAC
CGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACA
ACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAG
ATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCAT
CAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCCCGCG
AGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAG
CGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCT
GCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACA
ACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAAC
AAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAG
CAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCA
GAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAA
GACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAG
CTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGC
GGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATC
ATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAA
CCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGA
TCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTG
GACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAA
CACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCA
ACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGAC
CAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTGAACG
CCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCC
TACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGAT
CAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCA
GCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTAT
AGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCTACCG
CGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCACCCT
CAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGgg
atcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttaccCatacg
atgttccagattacgcttaaGaattctagcaataaaggatcgtttatttttcattggaagcgtgtgttg
gttttttgatcaggcgcgGGTACCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAG
TGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAG
TGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTG
GATTTGGGAATCTTATAAGTTCTGTATGAGACCACATATAGTAATGAAATTATTGGCACGTTTTAGTA
CTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGAT
TTTTGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactga
ggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgc
gcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgc
atacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacg
cgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttct
cgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtg
ctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctga
tagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactgg
aacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctatt
ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatt
ttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaac
acccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtct
ccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtg
atacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcg
gggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatga
gacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgt
gtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaa
agtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta

agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgt
ggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaa
tgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattat
gcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccg
aaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgga
gctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgc
gcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcg
gataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcg
tagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggt
gcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaa
acttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatccctt
aacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcct
ttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgcc
ggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactg
tccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgct
ctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaag
acgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgg
agcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaa
gggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttcc
agggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttt
tgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg
gccttttgctggccttttgctcacatgt

SEQ ID NO:15, Version 2 of vector 1

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGAT
AATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAAT
TTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGA
AAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGcatttgtatagagagg
aaatgtgttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtca
acttgttggcgagatttttCTCGAGCTAGACTAGCATGCTGCCCATGTAAGGAGGCAAGGCCTGGGGA
CACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTGCCCCCCCCCCCCCAACACCTGCTGCC
TCTAAAAATAACCCTGCATGCCATGTTCCCGGCGAAGGGCCAGCTGTCCCCCGCCAGCTAGACTCAGC
ACTTAGTTTAGGAACCAGTGAGCAAGTCAGCCCTTGGGGCAGCCCATACAAGGCCATGGGGCTGGGCA
AGCTGCACGCCTGGGTCCGGGGTGGGCACGGTGCCCGGGCAACGAGCTGAAAGCTCATCTGCTCTCAG
GGGCCCCTCCCTGGGGACAGCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATATAACCCAGGG
GCACAGGGGCTGCCCTCATTCTACCACCACCTCCACAGCACAGACAGACACTCAGGAGCCAGCCAGCa
ccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGG
AACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACG
GGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGA
GCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTG
TTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAA
GGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAG
GCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGC
CGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGG
CGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGA
AGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACC
CGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTA
CGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTG
GAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCA
GATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGC
CCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAG

AACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGA
AGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCT
ATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAAC
GACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAA
AGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGA
GCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCC
CGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAA
CGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCA
AGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTG
AACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAA
CAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGA
GCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAG
GGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCA
GAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGC
GGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTT
CTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCT
GATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGG
AAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAA
GAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCG
GGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGG
GCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTG
ATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCT
GATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACC
TGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTG
AACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAA
GCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATG
TGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAG
ATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCT
GTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCT
ACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAG
ACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCA
CCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAA
AGggatcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttacccCa
tacgatgttccagattacgcttaaGaattctagcaataaaggatcgtttatttttcattggaagcgtgt
gttggtttttgatcaggcgcgGGTACCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGC
CCAGTGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGG
GGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTC
TTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACCACATATAGTAATGAAATTATTGGCACGTTTT
AGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCG
AGATTTTTGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctca
ctgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcga
gcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcaca
ccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggt
tacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcct
ttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattt
agtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgcc
ctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaa
ctggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcc
tattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttac
aattttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgc
caacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgacc
gtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcct
cgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcactt
ttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctc
atgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacattt
ccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctgg

tgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagc
ggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgct
atgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctc
agaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaa
ttatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaac
cggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacg
ttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatgga
ggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaat
ctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgt
atcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagat
aggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatt
taaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatc
ccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgaga
tccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtt
tgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaat
actgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacct
cgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggact
caagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcc
cgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagc
ttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcga
tttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggtt
cctggccttttgctggccttttgctcacatgt

SEQ ID NO: 16, Version 1 of vector 2

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGC
CCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGG
GTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTGTCGTGAC
CGCGGCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGGAACT
ACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGAC
GTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAA
GAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCG
ACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAAGGGC
CTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGT
GCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGA
ACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAA
GTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGT
GCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGC
GGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAG
ATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGA
CCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAAT
ATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATC
GCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGA
GTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACG
CCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAA
CTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATAC

CGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACA
ACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAG
ATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCAT
CAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCCCGCG
AGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAG
CGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCT
GCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACA
ACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAAC
AAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAG
CAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCA
GAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAA
GACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAG
CTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGC
GGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATC
ATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAA
CCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGA
TCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTG
GACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAA
CACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCA
ACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGAC
CAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTGAACG
CCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCC
TACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGAT
CAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCA
GCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTAT
AGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCTACCG
CGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCACCCT
CAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGgg
atcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttaccCatacg
atgttccagattacgcttaaGaattctagcaataaaggatcgtttattttcattggaagcgtgtgttg
gttttttgatcaggcgcgGGTACCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAG
TGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAG
TGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTG
GATTTGGGAATCTTATAAGTTCTGTATGAGACCACATATAGTAATGAAATTATTGGCACGTTTTAGTA
CTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGAT
TTTTGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactga
ggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgc
gcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgc
atacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacg
cgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttct
cgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtg
ctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctga
tagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactgg
aacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctatt
ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatt
ttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaac
acccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtct
ccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtg
atacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcg
gggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatga
gacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgt
gtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaa
agtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta
agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgt

ggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaa
tgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattat
gcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccg
aaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgga
gctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgc
gcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcg
gataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcg
tagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggt
gcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaa
acttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatccctt
aacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcct
ttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgcc
ggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactg
tccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgct
ctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaag
acgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgg
agcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaa
gggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttcc
agggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttt
tgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg
gccttttgctggccttttgctcacatgt

SEQ ID NO: 17, Version 2 of vector 2

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGCTAGACTAGCATGCTGCCCATGTAAGGAGGCAAGGCCTGGGGA
CACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTGCCCCCCCCCCCCCAACACCTGCTGCC
TCTAAAAATAACCCTGCATGCCATGTTCCCGGCGAAGGGCCAGCTGTCCCCCGCCAGCTAGACTCAGC
ACTTAGTTTAGGAACCAGTGAGCAAGTCAGCCCTTGGGGCAGCCCATACAAGGCCATGGGGCTGGGCA
AGCTGCACGCCTGGGTCCGGGGTGGGCACGGTGCCCGGGCAACGAGCTGAAAGCTCATCTGCTCTCAG
GGGCCCCTCCCTGGGGACAGCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATATAACCCAGGG
GCACAGGGGCTGCCCTCATTCTACCACCACCTCCACAGCACAGACAGACACTCAGGAGCCAGCCAGCa
ccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGG
AACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACG
GGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGA
GCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTG
TTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAA
GGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAG
GCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGC
CGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGG
CGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGA
AGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACC
CGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTA
CGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTG
GAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCA
GATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGC
CCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAG
AACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGA

AGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCT
ATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAAC
GACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAA
AGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGA
GCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCC
CGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAA
CGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCA
AGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTG
AACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAA
CAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGA
GCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAG
GGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCA
GAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGC
GGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTT
CTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCT
GATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGG
AAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAA
GAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCG
GGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGG
GCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTG
ATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCT
GATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACC
TGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTG
AACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAA
GCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATG
TGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAG
ATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCT
GTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCT
ACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAG
ACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCA
CCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAA
AGggatcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttacccCa
tacgatgttccagattacgcttaaGaattctagcaataaaggatcgtttattttcattggaagcgtgt
gttggtttttgatcaggcgcgGGTACCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGC
CCAGTGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGG
GGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTC
TTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACCACATATAGTAATGAAATTATTGGCACGTTTT
AGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCG
AGATTTTTGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctca
ctgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcga
gcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcaca
ccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggt
tacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcct
ttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattt
agtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgcc
ctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaa
ctggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcc
tattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttac
aattttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgc
caacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgacc
gtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcct
cgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcactt
ttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctc
atgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacattt
ccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctgg
tgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagc

ggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgct
atgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctc
agaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaa
ttatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaac
cggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacg
ttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatgga
ggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaat
ctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgt
atcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagat
aggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatt
taaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatc
ccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgaga
tccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtt
tgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaat
actgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacct
cgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggact
caagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcc
cgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagc
ttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcga
ttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggtt
cctggccttttgctggccttttgctcacatgt

SEQ ID NO: 18, Version 3 of vector 2

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGGAGCTCCACCGCGGTGGCGGCCGTCCGCCttCGGCACCATCCT
CACGACACCCAAATATGGCGACGGGTGAGGAATGGTGGGGAGTTATTTTTAGAGCGGTGAGGAAGGTG
GGCAGGCAGCAGGTGTTGGCGCTCTAAAAATAACTCCCGGGAGTTATTTTTAGAGCGGAGGAATGGTG
GACACCCAAATATGGCGACGGTTCCTCACCCGTCGCCATATTTGGGTGTCCGCCCTCGGCCGGGGCCG
CATTCCTGGGGGCCGGGCGGTGCTCCCGCCCGCCTCGATAAAAGGCTCCGGGGCCGGCGGCGGCCCAC
GAGCTACCCGGAGGAGCGGGAGGCGCCAAGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCG
ATATaccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCA
AGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAG
ACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAG
GCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGC
TGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGA
GTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAG
AAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGA
TCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAA
GACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCT
GCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGG
AAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAA
TGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTA
CAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGA
AGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTG
AAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGG
CAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTA
TTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATC

CAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAA
GGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACA
CCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAG
CAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCAT
CCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGC
TGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAG
ACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAA
GATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATC
TGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGC
TTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTA
CCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGG
GCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCC
GTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCT
GCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCA
GCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGAC
GCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGT
GATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGT
ACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGC
CACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGA
CAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAA
AGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTG
AAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAA
CTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACA
AACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCC
CTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCT
GGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGA
AGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGC
GAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACAT
CACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCT
CCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAG
AAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAA
GAAAAAGggatcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgctt
accCatacgatgttccagattacgcttaaGaattctagcaataaaggatcgtttatttttcattggaag
cgtgtgttggtttttgatcaggcgcgGGTACCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCG
CGGGCCCAGTGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGG
ACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAA
ATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACCACATATAGTAATGAAATTATTGGCAC
GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGT
TGGCGAGATTTTTGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctc
gctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcg
agcgagcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatt
tcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtg
gtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttccc
ttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttcc
gatttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggcca
tcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgtt
ccaaactggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgattt
cggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacg
tttacaattttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgaca
cccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctg
tgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaag
ggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtgg
cacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatc
cgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaa
catttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaac
gctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctca

acagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagtt
ctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacacta
ttctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaa
gagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatc
ggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttg
ggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaa
caacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactgg
atggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctga
taaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccct
cccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgct
gagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagat
tgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgacca
aaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttct
tgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggt
ttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac
caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctaca
tacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggtt
ggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagc
ccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacg
cttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgag
ggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagc
gtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttta
cggttcctggccttttgctggccttttgctcacatgt

SEQ ID NO: 19, Version 4 of vector 2

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCctcgagGtttaaacaagcttgcatgtctaagctagacccttcagattaa
aaataactgaggtaagggcctgggtaggggaggtggtgtgagacgctcctgtctctcctctatctgcc
catcggccctttggggaggaggaatgtgcccaaggactaaaaaaaggccatggagccagaggggcgag
ggcaacagacctttcatgggcaaaccttggggccctgctgtctagcatgccccactacgggtctaggc
tgcccatgtaaggaggcaaggcctggggacacccgagatgcctggttataattaacccagacatgtgg
ctgccccccccccccaacacctgctgcctctaaaaataaccctgtccctggtggatcccctgcatgc
gaagatcttcgaacaaggctgtgggggactgagggcaggctgtaacaggcttgggggccagggcttat
acgtgcctgggactcccaaagtattactgttccatgttcccggcgaagggccagctgtcccccgccag
ctagactcagcacttagtttaggaaccagtgagcaagtcagcccttggggcagcccatacaaggccat
ggggctgggcaagctgcacgcctgggtccggggtgggcacggtgcccgggcaacgagctgaaagctca
tctgctctcaggggcccctccctggggacagcccctcctggctagtcacaccctgtaggctcctctat
ataacccaggggcacaggggctgccctcattctaccaccacctccacagcacagacagacactcagga
gccagccagcggcgcgcccaccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACG
GAGTCCCAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGC
ATCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGA
AAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCC
AGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAAC
CCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCT
GCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGT
CCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTG
GAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAA
AGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCT
ACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGG

AAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAG
CGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCA
GGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAG
AAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAG
AGTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCG
CCCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAG
AGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCA
GATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGG
ACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAG
GTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGT
GAAGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACG
ACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAG
AAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAA
GTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCA
TCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTG
TCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCG
GACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCC
TGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGAC
ATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAG
AGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCA
ATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAG
CACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGA
CAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCG
AAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAG
GACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTC
CACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACA
ATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAG
ACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTA
CGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAG
GTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGT
GACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGG
AAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTG
ATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGT
GAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTAT
GAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGC
CGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacgatg
ttccagattacgcttaccCatacgatgttccagattacgcttaaGaattctagcaataaaggatcgtt
tattttcattggaagcgtgtgttggtttttgatcaggcgcgGGTACCGAACGCTGACGTCATCAACC
CGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGA
AGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAAT
CACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGACCACATATAGTAA
TGAAATTATTGGCACGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTT
ATCTCGTCAACTTGTTGGCGAGATTTTTGGTACCaggaacccctagtgatggagttggccactccctc
tctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccggg
cggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatgcggtattttctccttacg
catctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcatta
agcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc
tttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggc
tccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggt
tcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaa
tagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttataag
ggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatttt
aacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctgatgccgcatagtt
aagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccg

cttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaa
cgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttc
ttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatac
attcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaag
agtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttt
tgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttaca
tcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatg
agcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcgg
tcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacgg
atggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaactta
cttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaac
tcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgc
ctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaa
caattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctgg
ctggtttattgctgataaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggc
cagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactc
atatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttg
ataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaag
atcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccacc
gctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttca
gcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactct
gtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtc
gtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggg
gttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcta
tgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaac
aggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgcc
acctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagc
aacgcggccttttacggttcctggccttttgctggccttttgctcacatgt

SEQ ID NO: 20, Version 1 of vector 3

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGC
CCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGG
GTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTGTCGTGAC
CGCGGCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGGAACT
ACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGAC
GTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAA
GAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCG
ACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAAGGGC
CTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGT
GCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGA
ACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAA
GTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGT
GCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGC
GGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAG
ATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGA
CCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAAT

ATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATC
GCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGA
GTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACG
CCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAA
CTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATAC
CGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACA
ACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAG
ATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCAT
CAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCCCGCG
AGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAG
CGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCT
GCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACA
ACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAAC
AAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAG
CAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCA
GAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAA
GACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAG
CTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGC
GGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATC
ATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAA
CCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGA
TCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTG
GACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAA
CACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCA
ACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGAC
CAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTGAACG
CCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCC
TACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGAT
CAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCA
GCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTAT
AGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCTACCG
CGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCC
AGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCACCCT
CAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGgg
atcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttaccCatacg
atgttccagattacgcttaaGAATTCctagagctcgctgatcagcctcgactgtgccttctagttgcc
agccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctt
tcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggt
ggggcaggacagcaaggggaggattgggaagagaatagcaggcatgctggggaGGTACCGAACGCTG
ACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCGCGCGTGC
GCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGT
GTTCTGGGAAATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATGAGAC
CACATATAGTAATGAAATTATTGGCACGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAA
AATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTGGTACCaggaacccctagtgatggagtt
ggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgg
gctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatgcggtat
ttttcctttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgt
agcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccct
agcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctc
taaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgat
ttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtc
cacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattctt
ttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttt
aacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctga
tgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgc

tcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccg
tcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgat
aataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttat
ttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatat
tgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttg
ccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcac
gagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgt
tttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggca
agagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaa
agcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacact
gcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatggg
ggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtg
acaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactcta
gcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggc
ccttccggctggctggtttattgctgataaatctggagccggtgagcgtggaagccgcggtatcattg
cagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaact
atggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcaga
ccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtga
agatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagac
cccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaac
aaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaagg
taactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccac
ttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag
tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgg
gctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagataccta
cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcgg
cagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctg
tcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgg
aaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgt

SEQ ID NO: 21, Version 2 of vector 3

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGCTAGACTAGCATGCTGCCCATGTAAGGAGGCAAGGCCTGGGGA
CACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTGCCCCCCCCCCCCCAACACCTGCTGCC
TCTAAAAATAACCCTGCATGCCATGTTCCCGGCGAAGGGCCAGCTGTCCCCCGCCAGCTAGACTCAGC
ACTTAGTTTAGGAACCAGTGAGCAAGTCAGCCCTTGGGGCAGCCCATACAAGGCCATGGGGCTGGGCA
AGCTGCACGCCTGGGTCCGGGGTGGGCACGGTGCCCGGGCAACGAGCTGAAAGCTCATCTGCTCTCAG
GGGCCCCTCCCTGGGGACAGCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATATAACCCAGGG
GCACAGGGGCTGCCCTCATTCTACCACCACCTCCACAGCACAGACAGACACTCAGGAGCCAGCCAGCa
ccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGG
AACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACG
GGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGA
GCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTG
TTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAA
GGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAG
GCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGC
CGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGG
CGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGA
AGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACC

CGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTA
CGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACG
CCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTG
GAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCA
GATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGC
CCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAG
AACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGA
AGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCT
ATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAAC
GACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAA
AGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGA
GCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCC
CGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAA
CGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCA
AGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTG
AACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAA
CAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTACCTGA
GCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAG
GGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCA
GAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGC
GGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTT
CTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCT
GATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGG
AAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAA
GAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCG
GGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGG
GCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTG
ATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCT
GATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACC
TGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTG
AACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAA
GCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATG
TGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAG
ATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCT
GTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCT
ACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAG
ACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCA
CCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAA
AGggatcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttacccCa
tacgatgttccagattacgcttaaGAATTCctagagctcgctgatcagcctcgactgtgccttctagt
tgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgt
cctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtg
gggtggggcaggacagcaaggggaggattgggaagagaatagcaggcatgctggggaGGTACCGAAC
GCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCGCGC
GTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCT
ATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCTGTATG
AGACCACATATAGTAATGAAATTATTGGCACGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAG
GCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTGGTACCaggaaccctagtgatgg
agttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgc
ccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatgcg
gtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgcc
ctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcg
ccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaa
gctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaact
tgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttgg
agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctat

tcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaa
atttaacgcgaatttttaacaaaatattaacgtttacaatttttatggtgcactctcagtacaatctgct
ctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgt
ctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttc
accgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtca
tgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgt
ttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaata
atattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcat
tttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggt
gcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaaga
acgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccg
ggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcaca
gaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataa
cactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaaca
tgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgag
cgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttac
tctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgct
cggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtggaagccgcggtatc
attgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggc
aactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgt
cagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctag
gtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtc
agaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgc
aaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccg
aaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggcca
ccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctg
ccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcgg
tcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagata
cctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaa
gcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagt
cctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcct
atggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt

SEQ ID NO: 22, Version 3 of vector 3

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGGAGCTCCACCGCGGTGGCGGCCGTCCGCCtTCGGCACCATCCT
CACGACACCCAAATATGGCGACGGGTGAGGAATGGTGGGGAGTTATTTTTAGAGCGGTGAGGAAGGTG
GGCAGGCAGCAGGTGTTGGCGCTCTAAAAATAACTCCCGGGAGTTATTTTTAGAGCGGAGGAATGGTG
GACACCCAAATATGGCGACGGTTCCTCACCCGTCGCCATATTTGGGTGTCCGCCCTCGGCCGGGGCCG
CATTCCTGGGGGCCGGGCGGTGCTCCCGCCCGCCTCGATAAAAGGCTCCGGGGCCGGCGGCGGCCCAC
GAGCTACCCGGAGGAGCGGGAGGCGCCAAGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCG
ATATaccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCA
AGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAG
ACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAG
GCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGC
TGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGA
GTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAG
AAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGA
TCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAA

GACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCT
GCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCTGCTGG
AAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAA
TGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTA
CAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGA
AGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTG
AAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGG
CAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTA
TTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATC
CAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAA
GGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACA
CCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAG
CAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCAT
CCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGC
TGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAG
ACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAA
GATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATC
TGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGC
TTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCCAGTA
CCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGG
GCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCC
GTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCT
GCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCA
GCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGAC
GCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGT
GATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGT
ACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGC
CACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGA
CAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAA
AGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTG
AAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAA
CTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACA
AACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCC
CTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCT
GGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGA
AGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGC
GAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACAT
CACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCT
CCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAG
AAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAA
GAAAAAGggatcctacccatacgatgttccagattacgcttacccatacgatgttccagattacgctt
accCatacgatgttccagattacgcttaaGAATTCctagagctcgctgatcagcctcgactgtgccttt
ctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactccc
actgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggg
gggtggggtggggcaggacagcaaggggaggattgggaagagaatagcaggcatgctggggaGGTAC
CGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAG
CGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATG
TCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCT
GTATGAGACCACATATAGTAATGAAATTATTGGCACGTTTTAGTACTCTGGAAACAGAATCTACTAAA
ACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTGGTACCaggaacccctagt
gatggagttggccactccctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgccc
gacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctg
atgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtac
gcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgc
cagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttcccc
gtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaa

aaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgac
gttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgg
gctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaa
caaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaat
ctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacggg
cttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagagg
ttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaa
tgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccCta
tttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgctt
caataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgc
ggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagt
tgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgcccc
gaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattga
cgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccag
tcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagt
gataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgca
caacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacg
acgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaacta
cttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttct
gcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtggaagccgcg
gtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagt
caggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggta
actgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggta
tctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactga
gcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctg
cttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttt
ttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttta
ggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggc
tgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgc
agcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactg
agatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatcc
ggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatcttt
atagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcgg
agcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctca
catgt

SEQ ID NO: 23, Version 4 of vector 3

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGAaaaaatctcgccaacaagttgacgagataaacacggcattttgccttgttttagtagattct
gtttccagagtactaaaacacatttcctctctatacaaatgCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGT
GTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCCTCGAGGtttaaacaagcttgcatgtctaagctagacccttcagattaa
aaataactgaggtaagggcctgggtaggggaggtggtgtgagacgctcctgtctctcctctatctgcc
catcggccctttggggaggaggaatgtgcccaaggactaaaaaaaggccatggagccagaggggcgag
ggcaacagacctttcatgggcaaaccttggggccctgctgtctagcatgccccactacgggtctaggc
tgcccatgtaaggaggcaaggcctggggacacccgagatgcctggttataattaacccagacatgtgg
ctgccccccccccccccaacacctgctgcctctaaaaataaccctgtccctggtggatcccctgcatgc
gaagatcttcgaacaaggctgtgggggactgagggcaggctgtaacaggcttgggggccagggcttat
acgtgcctgggactcccaaagtattactgttccatgttcccggcgaagggccagctgtcccccgccag
ctagactcagcacttagtttaggaaccagtgagcaagtcagcccttggggcagcccatacaaggccat
ggggctgggcaagctgcacgcctgggtccggggtgggcacggtgcccgggcaacgagctgaaagctca
tctgctctcaggggcccctccctggggacagcccctcctggctagtcacaccctgtaggctcctctat

ataacccaggggcacaggggctgccctcattctaccaccacctccacagcacagacagacactcagga
gccagccagcggcgcgcccaccggtgccaccATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACG
GAGTCCCAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGC
ATCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGA
AAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCC
AGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAAC
CCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCT
GCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGT
CCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTG
GAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAA
AGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCT
ACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGG
AAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAG
CGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCA
GGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAG
AAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAG
AGTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCG
CCCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAG
AGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCA
GATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGG
ACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAG
GTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGT
GAAGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACG
ACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAG
AAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAA
GTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCA
TCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTG
TCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCG
GACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCC
TGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGAC
ATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAG
AGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCA
ATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAG
CACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGA
CAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCG
AAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAG
GACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTC
CACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACA
ATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAG
ACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTA
CGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAG
GTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGT
GACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGG
AAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTG
ATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGT
GAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTAT
GAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGC
CGGCCAGGCAAAAAAGAAAAAGggatcctacccatacgatgttccagattacgcttacccatacgatg
ttccagattacgcttaccCatacgatgttccagattacgcttaaGAATTCctagagctcgctgatcag
cctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctg
gaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtg
tcattctattctgggggtggggtggggcaggacagcaaggggaggattgggaagagaatagcaggc
atgctggggaGGTACCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTA
GGCGGGAACACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCC

TGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTGGATTTGGG
AATCTTATAAGTTCTGTATGAGACCACATATAGTAATGAAATTATTGGCACGTTTTAGTACTCTGGAA
ACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTGGTA
CCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggc
gaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgc
ctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtca
aagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgt
gaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgt
tcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacgg
cacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggt
tttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacac
tcaaccctatctcgggctattcttttgatttataagggatttgccgatttcggcctattggttaaaa
aatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtg
cactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctg
acgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagc
tgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcct
atttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatg
tgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgccct
tattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaag
atgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcctt
gagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggt
attatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttgg
ttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgct
gccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagct
aaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatg
aagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaacta
ttaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagt
tgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtg
agcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatc
tacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcact
gattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatt
tttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgag
ttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttct
gcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaag
agctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttcta
gtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaat
cctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagt
taccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacg
acctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaa
ggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaa
acgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgc
tcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttg
ctggccttttgctcacatgt

SEQ ID NO: 24, polynucleotide encoding SV40 intron

tctagaggatccggtactcgaggaactgaaaaaccagaaagttaactggtaagtttagtctt
tttgtcttttatttcaggtcccggatccggtggtggtgcaaatcaaagaactgctcctcagt
ggatgttgcctttacttctaggcctgtacggaagtgttac

SEQ ID NO: 25

NNGRR (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 26
NNGRRN (R = A or G, N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 27
NNGRRT (R = A or G, N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 28
NNGRRV (R = A or G, N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 29, Version 3 of vector 5
cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGACTCGAGGAGCTCCACCGCGGTGGCGGCCGTCCGCCttCGGCACCATCCTCACGACACCCAAA
TATGGCGACGGGTGAGGAATGGTGGGGAGTTATTTTTAGAGCGGTGAGGAAGGTGGGCAGGCAGCAGG
TGTTGGCGCTCTAAAAATAACTCCCGGGAGTTATTTTTAGAGCGGAGGAATGGTGGACACCCAAATAT
GGCGACGGTTCCTCACCCGTCGCCATATTTGGGTGTCCGCCCTCGGCCGGGGCCGCATTCCTGGGGGC
CGGGCGGTGCTCCCGCCCGCCTCGATAAAAGGCTCCGGGGCCGGCGGCGGCCCACGAGCTACCCGGAG
GAGCGGGAGGCGCCAAGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCCATGGtct
agaggatccggtactcgaggaactgaaaaaccagaaagttaactggtaagtttagtctttttgtctttt
tatttcaggtcccggatccggtggtggtgcaaatcaaagaactgctcctcagtggatgttgcctttac
ttctaggcctgtacggaagtgttacgccaCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACG
GAGTCCCAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGC
ATCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGA
AAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCC
AGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAAC
CCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCT
GCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGT
CCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTG
GAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAA
AGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCT
ACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGG
AAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAG
CGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCA
GGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAG
AAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAG
AGTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCG
CCCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAG
AGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCA
GATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGG
ACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAG
GTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGT
GAAGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACG
ACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAG
AAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAA
GTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCA
TCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTG
TCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCG
GACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCC
TGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGAC
ATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAG
AGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCA
ATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAG

CACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGA
CAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCG
AAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAG
GACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTC
CACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACA
ATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAG
ACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTA
CGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTA
AGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAG
GTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGT
GACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGG
AAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTG
ATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGT
GAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTAT
GAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGC
CGGCCAGGCAAAAAAGAAAAAGggatccGAATTCtagcaataaaggatcgtttatttttcattggaagc
gtgtgttggttttttgatcaggcgcgGGTACCAAAAATCTCGCCAACAAGTTGACGAGATAAACACGG
CATTTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAAACacatttcctctctatacaaatgC
GGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCA
TATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGT
CACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTT
GTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCCTCGACTAGTAGAAAAATCTCGC
CAACAAGTTGACGAGATAAACACGGCATTTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAA
ACGTGCCAATAATTTCATTACTATATCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAAT
CGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAAC
TACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATT
AATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGC
CCTCGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactga
ggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgc
gcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgc
atacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacg
cgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttct
cgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtg
ctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctga
tagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactgg
aacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctatt
ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatt
ttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaac
acccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtct
ccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtg
atacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcg
gggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatga
gacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgt
gtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaa
agtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta
agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgt
ggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaa
tgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattat
gcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccg
aaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgga
gctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgc
gcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcg
gataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcg
tagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggt

gcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaa
acttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatccctt
aacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcct
ttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgcc
ggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactg
tccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgct
ctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaag
acgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgg
agcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaa
gggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttcc
agggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttt
tgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg
gccttttgctggccttttgctcacatgt

SEQ ID NO: 30, Version 4 of vector 5

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGACTCGAGagcttgcatgtctaagctagacccttcagattaaaaataactgaggtaagggcctg
ggtaggggaggtggtgtgagacgctcctgtctctcctctatctgcccatcggccctttggggaggagg
aatgtgcccaaggactaaaaaaaggccatggagccagaggggcgagggcaacagacctttcatgggca
aaccttggggccctgctgtctagcatgccccactacgggtctaggctgcccatgtaaggaggcaaggc
ctggggacacccgagatgcctggttataattaacccagacatgtggctgccccccccccccccaacacc
tgctgcctctaaaaataaccctgtccctggtggatcccctgcatgcgaagatcttcgaacaaggctgt
gggggactgagggcaggctgtaacaggcttgggggccagggcttatacgtgcctgggactcccaaagt
attactgttccatgttcccggcgaagggccagctgtcccccgccagctagactcagcacttagtttag
gaaccagtgagcaagtcagcccttggggcagcccatacaaggccatggggctgggcaagctgcacgcc
tgggtccggggtgggcacggtgcccgggcaacgagctgaaagctcatctgctctcaggggcccctccc
tggggacagcccctcctggctagtcacaccctgtaggctcctctatataacccaggggcacaggggct
gccctcattctaccaccacctccacagcacagacagacactcaggagccagccagcCCATGGtctaga
ggatccggtactcgaggaactgaaaaaccagaaagttaactggtaagtttagtctttttgtcttttat
ttcaggtcccggatccggtggtggtgcaaatcaaagaactgctcctcagtggatgttgcctttacttc
taggcctgtacggaagtgttacgccaCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAG
TCCCAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATC
ATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAA
CAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGA
GAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCC
TACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCA
CCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCA
CCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAA
CGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGA
AGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACA
TCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAG
GACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGT
GAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGG
ACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAG
AAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGT
GACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCC
GGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGC
AGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGAT
CTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACG
AGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTG
GACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAA
GAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACA
TCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAG
CGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTA

CCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCC
CTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCC
TTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGAC
CCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGA
ATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATC
AACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGG
CCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATG
GCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCAC
CACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAA
GGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAA
CCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGAC
TACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCAC
CCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATG
ACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACC
TACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGA
GGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGT
ATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTC
GTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGAC
CGTGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAG
CTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATC
AAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAA
CATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTA
AGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAA
GTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGG
CCAGGCAAAAAAGAAAAAGggatccGAATTCtagcaataaaggatcgtttattttcattggaagcgtg
tgttggtttttgatcaggcgcgGGTACCAAAAATCTCGCCAACAAGTTGACGAGATAAACACGGCAT
TTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAAACacatttcctctctatacaaatgCGGT
GTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATAT
GATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCAC
GTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTA
TCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCCTCGACTAGTAGAAAAATCTCGCCAA
CAAGTTGACGAGATAAACACGGCATTTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAAACG
TGCCAATAATTTCATTACTATATCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGA
AATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTAC
CCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAAT
TCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCT
CGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggc
cgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgca
gctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcata
cgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgc
agcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgc
cacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctt
tacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatag
acggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaac
aacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctattggt
taaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttta
tggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacc
cgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccg
ggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgata
cgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcgggg
aaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagac
aataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtc
gcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagt
aaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaaga
tccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggc
gcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatga

cttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgca
gtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaag
gagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagct
gaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgca
aactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggat
aaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagc
cggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtag
ttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcc
tcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaact
tcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaac
gtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttt
tttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccgga
tcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtcc
ttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctg
ctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacg
atagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagc
gaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggg
agaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagg
gggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgt
gatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggcc
ttttgctggccttttgctcacatgt

SEQ ID NO: 31, codon optimized polynucleotide encoding *S. aureas* Cas9

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc
aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatttt
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tggtcaacga agaggacatc aagggctacc gggtgacaag cactggaaaa
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg
gtggacgatt tcattctgtc acccgtggtc aagcggagct tcatccagag catcaaagtg
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc
tccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc
agaagcgtgt ccttcgacaa ttcctttaac aacaaggtgc tggtcaagca ggaagagaac
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag
accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat

tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag
ctggacaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct
atgcccgaaa tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc
aagcatatca aggatttcaa ggactacaag tactctcacc gggtggataa aaagcccaac
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgatc
aacaaaagtc ccgagaagct gctgatgtac caccatgatc ctcagacata tcagaaactg
aagctgatta tggagcagta cggcgacgag aagaacccac tgtataagta ctatgaagag
actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc
aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
cgcaacaagg tggtcaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac
ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca
gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
gtcatcgggg tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact
taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat caaaacaatt
gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
gtgaagagca aaaagcaccc tcagattatc aaaaagggc [SEQ ID NO: 31]

SEQ ID NO: 32, codon optimized polynucleotide encoding *S. aureas* Cas9

atgaagcgga actacatcct gggcctggac atcggcatca ccagcgtggg ctacggcatc
atcgactacg agacacggga cgtgatcgat gccggcgtgc ggctgttcaa agaggccaac
gtggaaaaca acgagggcag gcggagcaag agaggcgcca gaaggctgaa gcggcggagg
cggcatagaa tccagagagt gaagaagctg ctgttcgact acaacctgct gaccgaccac
agcgagctga gcggcatcaa cccctacgag gccagagtga agggcctgag ccagaagctg
agcgaggaag agttctctgc cgccctgctg cacctggcca agagaagagg cgtgcacaac
gtgaacgagg tggaagagga caccggcaac gagctgtcca ccaaagagca gatcagccgg
aacagcaagg ccctggaaga gaaatacgtg gccgaactgc agctggaacg gctgaagaaa
gacggcgaag tgcggggcag catcaacaga ttcaagacca gcgactacgt gaaagaagcc
aaacagctgc tgaaggtgca gaaggcctac caccagctgg accagagctt catcgacacc
tacatcgacc tgctggaaac ccggcggacc tactatgagg gacctggcga gggcagcccc
ttcggctgga aggacatcaa agaatggtac gagatgctga tgggccactg cacctacttc
cccgaggaac tgcggagcgt gaagtacgcc tacaacgccg acctgtacaa cgccctgaac
gacctgaaca atctcgtgat caccagggac gagaacgaga agctggaata ttacgagaag
ttccagatca tcgagaacgt gttcaagcag aagaagaagc ccaccctgaa gcagatcgcc
aaagaaatcc tcgtgaacga agaggatatt aagggctaca gagtgaccag caccggcaag
cccgagttca ccaacctgaa ggtgtaccac gacatcaagg acattaccgc ccggaaagag
attattgaga acgccgagct gctggatcag attgccaaga tcctgaccat ctaccagagc
agcgaggaca tccaggaaga actgaccaat ctgaactccg agctgaccca ggaagagatc
gagcagatct ctaatctgaa gggctatacc ggcacccaca acctgagcct gaaggccatc
aacctgatcc tggacgagct gtggcacacc aacgacaacc agatcgctat cttcaaccgg
ctgaagctgg tgcccaagaa ggtggacctg tcccagcaga aagagatccc caccaccctg
gtggacgact tcatcctgag ccccgtcgtg aagagaagct tcatccagag catcaaagtg
atcaacgcca tcatcaagaa gtacggcctg cccaacgaca tcattatcga gctggcccgc
gagaagaact ccaaggacgc ccagaaaatg atcaacgaga tgcagaagcg gaaccggcag
accaacgagc ggatcgagga aatcatccgg accaccggca aagagaacgc caagtacctg
atcgagaaga tcaagctgca cgacatgcag gaaggcaagt gcctgtacag cctggaagcc
atccctctgg aagatctgct gaacaacccc ttcaactatg aggtggacca catcatcccc
agaagcgtgt ccttcgacaa cagcttcaac aacaaggtgc tcgtgaagca ggaagaaaac
agcaagaagg gcaaccggac cccattccag tacctgagca gcagcgacag caagatcagc
tacgaaacct tcaagaagca catcctgaat ctggccaagg gcaagggcag aatcagcaag
accaagaaag agtatctgct ggaagaacgg gacatcaaca ggttctccgt gcagaaagac

ttcatcaacc ggaacctggt ggataccaga tacgccacca gaggcctgat gaacctgctg
cggagctact tcagagtgaa caacctggac gtgaaagtga agtccatcaa tggcggcttc
accagctttc tgcggcggaa gtggaagttt aagaaagagc ggaacaaggg gtacaagcac
cacgccgagg acgccctgat cattgccaac gccgatttca tcttcaaaga gtggaagaaa
ctggacaagg ccaaaaaagt gatggaaaac cagatgttcg aggaaaagca ggccgagagc
atgcccgaga tcgaaaccga gcaggagtac aaagagatct tcatcacccc ccaccagatc
aagcacatta aggacttcaa ggactacaag tacagccacc gggtggacaa gaagcctaat
agagagctga ttaacgacac cctgtactcc acccggaagg acgacaaggg caacaccctg
atcgtgaaca atctgaacgg cctgtacgac aaggacaatg acaagctgaa aaagctgatc
aacaagagcc ccgaaaagct gctgatgtac caccacgacc cccagaccta ccagaaactg
aagctgatta tggaacagta cggcgacgag aagaatcccc tgtacaagta ctacgaggaa
accgggaact acctgaccaa gtactccaaa aaggacaacg gccccgtgat caagaagatt
aagtattacg gcaacaaact gaacgcccat ctggacatca ccgacgacta ccccaacagc
agaaacaagg tcgtgaagct gtccctgaag ccctacagat tcgacgtgta cctggacaat
ggcgtgtaca agttcgtgac cgtgaagaat ctggatgtga tcaaaaaaga aaactactac
gaagtgaata gcaagtgcta tgaggaagct aagaagctga agaagatcag caaccaggcc
gagtttatcg cctccttcta caacaacgat ctgatcaaga tcaacggcga gctgtataga
gtgatcggcg tgaacaacga cctgctgaac cggatcgaag tgaacatgat cgacatcacc
taccgcgagt acctggaaaa catgaacgac aagaggcccc ccaggatcat taagacaatc
gcctccaaga cccagagcat taagaagtac agcacagaca ttctgggcaa cctgtatgaa
gtgaaatcta agaagcaccc tcagatcatc aaaaagggc [SEQ ID NO: 32]

SEQ ID NO: 33, codon optimized polynucleotide encoding *S. aureas* Cas9

atgaagcgca actacatcct cggactggac atcggcatta cctccgtggg atacggcatc
atcgattacg aaactaggga tgtgatcgac gctggagtca ggctgttcaa agaggcgaac
gtggagaaca acgaggggcg gcgctcaaag agggggccc gccggctgaa gcgccgccgc
agacatagaa tccagcgcgt gaagaagctg ctgttcgact acaaccttct gaccgaccac
tccgaacttt ccggcatcaa cccatatgag gctagagtga agggattgtc ccaaaagctg
tccgaggaag agttctccgc cgcgttgctc cacctcgcca agcgcagggg agtgcacaat
gtgaacgaag tggaagaaga taccggaaac gagctgtcca ccaaggagca gatcagccgg
aactccaagg ccctggaaga gaaatacgtg gcggaactgc aactggagcg gctgaagaaa
gacggagaag tgcgcggctc gatcaaccgc ttcaagacct cggactacgt gaaggaggcc
aagcagctcc tgaaagtgca aaaggcctat caccaacttg accagtcctt tatcgatacc
tacatcgatc tgctcgagac tcggcggact tactacgagg gtccagggga gggctcccca
tttggttgga aggatattaa ggagtggtac gaaatgctga tgggacactg cacatacttc
cctgaggagc tgcggagcgt gaaatacgca tacaacgcag acctgtacaa cgcgctgaac
gacctgaaca atctcgtgat cacccgggac gagaacgaaa agctcgagta ttacgaaaag
ttccagatta ttgagaacgt gttcaaacag aagaagaagc cgacactgaa gcagattgcc
aaggaaatcc tcgtgaacga agaggacatc aagggctatc gagtgacctc aacgggaaag
ccggagttca ccaatctgaa ggtctaccac gacatcaaag acattaccgc ccggaaggag
atcattgaga acgcggagct gttggaccag attgcgaaga ttctgaccat ctaccaatcc
tccgaggata ttcaggaaga actcaccaac ctcaacagcg aactgaccca ggaggagata
gagcaaatct ccaacctgaa gggctacacc ggaactcata acctgagcct gaaggccatc
aacttgatcc tggacgagct gtggcacacc aacgataacc agatcgctat tttcaatcgg
ctgaagctgg tccccaagaa agtggacctc tcacaacaaa aggagatccc tactaccctt
gtggacgatt tcattctgtc ccccgtggtc aagagaagct tcatacagtc aatcaaagtg
atcaatgcca ttatcaagaa atacggtctg cccaacgaca ttatcattga gctcgcccgc
gagaagaact cgaaggacgc ccagaagatg attaacgaaa tgcagaagag gaaccgacag
actaacgaac ggatcgaaga aatcatccgg accaccggga aggaaaacgc gaagtacctg
atcgaaaaga tcaagctcca tgacatgcag gaaggaaagt gtctgtactc gctggaggcc
attccgctgg aggacttgct gaacaaccct tttaactacg aagtggatca tatcattccg
aggagcgtgt cattcgacaa ttccttcaac aacaaggtcc tcgtgaagca ggaggaaaac
tcgaagaagg gaaaccgcac gccgttccag tacctgagca gcagcgactc caagatttcc
tacgaaacct tcaagaagca catcctcaac ctggcaaagg ggaagggtcg catctccaag
accaagaagg aatatctgct ggaagaaaga gacatcaaca gattctccgt gcaaaaggac

ttcatcaacc gcaacctcgt ggatactaga tacgctactc ggggtctgat gaacctcctg
agaagctact ttagagtgaa caatctggac gtgaaggtca agtcgattaa cggaggtttc
acctccttcc tgcggcgcaa gtggaagttc aagaaggaac ggaacaaggg ctacaagcac
cacgccgagg acgccctgat cattgccaac gccgacttca tcttcaaaga atggaagaaa
cttgacaagg ctaagaaggt catggaaaac cagatgttcg aagaaaagca ggccgagtct
atgcctgaaa tcgagactga acaggagtac aaggaaatct ttattacgcc acaccagatc
aaacacatca aggatttcaa ggattacaag tactcacatc gcgtggacaa aaagccgaac
agggaactga tcaacgacac cctctactcc acccggaagg atgacaaagg gaataccctc
atcgtcaaca accttaacgg cctgtacgac aaggacaacg ataagctgaa gaagctcatt
aacaagtcgc ccgaaaagtt gctgatgtac caccacgacc ctcagactta ccagaagctc
aagctgatca tggagcagta tggggacgag aaaaaccccgt tgtacaagta ctacgaagaa
actgggaatt atctgactaa gtactccaag aaagataacg gccccgtgat taagaagatt
aagtactacg gcaacaagct gaacgcccat ctggacatca ccgatgacta ccctaattcc
cgcaacaagg tcgtcaagct gagcctcaag ccctaccggt ttgatgtgta ccttgacaat
ggagtgtaca agttcgtgac tgtgaagaac cttgacgtga tcaagaagga gaactactac
gaagtcaact ccaagtgcta cgaggaagca aagaagttga agaagatctc gaaccaggcc
gagttcattg cctccttcta taacaacgac ctgattaaga tcaacggcga actgtaccgc
gtcattggcg tgaacaacga tctcctgaac cgcatcgaag tgaacatgat cgacatcact
taccgggaat acctggagaa tatgaacgac aagcgcccgc cccggatcat taagactatc
gcctcaaaga cccagtcgat caagaagtac agcaccgaca tcctgggcaa cctgtacgag
gtcaaatcga agaagcaccc ccagatcatc aagaaggga [SEQ ID NO: 33]

SEQ ID NO: 34, codon optimized polynucleotide encoding *S. aureas* Cas9

atggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccaagcggaactacatcct
gggcctggacatcggcatcaccagcgtgggctacggcatcatcgactacgagacacgggacgtgatcg
atgccggcgtgcggctgttcaaagaggccaacgtggaaaacaacgagggcaggcggagcaagagaggc
gccagaaggctgaagcggcggaggcggcatagaatccagagagtgaagaagctgctgttcgactacaa
cctgctgaccgaccacagcgagctgagcggcatcaaccctacgaggccagagtgaagggcctgagcc
agaagctgagcgaggaagagttctctgccgccctgctgcacctggccaagagaagaggcgtgcacaac
gtgaacgaggtggaagaggacaccggcaacgagctgtccaccagagagcagatcagccggaacagcaa
ggccctggaagagaaatacgtggccgaactgcagctggaacggctgaagaaagacggcgaagtgcggg
gcagcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagctgctgaaggtgcagaag
gcctaccaccagctggaccagagcttcatcgacacctacatcgacctgctggaaacccggcggaccta
ctatgagggacctggcgagggcagccccttcggctggaaggacatcaaagaatggtacgagatgctga
tgggccactgcacctacttccccgaggaactgcggagcgtgaagtacgcctacaacgccgacctgtac
aacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgagaagctggaatattacga
gaagttccagatcatcgagaacgtgttcaagcagaagaagaagcccaccctgaagcagatcgccaaag
aaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccggcaagcccgagttcacc
aacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagattattgagaacgccgagct
gctggatcagattgccaagatcctgaccatctaccagagcagcgaggacatccaggaagaactgacca
atctgaactccgagctgacccaggaagagatcgagcagatctctaatctgaagggctataccggcacc
cacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagat
cgctatcttcaaccggctgaagctggtgcccaagaaggtggacctgtcccagcagaaagagatcccca
ccaccctggtggacgacttcatcctgagccccgtcgtgaagagaagcttcatccagagcatcaaagtg
atcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcgagctggcccgcgagaagaa
ctccaaggacgcccagaaaatgatcaacgagatgcagaagcggaaccggcagaccaacgagcggatcg
aggaaatcatccggaccaccggcaaagagaacgccaagtacctgatcgagaagatcaagctgcacgac
atgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagatctgctgaacaacccctt
caactatgaggtggaccacatcatccccagaagcgtgtccttcgacaacagcttcaacaacaaggtgc
tcgtgaagcaggaagaaaacagcaagaagggcaaccggacccccattccagtacctgagcagcagcgac
agcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaagggcaagggcagaatcag
caagaccaagaaagagtatctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttca
tcaaccggaacctggtggataccagatacgccaccagaggcctgatgaacctgctgcggagctacttc
agagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttcaccagctttctgcggcggaa
gtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgaggacgccctgatcattgcca

acgccgatttcatcttcaaagagtggaagaaactggacaaggccaaaaaagtgatggaaaaccagatg
ttcgaggaaaggcaggccgagagcatgcccgagatcgaaaccgagcaggagtacaaagagatcttcat
caccccccaccagatcaagcacattaaggacttcaaggactacaagtacagccaccgggtggacaaga
agcctaatagagagctgattaacgacaccctgtactccacccggaaggacgacaagggcaacaccctg
atcgtgaacaatctgaacggcctgtacgacaaggacaatgacaagctgaaaaagctgatcaacaagag
ccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaactgaagctgattatggaac
agtacggcgacgagaagaatcccctgtacaagtactacgaggaaaccgggaactacctgaccaagtac
tccaaaaaggacaacggccccgtgatcaagaagattaagtattacggcaacaaactgaacgcccatct
ggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagctgtccctgaagccctacagat
tcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtgatcaaaaaa
gaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaacca
ggccgagtttatcgcctccttctacaacaacgatctgatcaagatcaacggcgagctgtatagagtga
tcggcgtgaacaacgacctgctgaaccggatcgaagtgaacatgatcgacatcacctaccgcgagtac
ctggaaaacatgaacgacaagaggcccccccaggatcattaagacaatcgcctccaagacccagagcat
taagaagtacagcacagacattctgggcaacctgtatgaagtgaaatctaagaagcaccctcagatca
tcaaaaagggcaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag [SEQ ID
NO: 34]

SEQ ID NO: 35, codon optimized polynucleotide encoding *S. aureas* Cas9

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc
aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc
gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg
ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc
ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac
ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc
ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc
cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag
gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg
gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac
tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag
agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca
ggagaaggga gcccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga
cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatct
tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg
gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca
ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg
acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc
acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg
actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg
acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg
tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt
gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag
atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc
cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt
atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag
aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag
aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg
tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc
gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc
aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca
gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag
ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc
tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc
ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc
atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac

aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt
aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag
aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc
actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg
gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat
aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag
ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag
acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat
aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc
gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac
gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat
gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa
aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag
attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat
ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat
atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccctcga
attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg
ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa
ttc [SEQ ID NO: 35]

SEQ ID NO: 36, codon optimized polynucleotide encoding *S. aureas* Cas9

atggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccaagcggaactacatcct
gggcctggacatcggcatcaccagcgtgggctacggcatcatcgactacgagacacgggacgtgatcg
atgccggcgtgcggctgttcaaagaggccaacgtggaaaacaacgagggcaggcggagcaagagaggc
gccagaaggctgaagcggcggaggcggcatagaatccagagagtgaagaagctgctgttcgactacaa
cctgctgaccgaccacagcgagctgagcggcatcaacccctacgaggccagagtgaagggcctgagcc
agaagctgagcgaggaagagttctctgccgccctgctgcacctggccaagagaagaggcgtgcacaac
gtgaacgaggtggaagaggacaccggcaacgagctgtccaccaaagagcagatcagccggaacagcaa
ggccctggaagagaaatacgtggccgaactgcagctggaacggctgaagaaagacggcgaagtgcggg
gcagcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagctgctgaaggtgcagaag
gcctaccaccagctggaccagagcttcatcgacacctacatcgacctgctggaaacccggcggaccta
ctatgagggacctggcgagggcagccccttcggctggaaggacatcaaagaatggtacgagatgctga
tgggccactgcacctacttccccgaggaactgcggagcgtgaagtacgcctacaacgccgacctgtac
aacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgagaagctggaatattacga
gaagttccagatcatcgagaacgtgttcaagcagaagaagaagcccaccctgaagcagatcgccaaag
aaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccggcaagcccgagttcacc
aacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagattattgagaacgccgagct
gctggatcagattgccaagatcctgaccatctaccagagcagcgaggacatccaggaagaactgacca
atctgaactccgagctgacccaggaagagatcgagcagatctctaatctgaagggctataccggcacc
cacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcacaccaacgacaaccagat
cgctatcttcaaccggctgaagctggtgcccaagaaggtggacctgtcccagcagaaagagatcccca
ccaccctggtggacgacttcatcctgagccccgtcgtgaagagaagcttcatccagagcatcaaagtg
atcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcgagctggcccgcgagaagaa
ctccaaggacgcccagaaaatgatcaacgagatgcagaagcggaaccggcagaccaacgagcggatcg
aggaaatcatccggaccaccggcaaagagaacgccaagtacctgatcgagaagatcaagctgcacgac
atgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagatctgctgaacaacccctt
caactatgaggtggaccacatcatccccagaagcgtgtccttcgacaacagcttcaacaacaaggtgc
tcgtgaagcaggaagaaaacagcaagaagggcaaccggacccccattccagtacctgagcagcagcgac
agcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaagggcaagggcagaatcag
caagaccaagaaagagtatctgctggaagaacgggacatcaacaggttctccgtgcagaaagacttca
tcaaccggaacctggtggataccagatacgccaccagaggcctgatgaacctgctgcggagctacttc
agagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttcaccagctttctgcggcggaa
gtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgaggacgccctgatcattgcca
acgccgatttcatcttcaaagagtggaagaaactggacaaggccaaaaaagtgatggaaaaccagatg
ttcgaggaaaagcaggccgagagcatgcccgagatcgaaaccgagcaggagtacaaagagatcttcat

cacccccaccagatcaagcacattaaggacttcaaggactacaagtacagccaccgggtggacaaga
agcctaatagagagctgattaacgacaccctgtactccacccggaaggacgacaagggcaacaccctg
atcgtgaacaatctgaacggcctgtacgacaaggacaatgacaagctgaaaaagctgatcaacaagag
ccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaactgaagctgattatggaac
agtacggcgacgagaagaatcccctgtacaagtactacgaggaaaccgggaactacctgaccaagtac
tccaaaaaggacaacggccccgtgatcaagaagattaagtattacggcaacaaactgaacgcccatct
ggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagctgtccctgaagccctacagat
tcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatctggatgtgatcaaaaaa
gaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaagctgaagaagatcagcaacca
ggccgagtttatcgcctccttctacaacaacgatctgatcaagatcaacggcgagctgtatagagtga
tcggcgtgaacaacgacctgctgaaccggatcgaagtgaacatgatcgacatcacctaccgcgagtac
ctggaaaacatgaacgacaagaggcccccaggatcattaagacaatcgcctccaagacccagagcat
taagaagtacagcacagacattctgggcaacctgtatgaagtgaaatctaagaagcaccctcagatca
tcaaaaagggcaaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag [seq id
no: 36]

SEQ ID NO: 37, polynucleotide sequence of *S. aureus* Cas9

aagcggaactacatcctgggcctggacatcggcatcaccagcgtgggctacggcatcatcgactacga
gacacgggacgtgatcgatgccggcgtgcggctgttcaaagaggccaacgtggaaaacaacgagggca
ggcggagcaagagaggcgccagaaggctgaagcggcggaggcggcatagaatccagagagtgaagaag
ctgctgttcgactacaacctgctgaccgaccacagcgagctgagcggcatcaacccctacgaggccag
agtgaagggcctgagccagaagctgagcgaggaagagttctctgccgccctgctgcacctggccaaga
gaagaggcgtgcacaacgtgaacgaggtggaagaggacaccggcaacgagctgtccaccaaagagcag
atcagccggaacagcaaggccctggaagagaaatacgtggccgaactgcagctggaacggctgaagaa
agacggcgaagtgcggggcagcatcaacagattcaagaccagcgactacgtgaaagaagccaaacagc
tgctgaaggtgcagaaggcctaccaccagctggaccagagcttcatcgacacctacatcgacctgctg
gaaacccggcggacctactatgagggacctggcgagggcagcccctteggctggaaggacatcaaaga
atggtacgagatgctgatgggccactgcacctacttccccgaggaactgcggagcgtgaagtacgcct
acaacgccgacctgtacaacgccctgaacgacctgaacaatctcgtgatcaccagggacgagaacgag
aagctggaatattacgagaagttccagatcatcgagaacgtgttcaagcagaagaagaagcccaccct
gaagcagatcgccaaagaaatcctcgtgaacgaagaggatattaagggctacagagtgaccagcaccg
gcaagcccgagttcaccaacctgaaggtgtaccacgacatcaaggacattaccgcccggaaagagatt
attgagaacgccgagctgctggatcagattgccaagatcctgaccatctaccagagcagcgaggacat
ccaggaagaactgaccaatctgaactccgagctgacccaggaagagatcgagcagatctctaatctga
agggctataccggcacccacaacctgagcctgaaggccatcaacctgatcctggacgagctgtggcac
accaacgacaaccagatcgctatcttcaaccggctgaagctggtgcccaagaaggtggacctgtccca
gcagaaagagatccccaccaccctggtggacgacttcatcctgagccccgtcgtgaagagaagcttca
tccagagcatcaaagtgatcaacgccatcatcaagaagtacggcctgcccaacgacatcattatcgag
ctggcccgcgagaagaactccaaggacgcccagaaaatgatcaacgagatgcagaagcggaaccggca
gaccaacgagcggatcgaggaaatcatccggaccaccggcaaagagaacgccaagtacctgatcgaga
agatcaagctgcacgacatgcaggaaggcaagtgcctgtacagcctggaagccatccctctggaagat
ctgctgaacaacccccttcaactatgaggtggaccacatcatccccagaagcgtgtccttcgacaacag
cttcaacaacaaggtgctcgtgaagcaggaagaaaacagcaagaagggcaaccggacccccattccagt
acctgagcagcagcgacagcaagatcagctacgaaaccttcaagaagcacatcctgaatctggccaag
ggcaagggcagaatcagcaagaccaagaaagagtatctgctggaagaacgggacatcaacaggttctc
cgtgcagaaagacttcatcaaccggaacctggtggataccagatacgccaccagaggcctgatgaacc
tgctgcggagctacttcagagtgaacaacctggacgtgaaagtgaagtccatcaatggcggcttcacc
agctttctgcggcggaagtggaagtttaagaaagagcggaacaaggggtacaagcaccacgccgagga
cgccctgatcattgccaacgccgatttcatcttcaaagagtggaagaaactggacaaggccaaaaaag
tgatggaaaaccagatgttcgaggaaaagcaggccgagagcatgcccgagatcgaaaccgagcaggag
tacaaagagatcttcatcacccccaccagatcaagcacattaaggacttcaaggactacaagtacag
ccaccgggtggacaagaagcctaatagagagctgattaacgacaccctgtactccacccggaaggacg
acaagggcaacaccctgatcgtgaacaatctgaacggcctgtacgacaaggacaatgacaagctgaaa
aagctgatcaacaagagccccgaaaagctgctgatgtaccaccacgacccccagacctaccagaaact
gaagctgattatggaacagtacggcgacgagaagaatcccctgtacaagtactacgaggaaaccggga

actacctgaccaagtactccaaaaaggacaacggccccgtgatcaagaagattaagtattacggcaac
aaactgaacgcccatctggacatcaccgacgactaccccaacagcagaaacaaggtcgtgaagctgtc
cctgaagccctacagattcgacgtgtacctggacaatggcgtgtacaagttcgtgaccgtgaagaatc
tggatgtgatcaaaaaagaaaactactacgaagtgaatagcaagtgctatgaggaagctaagaagctg
aagaagatcagcaaccaggccgagtttatcgcctccttctacaacaacgatctgatcaagatcaacgg
cgagctgtatagagtgatcggcgtgaacaacgacctgctgaaccggatcgaagtgaacatgatcgaca
tcacctaccgcgagtacctggaaaacatgaacgacaagaggcccccccaggatcattaagacaatcgcc
tccaagacccagagcattaagaagtacagcacagacattctgggcaacctgtatgaagtgaaatctaa
gaagcaccctcagatcatcaaaaagggc [SEQ ID NO: 37]

SEQ ID NO: 38, pDO242 (SaCas9 used in all JCR89/91 projects and JCR157/160 projects for in vitro work; SaCas9 in uppercase)

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttta
accaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgtt
gttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgt
ctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttttggggtcgaggtgccgta
aagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcgaacgtg
gcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgct
gcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggc
tgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggga
tgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggc
cagtgagcgcgcgtaatacgactcactatagggcgaattgggtacCtttaattctagtactatgcaTg
cgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccata
tatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcc
cattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgg
gtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccc
tattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttc
ctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatc
aatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggag
tttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaa
tgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactaccggtgccacc
ATGAAAAGGAACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTA
TGAAACAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAACAATGAGG
GACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGACGGAGAAGGCACAGAATCCAGAGGGTGAAG
AAACTGCTGTTCGATTACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAATTAATCCTTATGAAGC
CAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTA
AGCGCCGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAA
CAGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGGAACGGCTGAA
GAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAAGCGACTACGTCAAAGAAGCCAAGC
AGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGATACTTATATCGACCTG
CTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTCGGATGGAAAGACATCAA
GGAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACG
CTTATAACGCAGATCTGTACAACGCCCTGAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAAC
GAGAAACTGGAATACTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTAC
ACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCA
CTGGAAAACCAGAGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAGAA
ATCATTGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGA
CATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATC
TGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGG
CATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTGGTCCCAAAAAAGGTGGACCTGAG
TCAGCAGAAAGAGATCCCAACCACACTGGTGGACGATTTCATTCTGTCACCCGTGGTCAAGCGGAGCT
TCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATTATC
GAGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCG
GCAGACCAATGAACGCATTGAAGAGATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTG
AAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTGGAG

GACCTGCTGAACAATCCATTCAACTACGAGGTCGATCATATTATCCCCAGAAGCGTGTCCTTCGACAA
TTCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAAGGGCAATAGGACTCCTTTCC
AGTACCTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTCTGAATCTGGCC
AAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATT
CTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTGGACACAAGATACGCTACTCGCGGCCTGATGA
ATCTGCTGCGATCCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTC
ACATCTTTTCTGAGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGA
AGATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGA
AAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGGCCGAATCTATGCCCGAAATCGAGACAGAACAG
GAGTACAAGGAGATTTTCATCACTCCTCACCAGATCAAGCATATCAAGGATTTCAAGGACTACAAGTA
CTCTCACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAAAAG
ACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAATGACAAGCTG
AAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAA
ACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTG
GGAACTACCTGACCAAGTATAGCAAAAAGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGG
AACAAGCTGAATGCCCATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGTCAAGCT
GTCACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACTGTCAAGA
ATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAG
CTGAAAAAGATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAA
TGGCGAACTGTATAGGGTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTG
ACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAGCGCCCCCCTCGAATTATCAAAACAATT
GCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTATGAGGTGAAGAG
CAAAAAGCACCCTCAGATTATCAAAAAGGGCagcggaggcaagcgtcctgctgctactaagaaagctg
gtcaagctaagaaaaagaaaggatcctacccatacgatgttccagattacgcttaagaattcctagag
ctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgcct
tccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattg
tctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaag
agaatagcaggcatgctggggaggtagcggccgcCCgcggtggagctccagcttttgttccctttagt
gagggttaattgcgcgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctc
acaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagcta
actcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcatt
aatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcact
gactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggtt
atccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaacc
gtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcga
cgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctc
cctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaa
gcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctg
ggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtc
caacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggt
atgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtattt
ggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaca
aaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctc
aagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggatt
ttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatc
aatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatct
cagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgg
gagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagattt
atcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctcca
tccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgtt
gttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc
ccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctc
cgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattct
cttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgaga
atagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagca
gaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctg

ttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccag
cgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaat
gttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagc
ggatacatatttgaatgtatttagaaaaataaacaaatagggqttccgcgcacatttccccgaaaagt
gccac [SEQ ID NO:38]

SEQ ID NO: 39, amino acid sequence of an *S. aureus* Cas9 molecule
MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRH
RIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEV
EEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKV
QKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVK
YAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDI
KGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLN
SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQK
EIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQK
RNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHI
IPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISK
TKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTS
FLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEI
ETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLN
GLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKY
SKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKN
LDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRI
EVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
[SEQ ID NO: 39]

SEQ ID NO: 40, amino acid sequence of an *S. aureus* Cas9
KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
[SEQ ID NO: 40]

SEQ ID NO: 41, Version 1 of vector 5
cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGACTCGAGTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGA
GAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAA
GTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTC

GCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTGTCGTGACCGCGGCCATGGtc
tagaggatccggtactcgaggaactgaaaaaccagaaagttaactggtaagtttagtctttttgtctt
ttatttcaggtcccggatccggtggtggtgcaaatcaaagaactgctcctcagtggatgttgccttta
cttctaggcctgtacggaagtgttacgccaCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCAC
GGAGTCCCAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGG
CATCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGG
AAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATC
CAGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAA
CCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC
TGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTG
TCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCT
GGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGA
AAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACC
TACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTG
GAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGA
GCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACC
AGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAA
GAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACA
GAGTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACC
GCCCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCA
GAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGC
AGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTG
GACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAA
GGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCG
TGAAGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAAC
GACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCA
GAAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCA
AGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGT
GTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACC
GGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATC
CTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGA
CATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCA
GAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATC
AATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAA
GCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGG
ACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATC
GAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAA
GGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACT
CCACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGAC
AATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCA
GACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACT
ACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATT
AAGTATTACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAA
GGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCG
TGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAG
GAAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCT
GATCAAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAG
TGAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATC
ATTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTA
TGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGG
CCGGCCAGGCAAAAAAGAAAAAGggatccGAATTCtagcaataaaggatcgtttattttcattggaag
cgtgtgttggtttttgatcaggcgcgGGTACCAAAAATCTCGCCAACAAGTTGACGAGATAAACACG
GCATTTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAAACacatttcctctctatacaaatg
CGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGC
ATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACG

TCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCT
TGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCCTCGACTAGTAGAAAAATCTCG
CCAACAAGTTGACGAGATAAACACGGCATTTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAA
AACGTGCCAATAATTTCATTACTATATCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAA
TCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAA
CTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAAT
TAATTCCAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGG
CCCTCGGTACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactg
aggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcg
cgcagctgcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccg
catacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttac
gcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttc
tcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagt
gctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctg
atagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactg
gaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctat
tggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaat
tttatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaa
cacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtc
tccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgt
gatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttc
ggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatg
agacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccg
tgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtga
aagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggt
aagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatg
tggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcaga
atgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaatta
tgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggacc
gaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgg
agctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttg
cgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggc
ggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctg
gagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatc
gtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagatagg
tgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaa
aacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatccct
taacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcc
ttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgc
cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatact
gtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgc
tctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaa
gacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttg
gagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccga
agggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttc
cagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgattt
ttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcct
ggccttttgctggccttttgctcacatgt

SEQ ID NO: 42, Version 2 of vector 5

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgccc
ggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcc
TCTAGACTCGAGCTAGACTAGCATGCTGCCCATGTAAGGAGGCAAGGCCTGGGGACACCCGAGATGCC
TGGTTATAATTAACCCAGACATGTGGCTGCCCCCCCCCCCCCAACACCTGCTGCCTCTAAAAATAACC
CTGCATGCCATGTTCCCGGCGAAGGGCCAGCTGTCCCCCGCCAGCTAGACTCAGCACTTAGTTTAGGA

ACCAGTGAGCAAGTCAGCCCTTGGGGCAGCCCATACAAGGCCATGGGGCTGGGCAAGCTGCACGCCTG
GGTCCGGGGTGGGCACGGTGCCCGGGCAACGAGCTGAAAGCTCATCTGCTCTCAGGGGCCCCTCCCTG
GGGACAGCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATATAACCCAGGGGCACAGGGGCTGC
CCTCATTCTACCACCACCTCCACAGCACAGACAGACACTCAGGAGCCAGCCAGCCATGGtctagagga
tccggtactcgaggaactgaaaaaccagaaagttaactggtaagtttagtctttttgtcttttatttc
aggtcccggatccggtggtggtgcaaatcaaagaactgctcctcagtggatgttgcctttacttctag
gcctgtacggaagtgttacgccaCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCC
CAGCAGCCAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATC
GACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAA
CGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAG
TGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTAC
GAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCT
GGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCA
AAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGG
CTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGC
CAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCG
ACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGAC
ATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAA
GTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAG
CCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGAC
CAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGA
AAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGC
GAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTC
TAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGC
TGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGAC
CTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAG
AAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCA
TTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGG
AACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCT
GATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTC
TGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTC
GACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCC
ATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATC
TGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAAC
AGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCT
GATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCG
GCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCAC
GCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGC
CAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCG
AGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTAC
AAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCG
GAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACA
AGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTAC
CAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGA
AACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATT
ACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTG
AAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGT
GAAGAATCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTA
AGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAG
ATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACAT
GATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGA
CAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTG
AAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCA
GGCAAAAAAGAAAAAGggatccGAATTCtagcaataaaggatcgtttattttcattggaagcgtgtgt
tggtttttgatcaggcgcgGGTACCAAAAATCTCGCCAACAAGTTGACGAGATAAACACGGCATTTT

GCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAAACacatttcctctctatacaaatgCGGTGTT
TCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGAT
AGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTA
TTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCCAATTATCTCTCTAACAGCCTTGTATCG
TATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCCTCGACTAGTAGAAAAATCTCGCCAACAA
GTTGACGAGATAAACACGGCATTTTGCCTTGTTTTAGTAGATTCTGTTTCCAGAGTACTAAAACGTGC
CAATAATTTCATTACTATATCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAAT
ACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCA
AGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCC
AATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCGG
TACCaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgg
gcgaccaaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagct
gcctgcaggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgt
caaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagc
gtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac
gttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttac
ggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacg
gtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaac
actcaaccctatctcgggctattcttttgatttataagggattttgccgatttcggcctattggttaa
aaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatgg
tgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgc
tgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggga
gctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgc
ctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaa
tgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaat
aaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcc
cttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaa
agatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcc
ttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcg
gtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgactt
ggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtg
ctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggag
ctaaccgtttttgcacaacatggggatcatgtaactcgccttgatcgttgggaaccggagctgaa
tgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaac
tattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaa
gttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccgg
tgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagtta
tctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctca
ctgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttca
tttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtg
agttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttt
ctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatca
agagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttc
tagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgcta
atcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgata
gttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaa
cgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggaga
aaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggg
aaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgat
gctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttt
tgctggccttttgctcacatgt

CLAIMS

What is claimed is:

1. A CRISPR-Cas system comprising one or more vectors encoding a composition, the composition comprising:
   (a) a first guide RNA (gRNA) molecule targeting intron 44 of dystrophin;
   (b) a second gRNA molecule targeting intron 55 of dystrophin; and
   (c) a Cas9 protein; and
   (d) one or more Cas9 gRNA scaffolds.

2. The system of claim 1, wherein the system comprises a single vector.

3. The system of claim 1, wherein the system comprises two or more vectors, wherein the two or more vectors comprises a first vector and a second vector.

4. The system of claim 3, wherein
   (a) the first vector encodes the first gRNA molecule and the second gRNA molecule; and
   (b) the second vector encodes the Cas9 protein.

5. The system of claim 3, wherein
   (a) the first vector encodes the first gRNA molecule; and
   (b) the second vector encodes the second gRNA molecule.

6. The system of claim 5, wherein the first vector further encodes the Cas9 protein.

7. The system of claim 5 or 6, wherein the second vector further encodes the Cas9 protein.

8. The system of any one of claims 1-7, wherein the expression of the Cas9 protein is driven by a constitutive promoter or a muscle-specific promoter.

9. The system of claim 8, where the muscle-specific promoter comprises a MHCK7 promoter, a CK8 promoter, or a Spc512 promoter.

10. The system of claim 2, wherein the single vector encodes the first gRNA molecule, the second gRNA molecule, and the Cas9 protein.

11. The system of any one of claims 1-10, wherein the vector comprises at least one bidirectional promoter.

12. The system of claim 11, wherein the bidirectional promoter comprises:

a first promoter driving expression of the first gRNA molecule and/or the second gRNA molecule; and

a second promoter driving expression of the Cas9 protein.

13. The system of any one of claims 1-12, wherein the first gRNA targets the polynucleotide of SEQ ID NO:2 or a 5' truncation thereof.

14. The system of any one of claims 1-13, wherein the second gRNA targets the polynucleotide of SEQ ID NO:3 or a 5' truncation thereof.

15. The system of any one of claims 1-14, wherein the Cas9 protein is SpCas9, SaCas9, or St1Cas9 protein.

16. The system of any one of claims 1-15, wherein the Cas9 gRNA scaffold is a SaCas9 gRNA scaffold.

17. The system of claim 16, wherein the SaCas9 gRNA scaffold comprises or is encoded by the polynucleotide of SEQ ID NO:4.

18. The system of any one of claims 1-17, wherein the Cas9 protein is a SaCas9 protein encoded by the polynucleotide of SEQ ID NO:11.

19. The system of any one of claims 1-18, wherein the vector comprises at least one polynucleotide selected from SEQ ID NOs: 1-13 and 24.

20. The system of any one of claims 1-19, wherein the vector comprises the polynucleotide sequence of SEQ ID NO: 24.

21. The system of any one of claims 1-20, wherein the vector comprises a polynucleotide sequence that is selected from SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID

NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 29, and SEQ ID NO: 30.

22. The system of any one of claims 1-21, wherein the vector is a viral vector.

23. The system of any one of claims 1-22, wherein the vector is an Adeno-associated virus (AAV) vector.

24. The system of claim 23, wherein the AAV vector is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV-10, AAV-11, AAV-12, AAV-13, or AAVrh.74.

25. The system of any one of claims 1-24, wherein the vector comprises a ubiquitous promoter or a tissue-specific promoter operably linked to the polynucleotide sequence encoding the first gRNA molecule, the second gRNA molecule, and/or the Cas9 protein.

26. The system of claim 25, wherein the tissue-specific promoter is a muscle specific promoter.

27. A cell comprising the system of any one of claims 1-26.

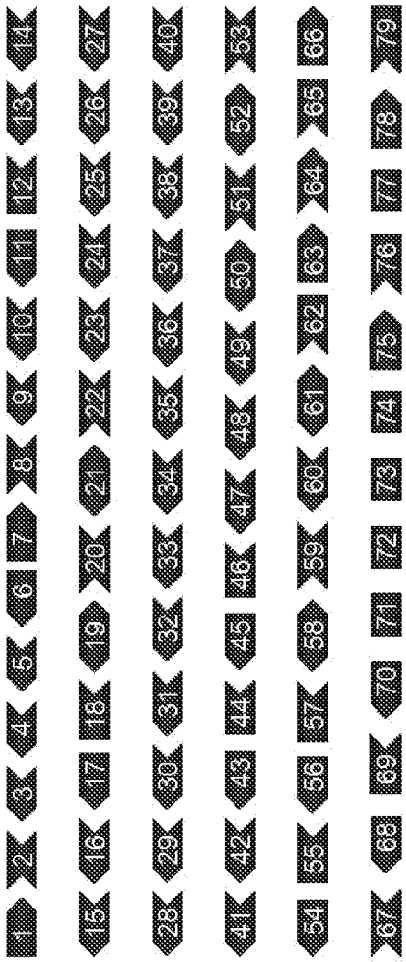
28. A kit comprising the system of any one of claims 1-26.

29. A method of correcting a mutant dystrophin gene in a cell, the method comprising administering to a cell the system of any one of claims 1-26.

30. A method of genome editing a mutant dystrophin gene in a subject, the method comprising administering to the subject the system of any one of claims 1-26 or the cell of claim 27.

31. A method of treating a subject having a mutant dystrophin gene, the method comprising administering to the subject the system of any one of claims 1-26 or the cell of claim 27.

32. The method of claim 30 or 31, wherein the system or the cell is administered to the subject intramuscularly, intravenously, or a combination thereof.



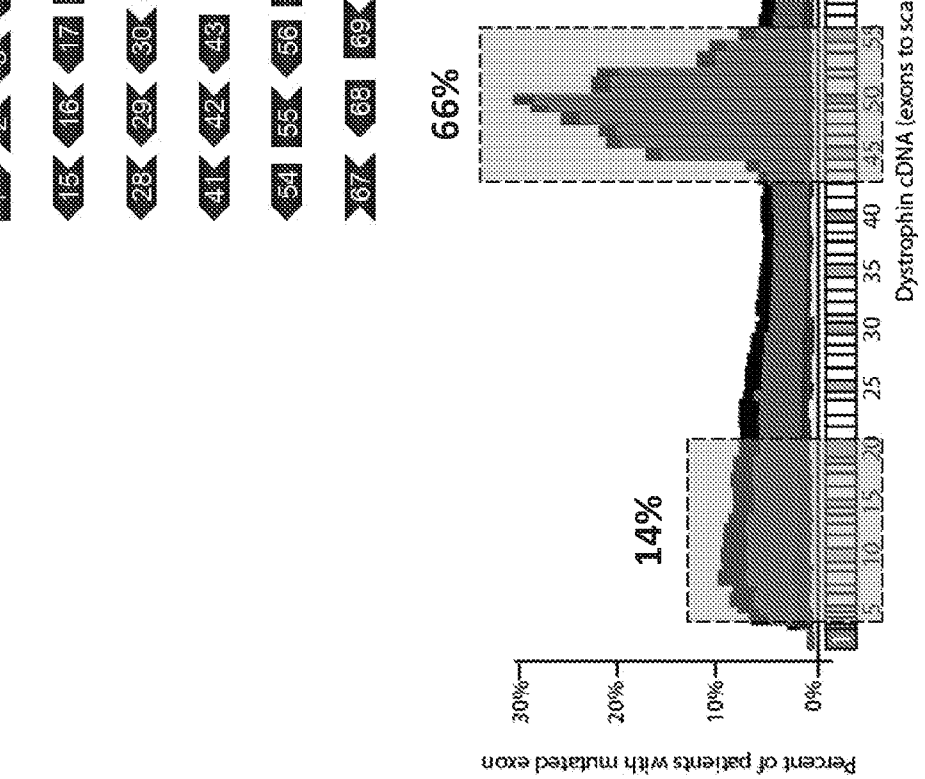

66%
14%
Total
Deletions
Duplications
Indel
1kb
Dystrophin cDNA (exons to scale)
Percent of patients with mutated exon
30%
20%
10%
0%

FIG. 1

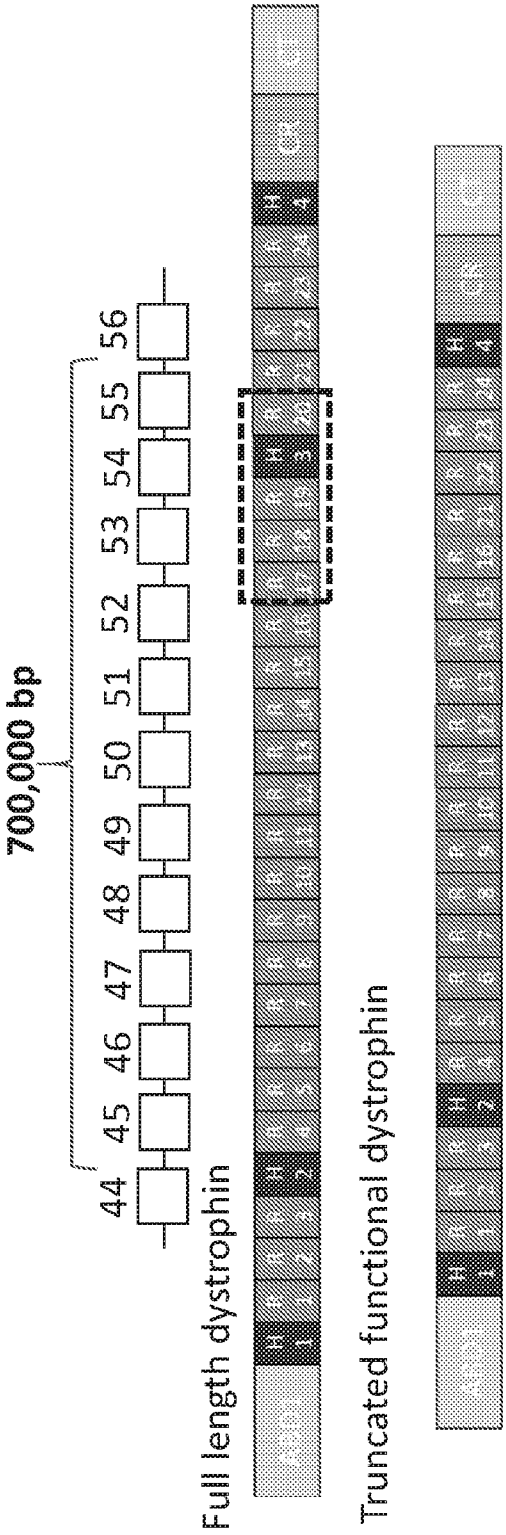

700,000 bp
44
45
46
47
48
49
50
51
52
53
54
55
56
Full length dystrophin
Truncated functional dystrophin

FIG. 2

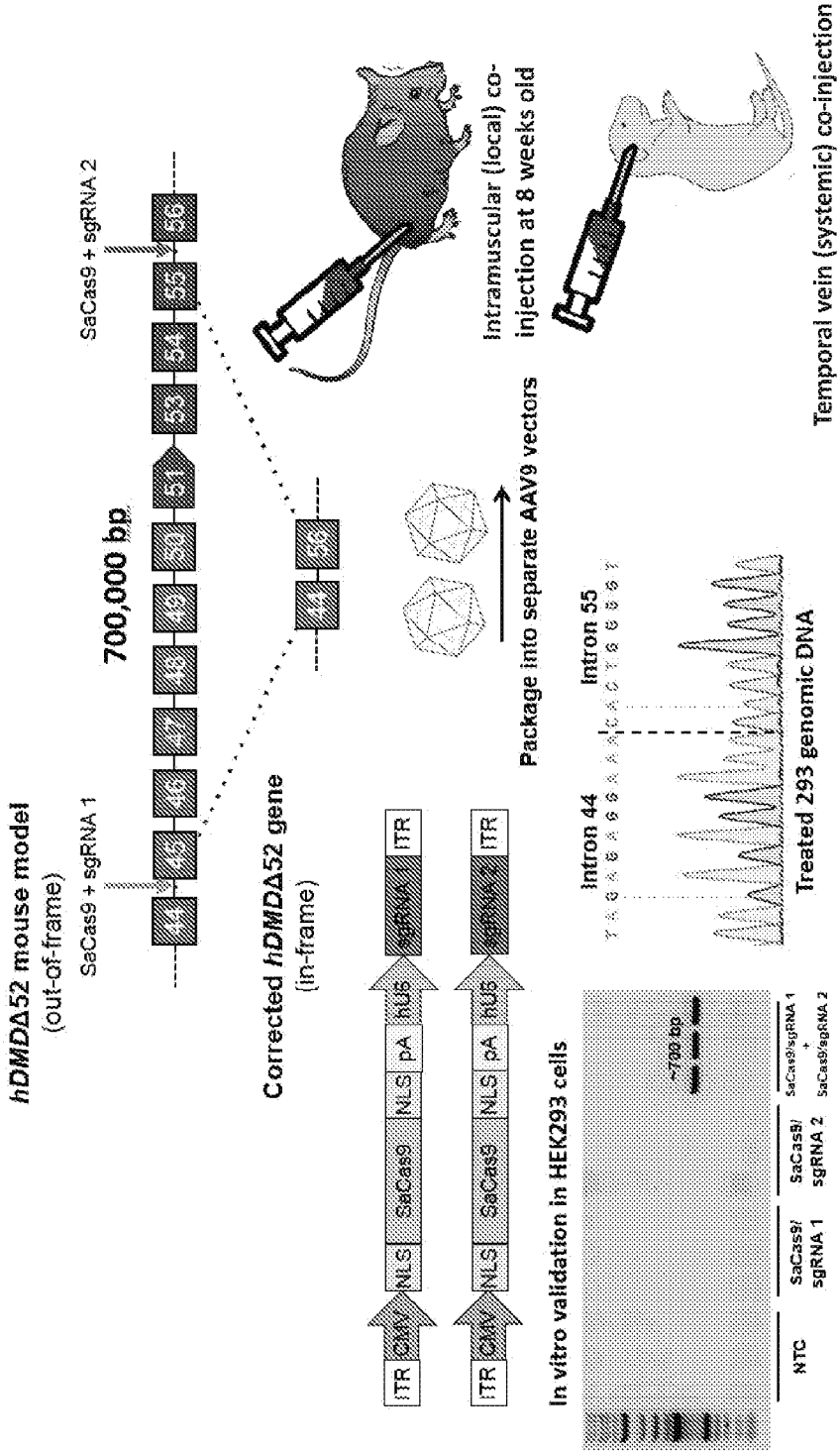

hDMDΔ52 mouse model
(out-of-frame)
SaCas9 + sgRNA 1
SaCas9 + sgRNA 2
700,000 bp
Corrected hDMDΔ52 gene
(in-frame)
Package into separate AAV9 vectors
Intramuscular (local) co-injection at 8 weeks old
Temporal vein (systemic) co-injection at 2 days postnatal
ITR
CMV
NLS
SaCas9
pA
hU6
sgRNA 1
sgRNA 2
In vitro validation in HEK293 cells
Intron 44
Intron 55
Treated 293 genomic DNA
~700 bp
NTC
SaCas9/ sgRNA 1
SaCas9/ sgRNA 2
SaCas9/sgRNA 1 + SaCas9/sgRNA 2

FIG. 3

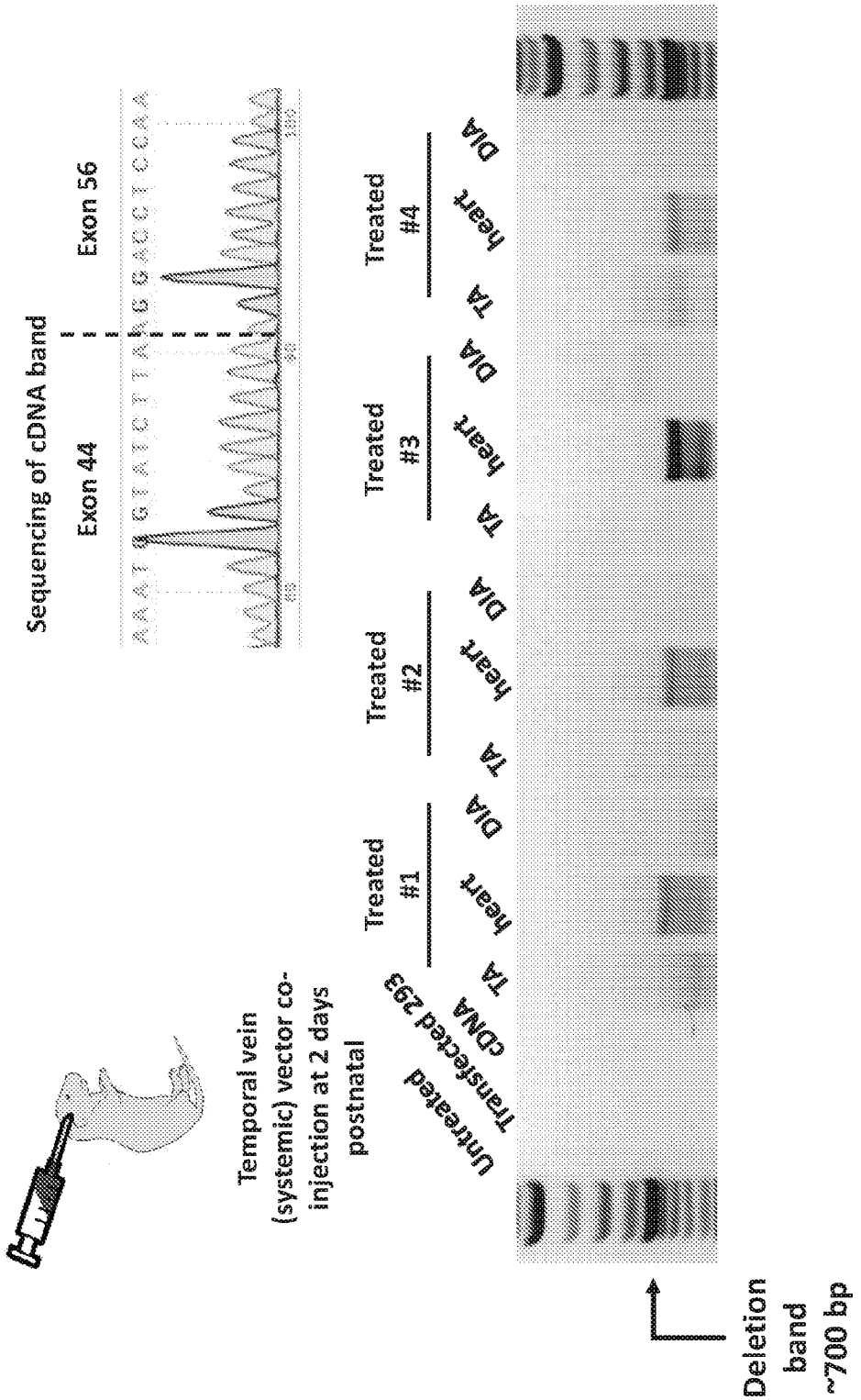

Sequencing of cDNA band
Exon 44
Exon 56
AAAT G TATCTTAT G GACCTCCAA
Temporal vein (systemic) vector co-injection at 2 days postnatal
Untreated
Transfected 293 cDNA
TA
Heart
DIA
Treated #1
TA
Heart
DIA
Treated #2
TA
Heart
DIA
Treated #3
TA
Heart
DIA
Treated #4
Deletion band ~700 bp

FIG. 4

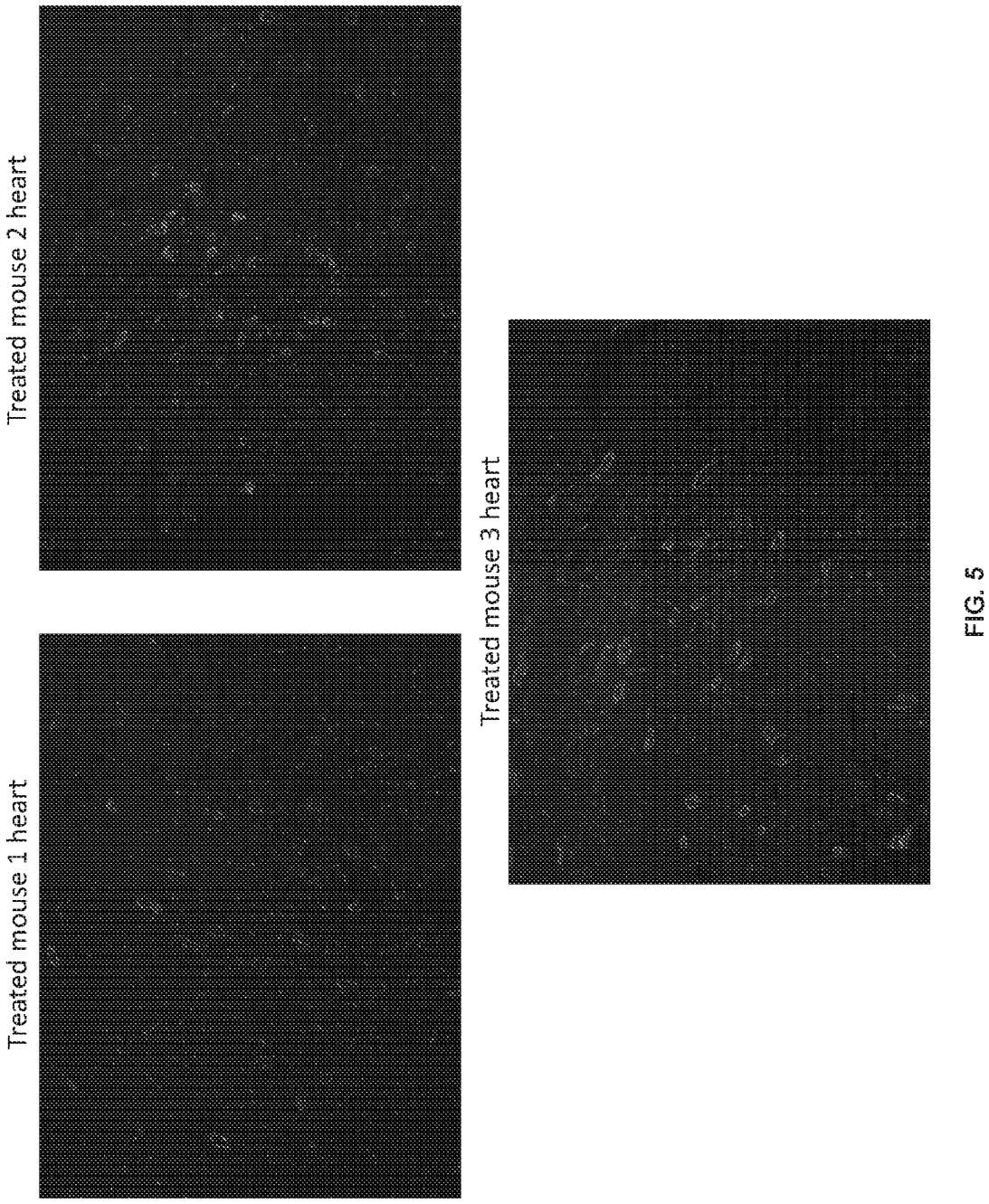

Treated mouse 2 heart
Treated mouse 1 heart
Treated mouse 3 heart

FIG. 5

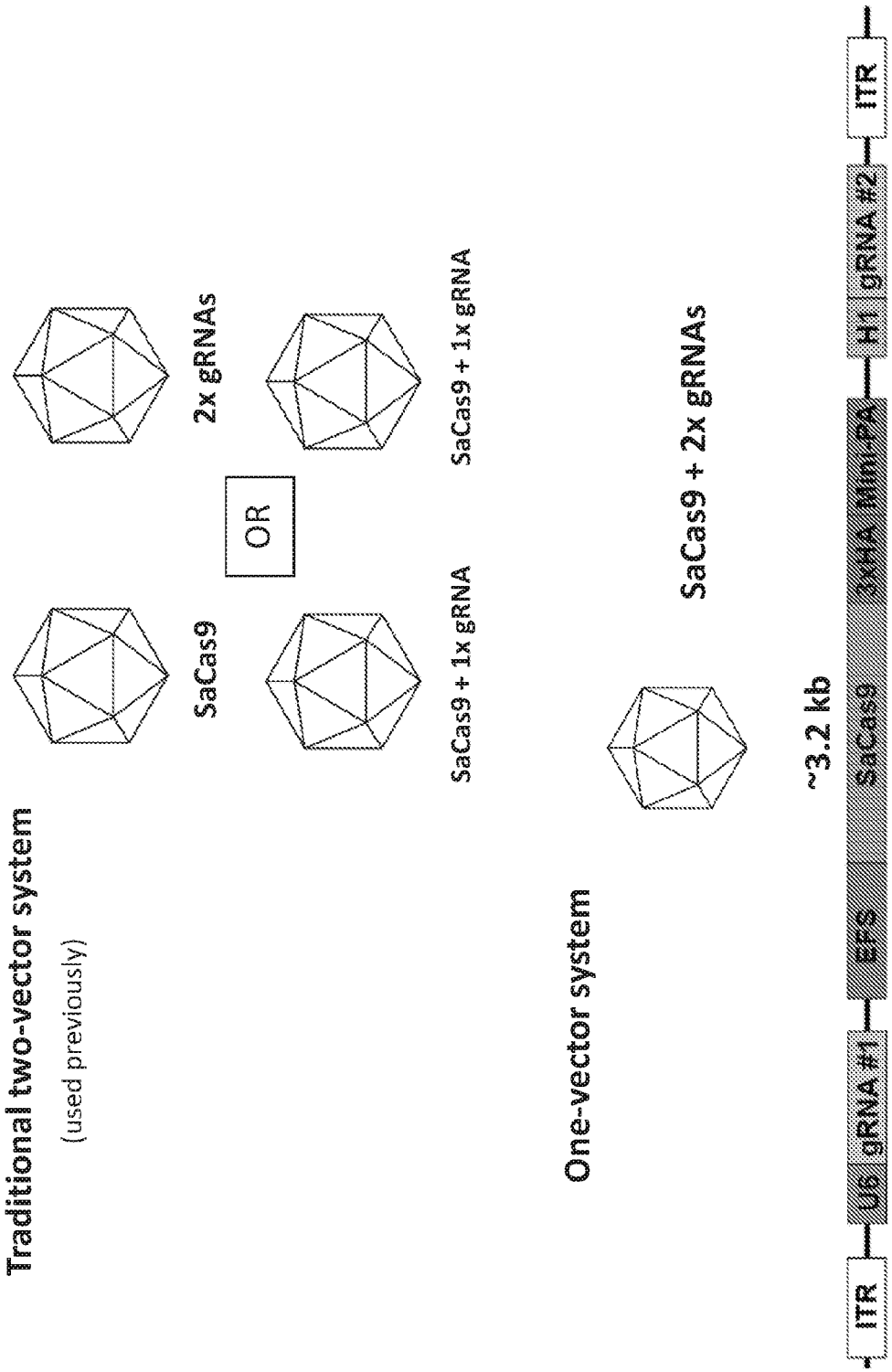

Traditional two-vector system
(used previously)
SaCas9
2x gRNAs
OR
SaCas9 + 1x gRNA
SaCas9 + 1x gRNA
One-vector system
SaCas9 + 2x gRNAs
~3.2 kb
ITR
U6
gRNA #1
EFS
SaCas9
3xHA
Mini-PA
H1
gRNA #2
ITR

FIG. 6

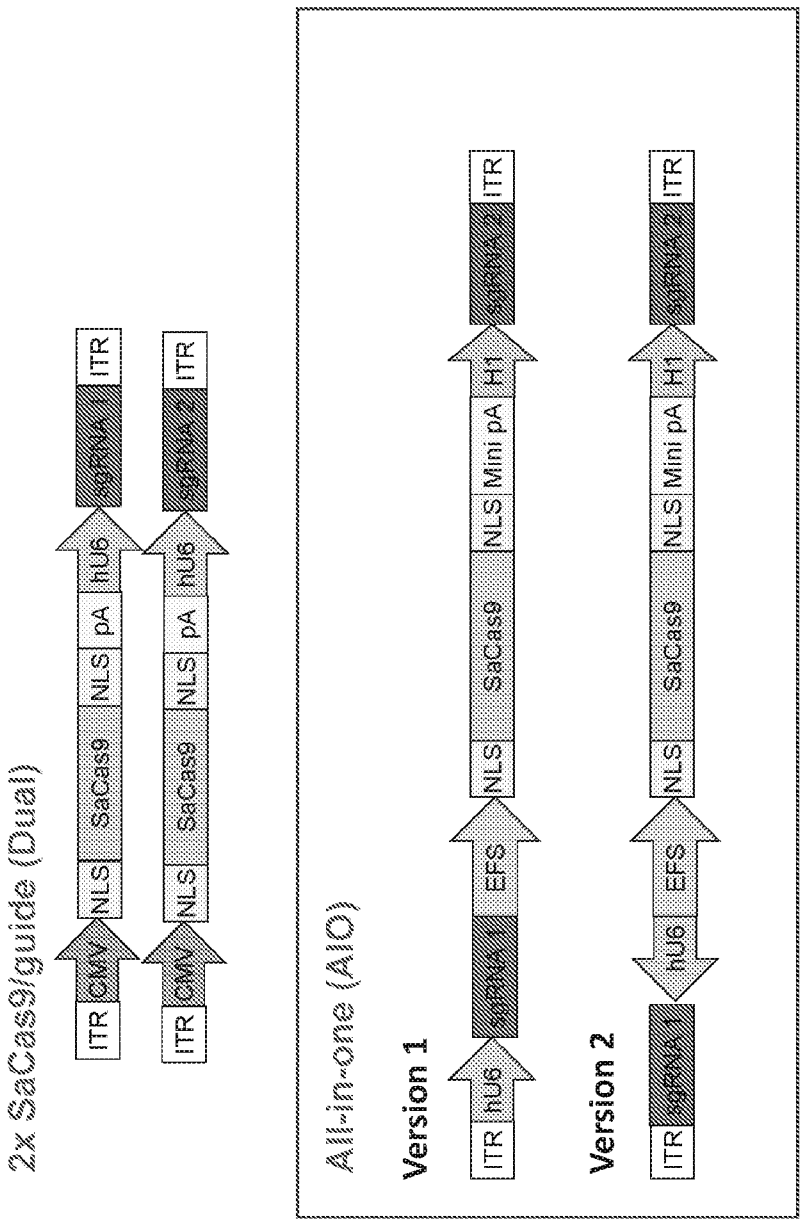

2x SaCas9/guide (Dual)
ITR
CMV
NLS
SaCas9
NLS
pA
hU6
gRNA 1
ITR
ITR
CMV
NLS
SaCas9
NLS
pA
hU6
gRNA 2
ITR
All-in-one (AIO)
Version 1
ITR
hU6
gRNA 1
EFS
NLS
SaCas9
NLS
Mini pA
H1
gRNA 2
ITR
Version 2
ITR
gRNA 1
hU6
EFS
NLS
SaCas9
NLS
Mini pA
H1
gRNA 2
ITR

FIG. 7

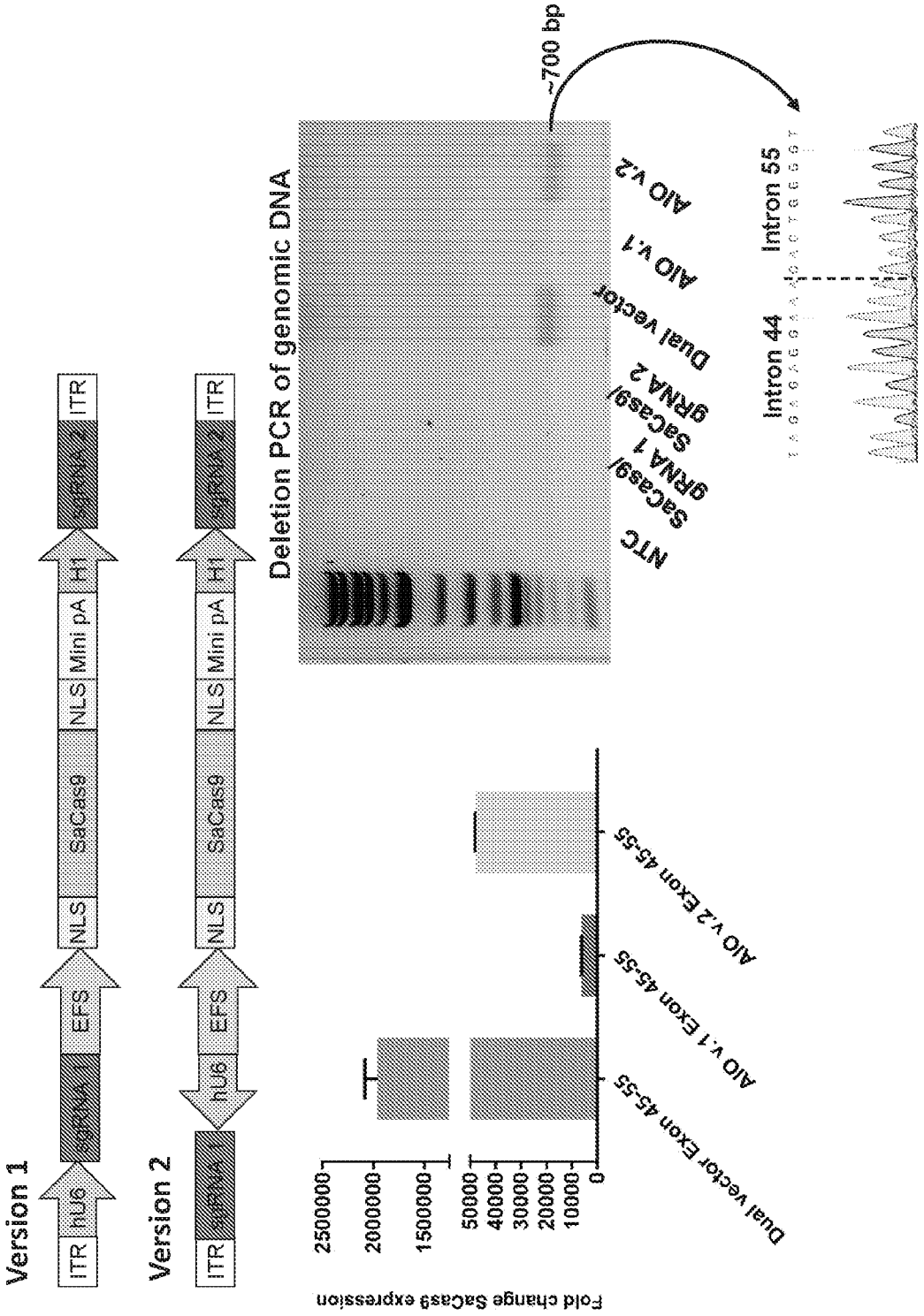

Version 1
ITR
hU6
gRNA 1
EFS
NLS
SaCas9
NLS
Mini pA
H1
gRNA 2
ITR
Version 2
ITR
gRNA 1
hU6
EFS
NLS
SaCas9
NLS
Mini pA
H1
gRNA 2
ITR
Deletion PCR of genomic DNA
~700 bp
NTC
SaCas9/ gRNA 1
SaCas9/ gRNA 2
Dual vector
AIO V.1
AIO V.2
Intron 44
Intron 55
Fold change SaCas9 expression
2500000
2000000
1500000
50000
40000
30000
20000
10000
0
Dual Vector Exon 45-55
AIO V.1 Exon 45-55
AIO V.2 Exon 45-55

FIG. 8

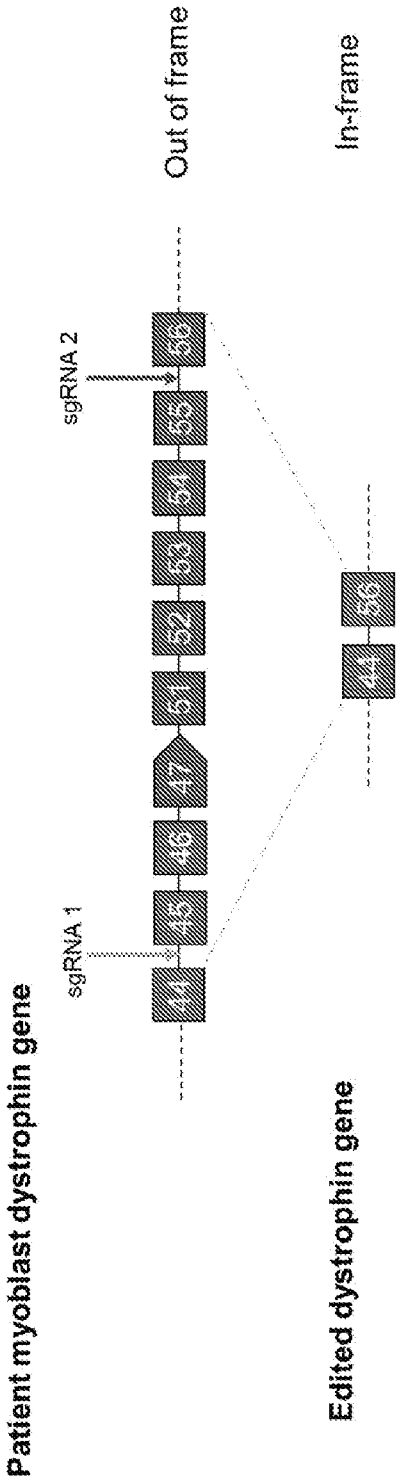

Out of frame
In-frame
sgRNA 2
sgRNA 1
Patient myobiast dystrophin gene
Edited dystrophin gene
44
45
46
47
51
52
53
54
55
56
44
56

FIG. 9A

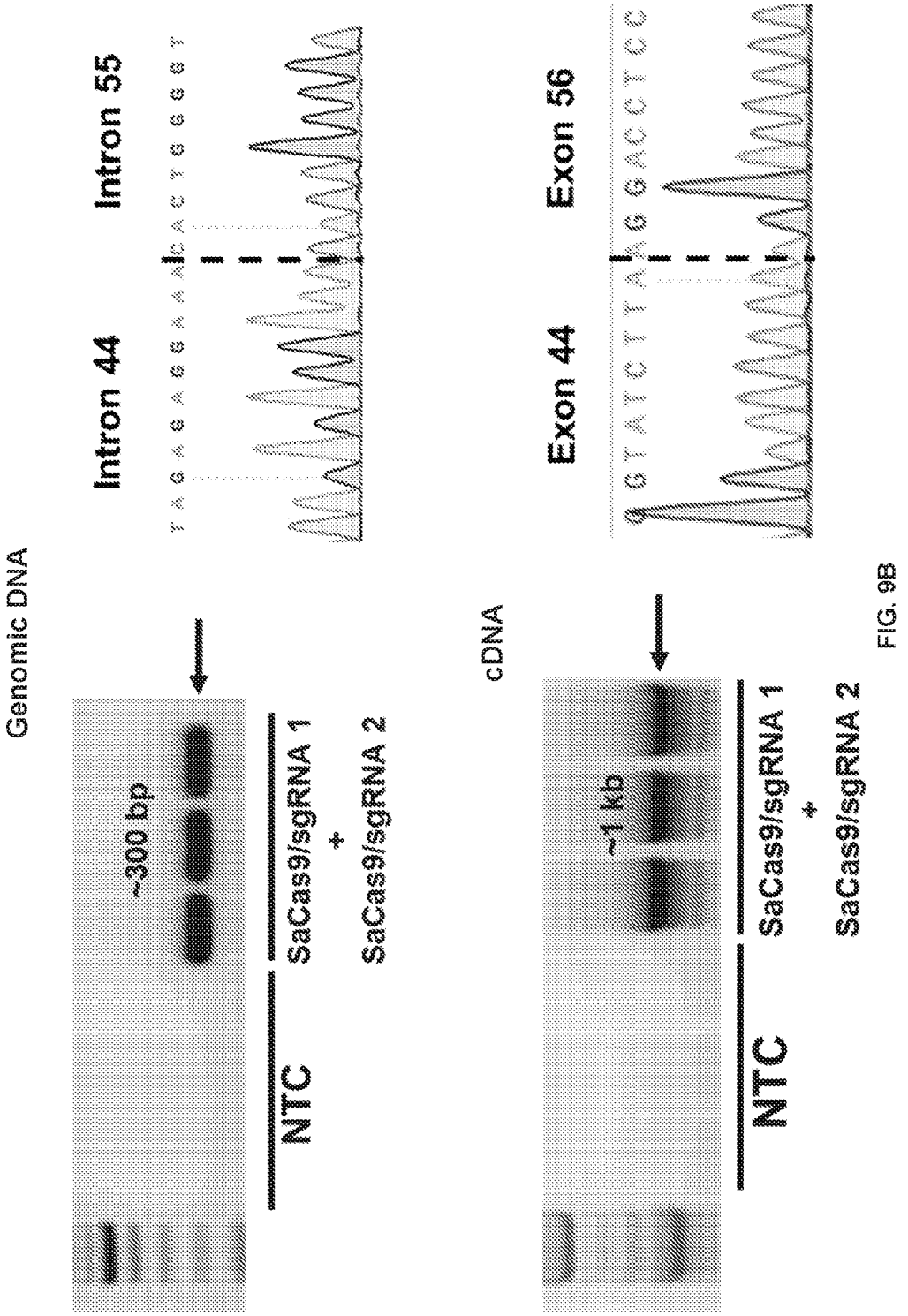

Genomic DNA
Intron 44
Intron 55
TAGAGAGGAAAACACTGGGGT
~300 bp
NTC
SaCas9/sgRNA 1
+
SaCas9/sgRNA 2
cDNA
Exon 44
Exon 56
GTATCTTACGGACCTCC
~1 kb
NTC
SaCas9/sgRNA 1
+
SaCas9/sgRNA 2

FIG. 9B

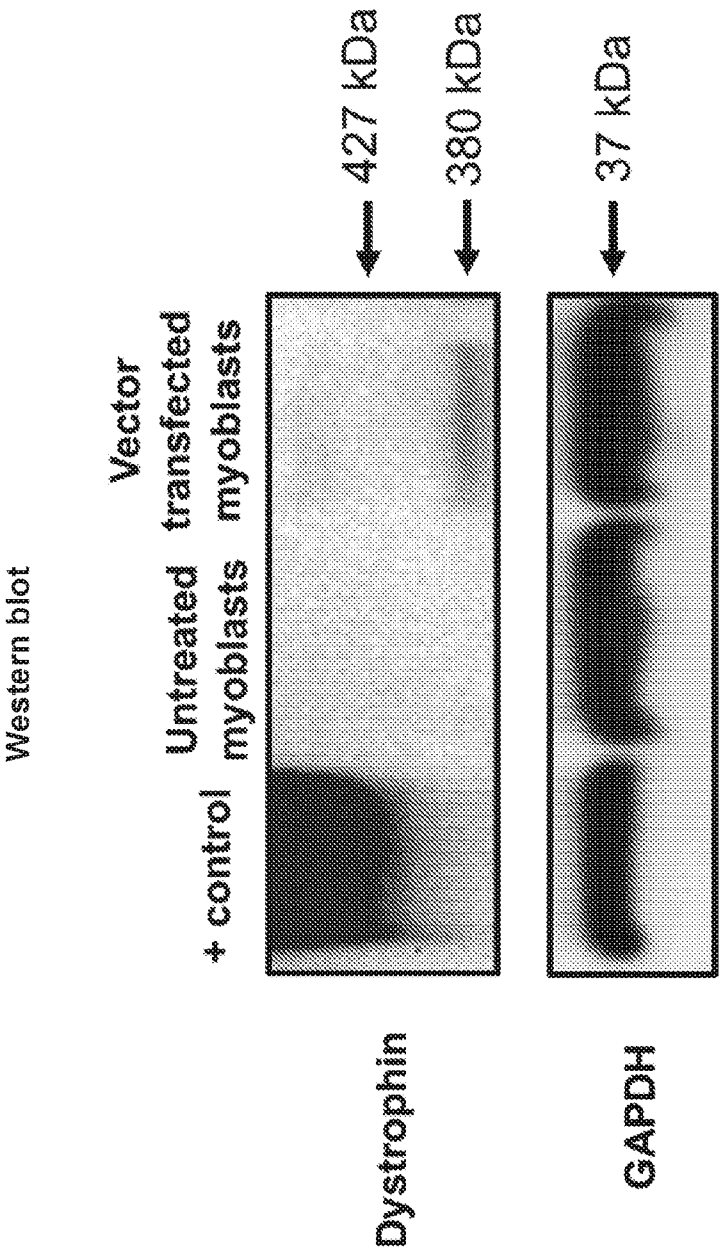

Western blot
+ control
Untreated myoblasts
Vector transfected myoblasts
427 kDa
380 kDa
37 kDa
Dystrophin
GAPDH

FIG. 9C

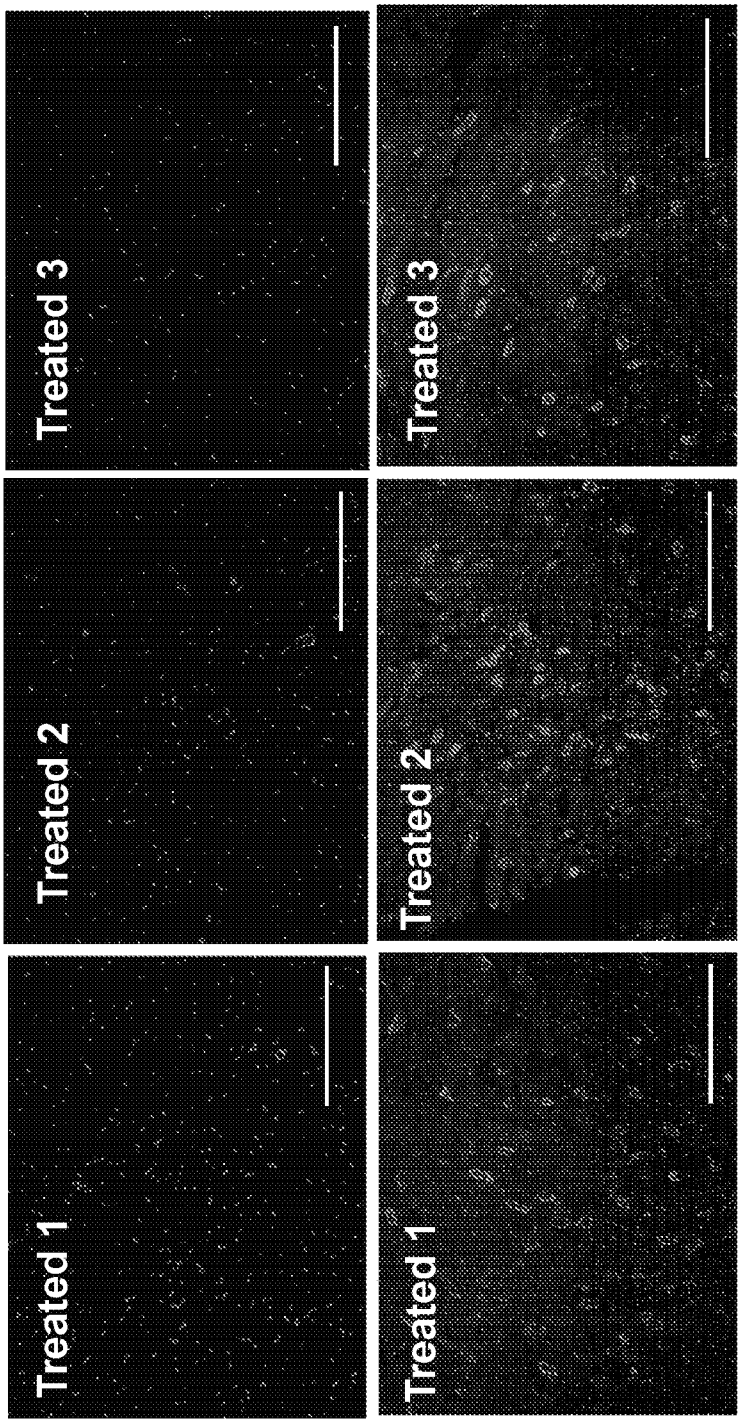

Treated 3
Treated 2
Treated 1
Treated 3
Treated 2
Treated 1
AAV9-CRISPR ROSA26
AAV9-CRISPR Exon 45-55
Dystrophin DAPI

FIG. 10

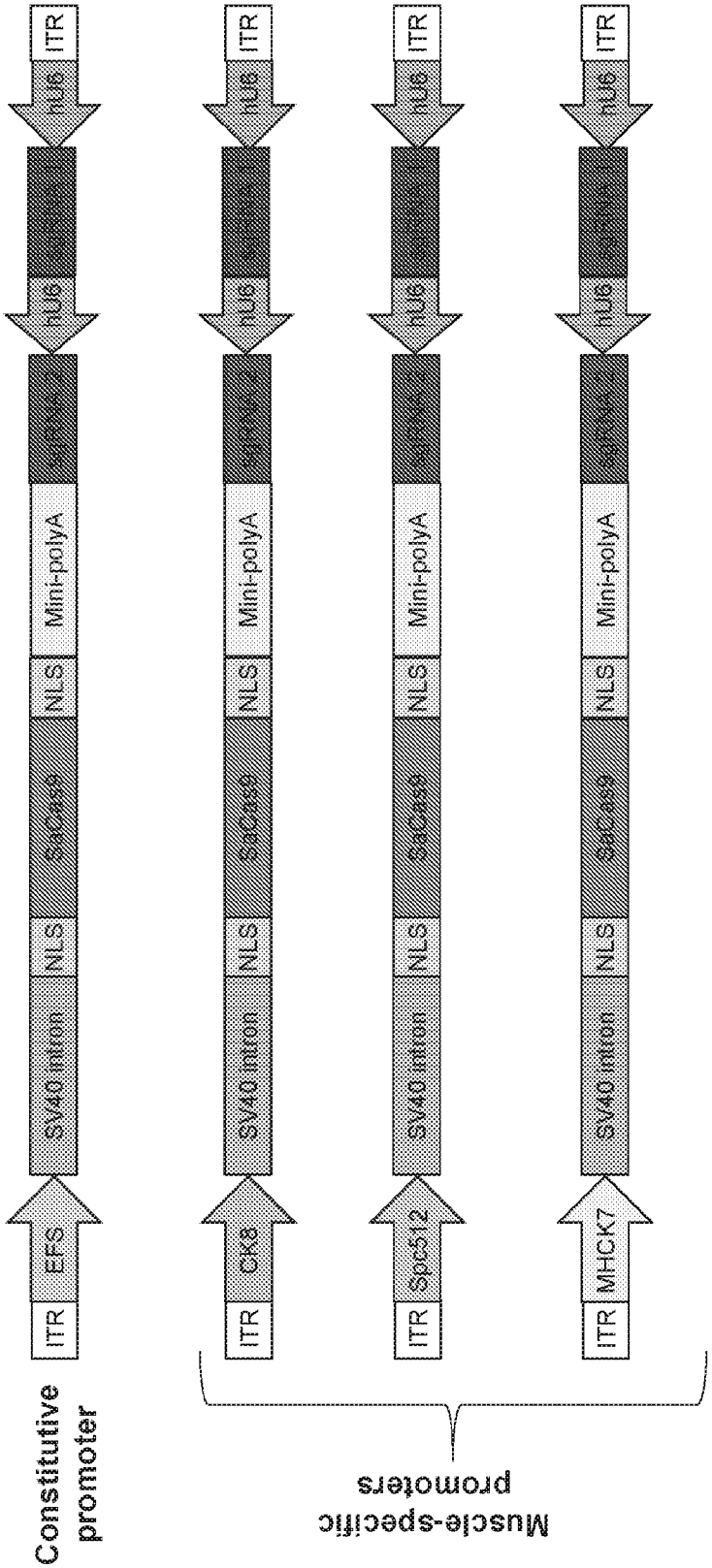

ITR
hU6
hU6
Mini-polyA
NLS
SaCas9
NLS
SV40 intron
EFS
ITR
Constitutive promoter
CK8
Spc512
MHCK7
Muscle-specific promoters

FIG. 11

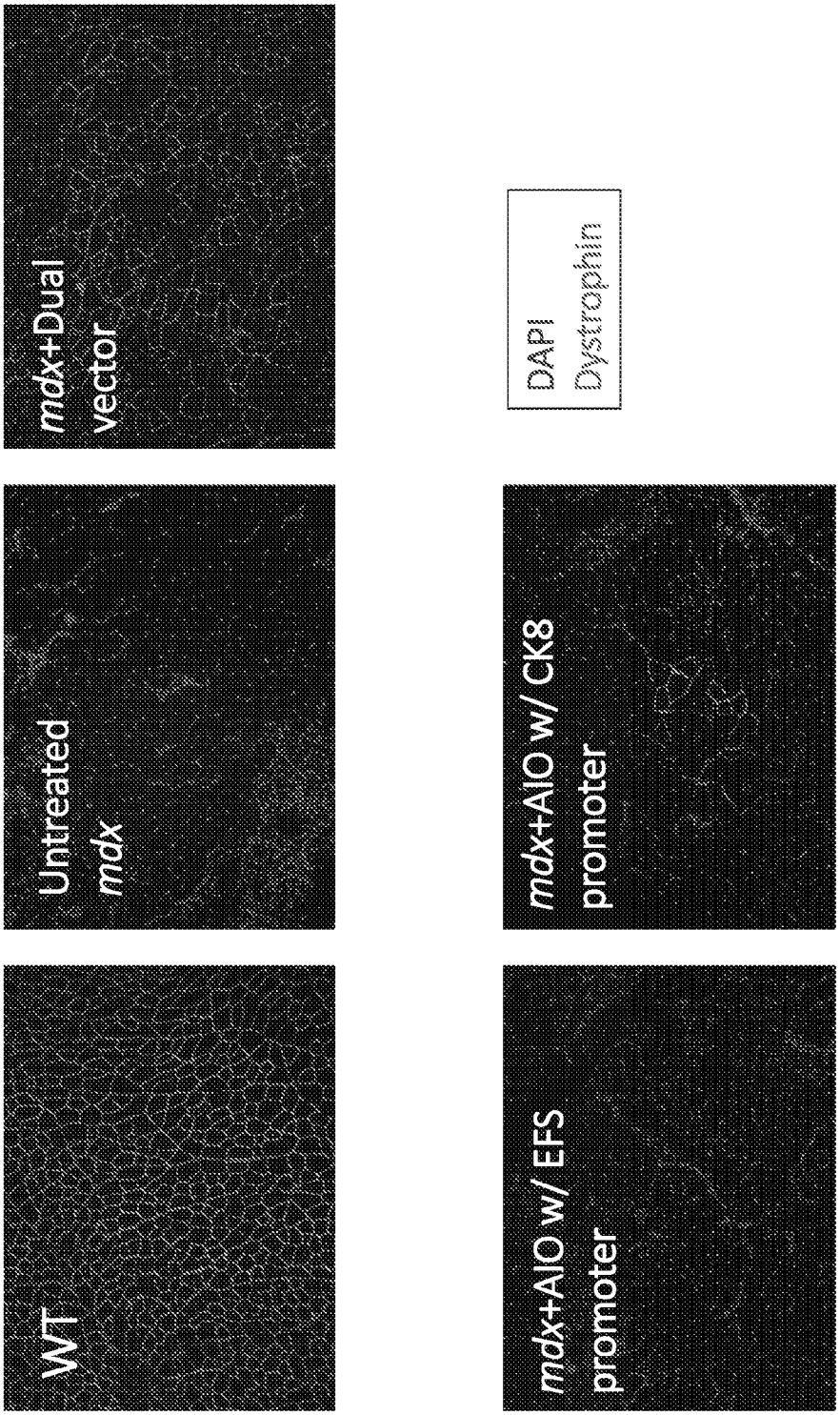

WT
Untreated mdx
mdx+Dual vector
mdx+AIO w/ EFS promoter
mdx+AIO w/ CK8 promoter
DAPI
Dystrophin

FIG. 12

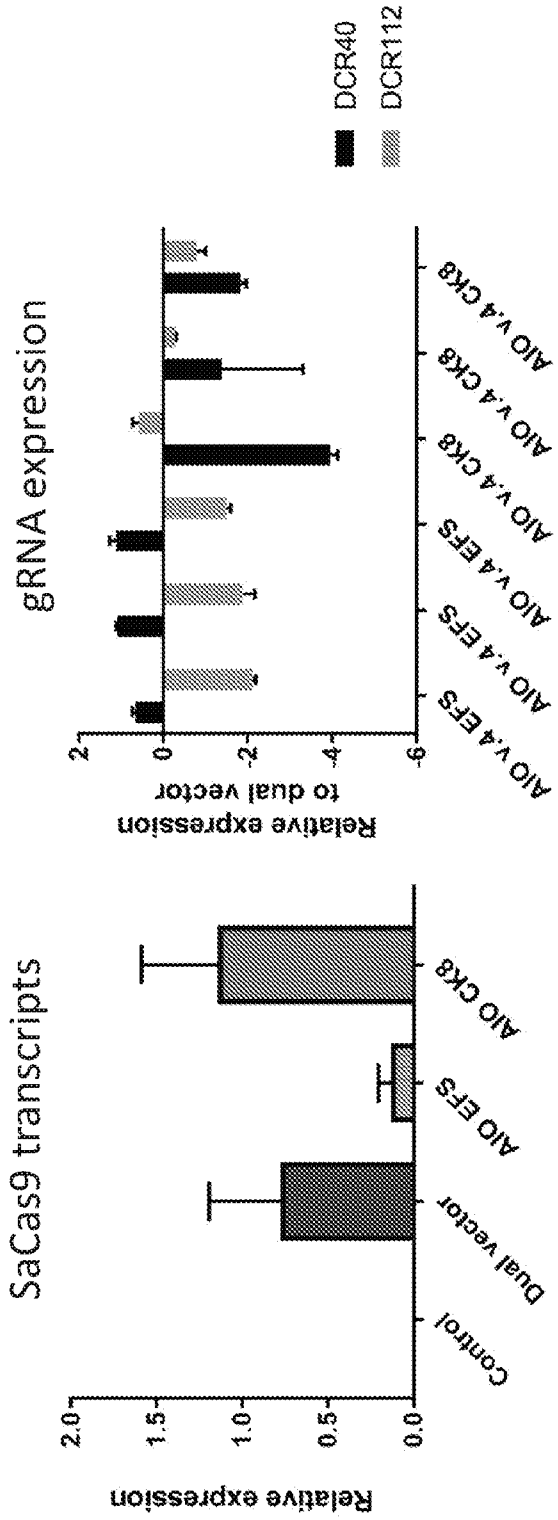

DCR40
DCR112
gRNA expression
Relative expression
to dual vector
2
0
-2
-4
-6
AIO V.4 EFS
AIO V.4 EFS
AIO V.4 EFS
AIO V.4 CK8
AIO V.4 CK8
AIO V.4 CK8
SaCas9 transcripts
Relative expression
2.0
1.5
1.0
0.5
0.0
Control
Dual vector
AIO EFS
AIO CK8

FIG. 13

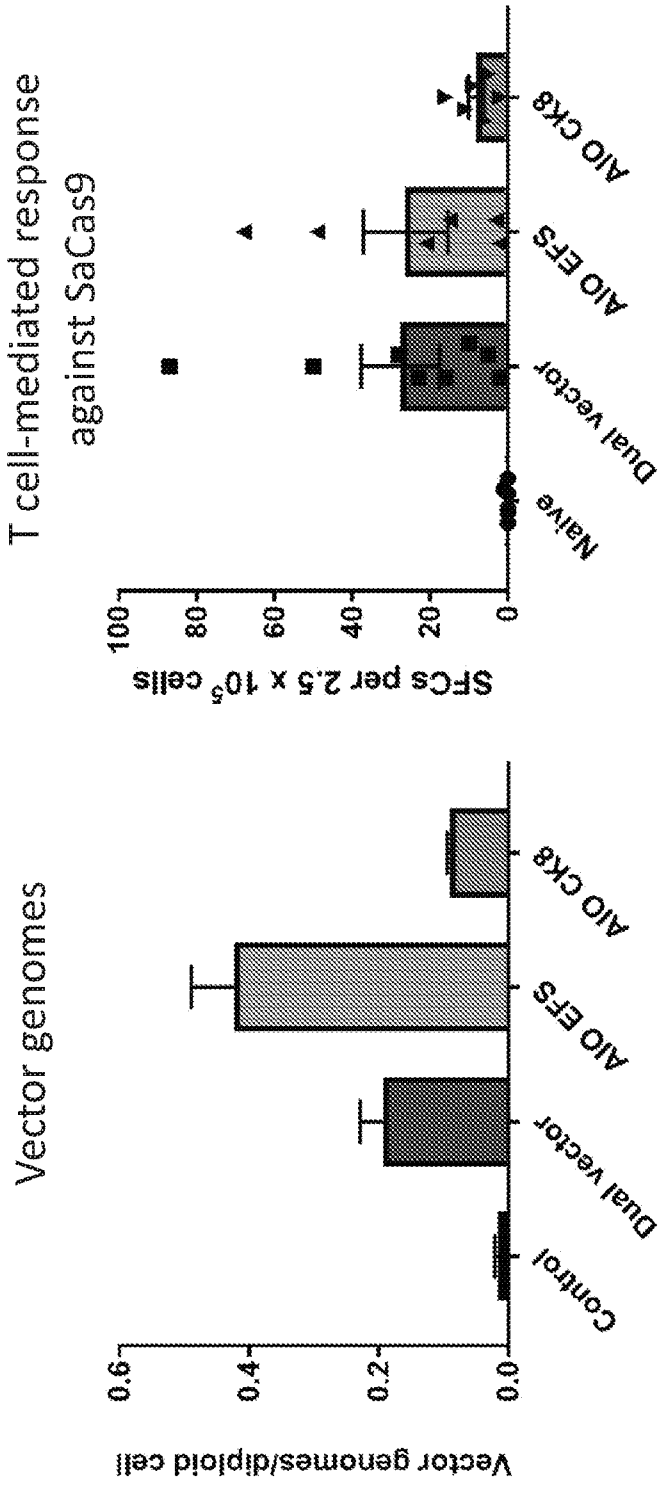

T cell-mediated response against SaCas9
SFCs per 2.5 x 10^5 cells
100
80
60
40
20
0
Naive
Dual vector
AIO EFS
AIO CK8
Vector genomes
Vector genomes/diploid cell
0.6
0.4
0.2
0.0
Control
Dual vector
AIO EFS
AIO CK8

FIG. 14

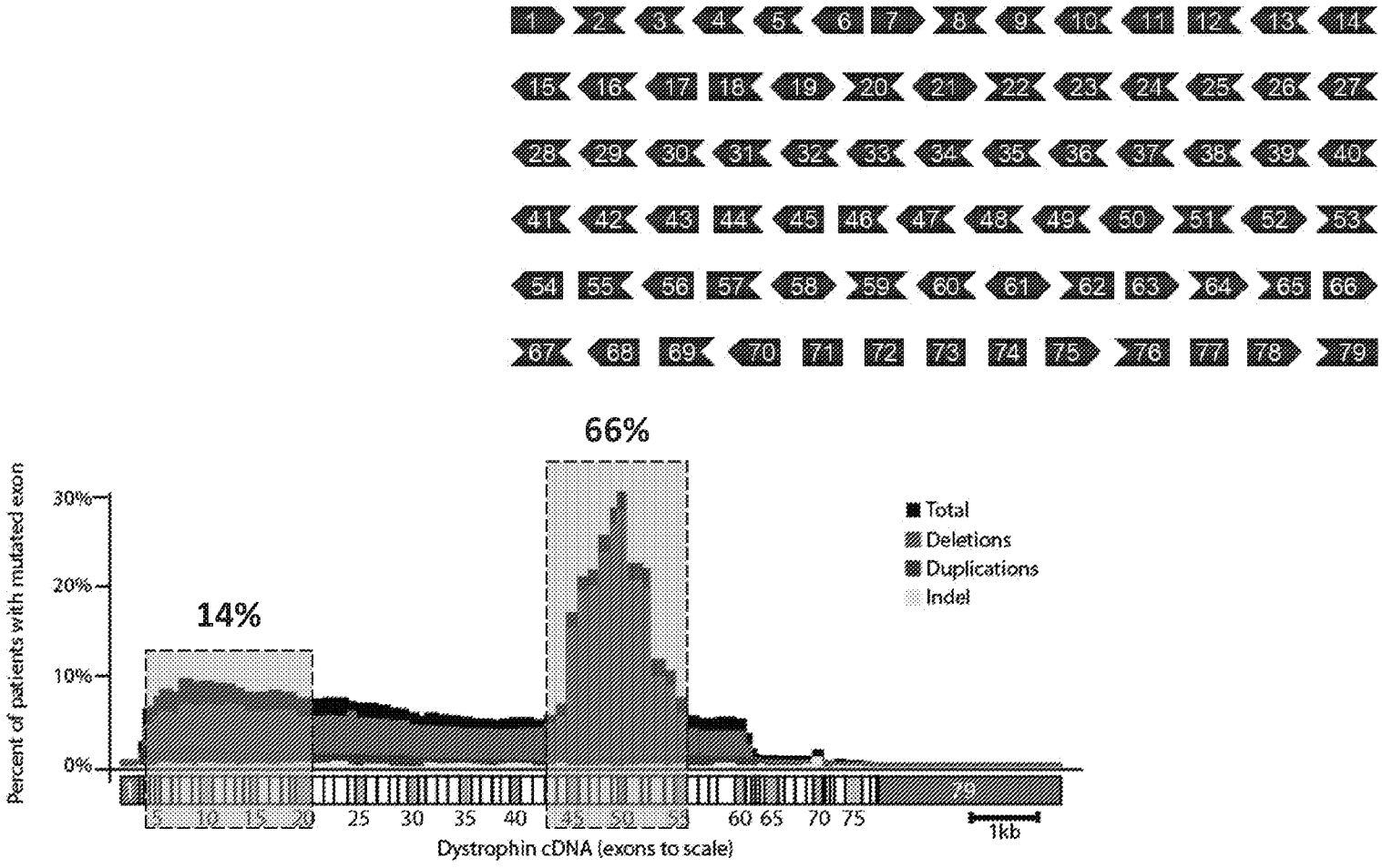

1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 21 22 23 24 25 26 27
28 29 30 31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50 51 52 53
54 55 56 57 58 59 60 61 62 63 64 65 66
67 68 69 70 71 72 73 74 75 76 77 78 79
66%
14%
30%
20%
10%
0%
Percent of patients with mutated exon
Total
Deletions
Duplications
Indel
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 79
1kb
Dystrophin cDNA (exons to scale)

FIG. 1