



(51) International Patent Classification:

C12N 9/22 (2006.01)

C12N 15/86 (2006.01)

C12N 9/78 (2006.01)

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.

(21) International Application Number:

PCT/US2021/017868

(22) International Filing Date:

12 February 2021 (12.02.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/976,590

14 February 2020 (14.02.2020) US

Published:

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

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

(71) Applicant: **OHIO STATE INNOVATION FOUNDATION** [US/US]; 1524 North High Street, Columbus, Ohio 43201 (US).

(72) Inventors: **HAN, Renzhi**; c/o Ohio State Innovation Foundation, 1524 North High Street, Columbus, Ohio 43201 (US). **XU, Li**; c/o Ohio State Innovation Foundation, 1524 North High Street, Columbus, Ohio 43201 (US). **GAO, Yandi**; c/o Ohio State Innovation Foundation, 1524 North High Street, Columbus, Ohio 43201 (US).

(74) Agent: **PRATHER, Donald M.** et al.; MEUNIER CARLIN & CURFMAN LLC, 999 Peachtree St. NE, Suite 1300, Atlanta, Georgia 30309 (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, IT, 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,

(54) Title: NUCLEOBASE EDITORS AND METHODS OF USE THEREOF



FIG. 6A

(57) Abstract: The present disclosure relates to nucleobase editors and methods of use thereof. Disclosed herein are fusion proteins, systems, and compositions for editing disease-associated mutations and methods of use thereof. In some aspects, disclosed herein is a fusion protein comprising a Cas9 nickase and a nucleotide deaminase, wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.



NUCLEOBASE EDITORS AND METHODS OF USE THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 62/976,590, filed February 14, 2020, which is expressly incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

This invention was made with government support under Grant No. R01 HL116546 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD

The present disclosure relates to nucleobase editors and methods of use thereof.

BACKGROUND

Many genetic diseases arise from single nucleotide alterations. Duchenne muscular dystrophy (DMD) is a fatal genetic muscle disease affecting approximately 1 in ~5000 male births worldwide, which is caused by mutations in the *DMD* gene. Most of the *DMD* mutations are due to deletions or duplications with over 500 point mutations accounting for ~10% of the cases. DMD codes for the dystrophin protein, a cytoskeletal protein that functions in the muscle force transmission and sarcolemmal stability of muscle fibers. Loss of dystrophin leads to progressive muscle weakness and wasting, loss of ambulation, respiratory impairment, cardiomyopathy, and eventual death. Previous studies showed that exon deletion through CRISPR genome editing can restore dystrophin expression and function. Although promising, this strategy has safety concerns as it relies on repair of the double strand DNA break (DSB) created by CRISPR/Cas9, which can cause unwanted large deletion and even chromosomal rearrangement.

Therefore, what is needed are compositions for precise correction of disease-associated mutations.

SUMMARY

Disclosed herein are fusion proteins, systems, and compositions for editing disease-associated mutations and methods of use thereof.

In some aspects, disclosed herein is a fusion protein comprising a Cas9 nickase and a nucleotide deaminase, wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when

compared to SEQ ID NO: 11, and wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.

In some embodiments, the first amino acid substitution is selected from the group consisting of A262T, R324L, S409I, E480K, E543D, M694I, and E1219V when compared to SEQ ID NO: 11. In some embodiments, the second amino acid substitution is selected from the group consisting of L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, R1335Q, R1335E, and T1337R when compared to SEQ ID NO: 11.

In some embodiments, the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

In some embodiments, the Cas9 nickase when in conjunction with a bound guide RNA (gRNA) specifically binds to a target nucleic acid sequence.

In some embodiments, the Cas9 nickase recognizes a NG protospacer adjacent motif (PAM) sequence.

In some embodiments, the nucleotide deaminase is a cytidine deaminase or an adenine deaminase. In some embodiments, the adenine deaminase is a dimeric adenine deaminase or a monomeric adenine deaminase domain. In some embodiments, the monomeric adenine deaminase comprises amino acid substitutions A56G and V82G when compared to SEQ ID NO: 47. In some embodiments, the adenine deaminase comprises the amino acid sequence set forth in SEQ ID NO: 22 or 23.

In some aspects, disclosed herein is a system for base editing comprising:

a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N; and

a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

wherein the first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.

In some embodiments, the N-terminal portion of the Cas9 nickase comprises the amino acid positions 2-573 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20 and the C-terminal portion of the Cas9 nickase comprises the amino acid positions 574-1368 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

5 In some embodiments, the first nucleotide sequence or the second nucleotide sequence further comprises a nucleotide encoding a guide RNA (gRNA).

In some aspects, disclosed herein is a composition comprising:

a first recombinant adeno-associated virus (AAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N;

10 and

a second recombinant AAV particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

wherein the first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

15 wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to
20 SEQ ID NO: 11.

In some embodiments, the first nucleotide sequence comprises the sequence as set forth in SEQ ID NO: 3 or SEQ ID NO: 5. In some embodiments, the second nucleotide sequence comprises the sequence as set forth in SEQ ID NO: 4 or SEQ ID NO: 6.

25 In some aspects, disclosed herein is a method of treating a genetic disease or disorder in a subject, comprising administering a therapeutically effective amount of the composition of any preceding aspect. In some embodiments, the genetic disease or disorder is Duchenne muscular dystrophy, dysferlinopathy, or spinal muscular atrophy. In some embodiments, the genetic disease or disorder is due to a point mutation in a causative gene from a G:C pair to an A:T pair.

30 BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying figures, which are incorporated in and constitute a part of this specification, illustrate several aspects described below.

Figs. 1a-1e show *in vitro* studies of *mdx*^{Δcv} mutation correction using ABE-NG. Fig. 1a shows genomic DNA, encoded amino acids and guide RNA with PAM (highlighted in blue)

sequences at the stop codon mutation site (red). Fig. 1b shows the reporter construct containing a puromycin resistance cassette fused with E2A peptide, *mdx*^{4cv} target sequence and ATG-removed EGFP. Correction of the stop codon within the target sequence would allow EGFP expression. Fig. 1c shows fluorescence microscopy images of HEK293 cells transfected with reporter alone, or reporter, gRNA and one of the base editors (ABEmax, ABE-x and ABE-NG). Scale bar: 500 μ m. Figs. 1d-1e shows flow cytometry analysis of EGFP expression in HEK293 cells transfected as described in Fig. 1c. * $p < 0.05$; **** $p < 0.0001$ (one-way ANOVA). The sequences in Fig. 1 are GAACAGCTGCAGAACAGGAGATAACAG (SEQ ID NO: 620), GTTATCTCCTGTTCTG CAGCTGT (SEQ ID NO: 621).

Figs. 2a-2f show optimization of the PAM-interacting domain to improve the editing efficiency of ABE-NG at the NGN sites. Figs. 2a-2f show measurements of the base editing efficiencies of different ABE variants on a NGG-PAM site and five different sites with NGH or GAT PAM. The gRNA sequences are listed on the top of each graph with the PAM sequences in red and the target adenines in blue. ns, not significant; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$ (one-way ANOVA). The sequences in Fig. 2 are ATGACAGGCAGGGGCACCGCGG (SEQ ID NO: 622), GAGCGAGCAGCGTCTTCGAGAGT (SEQ ID NO: 623), GCAGACGGCA GTCACTAGGGGGC (SEQ ID NO: 624), GTCGCAGGACAGCTTTTCCTAGA (SEQ ID NO: 625), GGAAGCTGGGTGAATGGAGCGA (SEQ ID NO: 626), GATCCAGGT GCTGCAGAAGGGAT (SEQ ID NO: 627).

Figs. 3a-3d show re-engineering of the adenine deaminase domain to improve the efficiency and specificity of ABE-NG. Fig. 3a shows schematics of the adenine deaminase domain used in ABE-NG editors. Fig. 3b shows quantification of the editing efficiency of different ABE-NG variants with modified TadA* domain at the *mdx*^{4cv} target site. **** $p < 0.0001$ (one-way ANOVA). Fig. 3c shows the number of off-target RNA editing events in Neuro-2a cells transfected with different ABE-NG variants. *** $p < 0.05$; *** $p < 0.001$ (one-way ANOVA). Fig. 3d shows quantification of the off-target RNA editing (A-to-I) activities on four RNA adenines previously identified as being efficiently modified by ABEmax in HEK293 cells. **** $p < 0.0001$ (one-way ANOVA).

Figs. 4a-4b show comparison of the editing efficiency of different ABEs at the *mdx*^{4cv} target site (Fig. 4a) and human dysferlin Q605X site (Fig. 4b). All base editors carry the SpCas9-NG and differ at the adenine deaminase domain. ns, not statistically significant; ** $p < 0.01$; **** $p < 0.0001$ (one-way ANOVA for Fig. 4a; two-way ANOVA for Fig. 4b). The sequence in Fig. 4 is GTTATCTCCTGTTCTGCAGC (SEQ ID NO: 570) and ATCCTACAGCATGGTGGCTG (SEQ ID NO: 590).

Figs. 5a-5b show the intein split of ABEmax had relatively low editing activity. Fig. 5a shows schematics showing the two halves of intein-split ABEmax. The TadA-TadA* was fused with Npu intein N-terminal fragment and SpCas9 nickase (nSpCas9) was fused with Npu intein C-terminal fragment. Fig. 5b shows genomic DNA PCR analysis of HEK293 cells at 5 days after transfection with S2-gRNA and different versions of ABEs.

Figs. 6a-6h show intein-mediated assembly of full-length ABE-NG. Fig. 6a shows schematics of the intein split ABE-NG. The N-terminal and C-terminal intein sequences reconstitute the full-length protein when co-expressed within cells. Fig. 6b shows Western blot analysis of HEK293 cell lysates transfected with different versions of ABEs. Fig. 6c shows fluorescence microscopy images of HEK293 cells transfected with reporter alone, or reporter, gRNA and one of the base editors (ABE-NG, split_v1 N+C or Split_v2 N+C). Scale bar: 500 μ m. Fig. 6d shows flow cytometry analysis of EGFP expression in HEK293 cells transfected as described in Fig. 6c. Fig. 6e shows Western blot analysis of HEK293 cell lysates transfected with full-length iABE-NGA, Gp41-1 or Npu split of iABE-NGA. FL, the full-length iABE-NGA band; N, the N-terminal fragment of the iABE-NGA. Fig. 6f shows densitometry quantification of the Western blot data shown in Fig. 6e. Fig. 6g shows the assembly efficiency of the Gp41-1 and Npu split of iABE-NGA (defined as the percentage of the full-length iABE-NGA bands). Fig. 6h shows quantification of the editing efficiency of full-length iABE-NGA, Gp41-1 split and Npu split of iABE-NGA at the *mdx*^{4cv} target site. Npu Split_1g is same as Npu Split except that only the C-terminal construct carries the gRNA. ns, not significant; **p* < 0.05; ****p* < 0.001; *****p* < 0.0001 (one-way ANOVA).

Fig. 7. Immunofluorescence staining of heart and gastrocnemius muscle sections with anti-dystrophin antibody and DAPI. The *mdx*^{4cv} mice at 5 weeks of age received either 5x10¹³, 1x10¹⁴ AAV9-iNG/*mdx*^{4cv}-gRNA or 1x10¹⁴ vg/kg AAV9-iNG/non-targeting-gRNA (NT-gRNA) through tail vein injection, and sacrificed at 10 weeks of age for immunofluorescence staining analysis. Scale bar: 100 μ m.

Figs. 8a-8k show widespread restoration of dystrophin expression in *mdx*^{4cv} mice following systemic administration of AAV9-iNG. Fig. 8a shows dystrophin and laminin- α 2 co-immunostaining of heart sections from WT and *mdx*^{4cv} mice (10 weeks or 10 months of age) with or without tail vein injection of AAV9-iNG (a total of 1x10¹⁴ vg/kg, 1:1 of the N and C-terminal half). Scale bar: 100 μ m. Figs. 8b-8d show quantification of dystrophin-positive fibers in the heart (Fig. 8b), gastrocnemius (GA, Fig. 8c) and diaphragm (Diaph, Fig. 8d) muscles. 10w, 10 weeks old; 10m, 10 months old. ***p* < 0.01; *****p* < 0.0001 (Student's *t*-test). Fig. 8e shows Western blot analysis of heart homogenates from 10-week-old mice with anti-dystrophin, Cas9 and Gapdh

antibodies. The WT muscle lysates were loaded at 5 μ g/lane while the *mdx*^{4cv} muscle lysates were loaded at 25 μ g/lane. Fig. 8f shows densitometry quantification of Western blot data shown in Fig. 8e. * p < 0.05 (one-way ANOVA). Fig. 8g shows Western blot analysis of heart homogenates from 10-month-old mice with anti-dystrophin and Gapdh antibodies. Fig. 8h shows densitometry quantification of Western blot data in Fig. 8g. ** p < 0.01; **** p < 0.0001 (one-way ANOVA). Fig. 8i shows representative sequencing trace of dystrophin transcripts of WT and *mdx*^{4cv} mouse hearts (10 weeks or 10 months old) with or without AAV9-iNG treatment. Figs. 8j and 8k show quantification of the targeted T-to-C editing efficiency in the *mdx*^{4cv} mouse hearts (Fig. 8j, 10 weeks; Fig. 8k, 10 months) as assayed by sequencing of dystrophin transcripts. **** p < 0.0001 (Student's t -test).

Fig. 9 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections of a WT mouse at 10 weeks of age. Scale bars: 0.5 mm.

Fig. 10 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections of a control *mdx*^{4cv} mouse at 10 weeks of age. Scale bars: 0.5 mm.

Fig. 11 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #142 five weeks after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Fig. 12 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #147 five weeks after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Fig. 13 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #148 five weeks after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Fig. 14 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #149 five weeks after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Fig. 15 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #150 five weeks after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Figs. 16a-16b show immunofluorescence staining of dystrophin and laminin α 2 in the gastrocnemius (Fig. 16a) and diaphragm (Fig. 16b) muscles from WT and *mdx*^{4cv} (10 weeks of age) treated with or without tail vein injection of AAV9-iNG. Scale bar: 100 μ m.

Fig. 17 shows Western blot analysis of gastrocnemius muscles from WT and *mdx*^{4cv} (10 weeks of age) treated with or without tail vein injection of AAV9-iNG.

Fig. 18 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections of a WT mouse at 10 months of age. Scale bars: 0.5 mm.

Fig. 19 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections of a control *mdx*^{4cv} mouse at 10 months of age. Scale bars: 0.5 mm.

5 Fig. 20 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #1976 9-10 months after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Fig. 21 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #1982 9-10 months after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

10 Fig. 22 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #1985 9-10 months after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

Fig. 23 shows stitched large images showing dystrophin and laminin- α 2 immunostaining of the entire heart sections *mdx*^{4cv} mouse #1984 9-10 months after intravenous injection of AAV9-iNG at 5 weeks of age. Mouse number is shown in yellow. Scale bars: 0.5 mm.

15 Figs. 24a-24c show restoration of dystrophin expression in the skeletal muscles of 10-month-old *mdx*^{4cv} mice after tail vein injection of AAV9-iNG at 5 weeks of age. Fig. 24a shows immunofluorescence staining of dystrophin in diaphragm and gastrocnemius muscles of WT and *mdx*^{4cv} mice with or without systemic AAV9-iNG delivery. Fig. 24b shows Western blot of dystrophin expression in gastrocnemius muscles. Fig. 24c shows quantification of Western blot data.

20 Figs. 25a-25m show systemic delivery of AAV9-iNG improved histopathology and contractility in *mdx*^{4cv} mice. Fig. 25a shows trichrome staining of muscle and heart sections showing the extensive fibrosis in diaphragm and gastrocnemius (GA) muscles of *mdx*^{4cv} mice (10 months of age), which were substantially reduced following systemic AAV9-iNG delivery. The *mdx*^{4cv} mouse heart had little fibrosis at 10 months of age. Scale bar: 200 μ m. Figs. 25b-25d show quantification of fibrotic area of the diaphragm, gastrocnemius and heart muscles. ns, not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$ (one-way ANOVA). Figs. 25e-25h show measurement of CNF in the diaphragm (Figs. 25e and 25f) and gastrocnemius (Figs. 25g and 25h) muscles of *mdx*^{4cv} mice with or without AAV9-iNG treatment at 10 weeks (Figs. 25e and 25g) or 10 months (Figs. 25f and 25h) of age (two-tailed, unpaired Student's *t*-test). Figs. 25i-25l show muscle fiber size measurement and distribution in diaphragm and gastrocnemius muscles of the mice (WT and *mdx*^{4cv} with or without AAV9-iNG) at 10 weeks of age. ns, not

significant; $**p < 0.01$ (one-way ANOVA). Fig. 25m shows tetanic torque measurements of the posterior compartment muscles. $*p < 0.05$; $****p < 0.0001$ (one-way ANOVA).

Figs. 26a-26d show measurement of muscle fiber size in diaphragm and gastrocnemius muscles at 10 months of age. Figs. 26a-26b show average CSA of diaphragm (Fig. 26a) and gastrocnemius (Fig. 26b) muscle fibers. c, d, Distribution of diaphragm (Fig. 26c) and gastrocnemius (Fig. 26d) muscle fiber areas. ns, not significant; $*p < 0.05$ (one-way ANOVA).

Figs. 27a-27j show host immune response, toxicity and off-target studies of AAV9-iNG therapy in *mdx^{4cv}* mice. Figs. 27a and 27b show host immune response to AAV9 capsid and the base editor transgene (anti-Cas9). IM, intramuscular injection; IV, intravenous injection. $****p < 0.0001$ (one-way ANOVA) compared to WT or *mdx^{4cv}*. Figs. 27c-27e show measurements of serum AST (Fig. 27c), ALT (Fig. 27d) and BUN (Fig. 27e) of mice treated with or without AAV9-iNG. ns, not statistically significant (one-way ANOVA). Fig. 27f shows predicted off-target sites with different number of mismatches from the *mdx^{4cv}* target gRNA, and the sequences of two most similar off-target sites located on chromosome 16 and 1, respectively. Figs. 27g and 27h show quantification of deep sequencing reads of the genomic DNA PCR amplicons of the chromosome 16 off-target site (Fig. 27g) or the chromosome 1 off-target site (Fig. 27h) from Neuro-2a cells transfected with ABE-NG, iABE-NGA or control plus the gRNA. ns, not significant (one-way ANOVA). Fig. 27i shows quantification of the A⁴-to-G editing in *mdx^{4cv}* mice treated with or without AAV9-iNG. $****p < 0.0001$ (student's *t*-test). Fig. 27j shows quantification of the bystander C⁶-to-T editing in *mdx^{4cv}* mice treated with or without AAV9-iNG. $***p < 0.001$ (student's *t*-test). The sequences in Fig. 27 are GTTATCTCCTGTTCTGCAGCTGT (SEQ ID NO: 621), GTTATCTCCTGCTCTGCAGCAGA (SEQ ID NO: 628), GATATCTCCTGTCTGCAGGAGA (SEQ ID NO: 629).

Fig. 28 shows off-target activities of AAV9-iNG. The nucleotide frequency at the on-target site of the four *mdx^{4cv}* mice at 10 months after treatment with AAV9-iNG. The desired edit at A⁴ is highlighted in green and the bystander C⁶ edit in red. The sequence in Fig. 28 is GTTATCTCCTGTTCTGCAGCTGT (SEQ ID NO: 621).

Figs. 29a-29c show off-target RNA editing activities in the *mdx^{4cv}* mouse hearts (10 weeks and 10 months of age) induced by systemic delivery of AAV9-iNG. Fig. 29a shows the number of RNA SNVs in three AAV9-iNG treated mouse hearts after filtering the RNA SNVs in three control mouse hearts. Fig. 29b shows the A-to-I editing rate of all RNA SNVs. Fig. 29c shows the A-to-I editing rate of the common RNA SNVs in the three AAV9-iNG treated mouse hearts.

Figs. 30a-30b show iABE-NGA editing-mediated exon skipping of human DMD exon 55 in DMD-hiPSC-derived cardiomyocytes. Fig. 30a shows Sanger sequencing showing that iABE-

NGA mediated efficient conversion of A to G at the splice acceptor site of human DMD exon 55. Fig. 30b shows RT-PCR analysis showing that iABE-NGA editing led to skipping of exon 55 in Δ E48-54 DMD-hiPSC-derived cardiomyocytes. The WT, Δ E48-54 (green arrow) and E55 skipped transcript (red arrow) were clearly discernable.

Figs. 31a-31c show iABE-NGA editing-mediated exon 7 inclusion of human SMN2. Fig. 31a shows sequences of SMN1 and SMN2 exon 7 with the different nucleotides at position 6 highlighted in red. Two gRNAs were designed to edit A36 to G with the PAM highlighted in purple. Fig. 31b shows RT-PCR analysis of Neuro-2a cells transfected with a SMN2 reporter gene plus or minus the base editors and gRNAs. The exon 7 inclusion (FL) or exclusion (Δ 7) transcripts were measured. Fig. 31c shows quantification of the RT-PCR products. ** $p < 0.01$; **** $p < 0.0001$; ns, not significant when compared with the control group, using one-way ANOVA with Turkey post test. The sequences in Fig. 31a are GGTTCAGACAAAATCA AAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGA (SEQ ID NO: 632), GGTTCAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGA (SEQ ID NO: 633), GTGCTCACATTCCTTAAATTAAGG (SEQ ID NO: 634), gGCTCACATTCCTTAAATTAAGGA (SEQ ID NO: 635).

Figs. 32a-32c show the iABE-NGA editing of mouse Angptl3 (Fig. 32a), Apoc3 (Fig. 32b) and Asgr1 (Fig. 32c) in Neuro-2a cells. ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

DETAILED DESCRIPTION

Fusing the CRISPR-Cas9 nickase with nucleobase deaminases (e.g. cytidine or adenine deaminase), a new paradigm-shifting class of genome editing technology, termed “base editors”, have recently been developed. DNA base editors, via catalyzing the conversion of one base to another, directly and precisely install point mutations into chromosomal DNA without making DSBs. Therefore, base editing can be developed as promising therapeutics to correct the genetic diseases without DNA cleavage. In particular, the adenine base editors (ABEs) show remarkable fidelity in mouse embryos and rice as compared to cytosine base editors (CBEs), making them highly attractive in therapeutic development. Moreover, nearly half of the point mutations causing human diseases are G-to-A or C-to-T, highlighting the potential of ABEs in correcting a large number of human diseases. In particular, 174 out of 508 pathogenic point mutations for DMD are due to G:C to A:T conversion (Table 5), which can be targeted by ABE editing.

In vivo base editing can correct a custom-made mouse model of Duchenne muscular dystrophy (DMD), which carries a nonsense mutation in exon 20 with a classical 5'-TGG protospacer adjacent motif (PAM) sequence in the noncoding strand for recognition by the Cas9

from *Streptococcus pyogenes* (SpCas9). In silico analysis of the ClinVar database showed that about 42.8% of the 53469 human disease-causing mutations can be potential targets for base editing correction; however, the majority (~72.4%) of these potential targets cannot be suitable for SpCas9 base editing due to the lack of the 5'-NGG PAM sequence within the suitable distance from the mutations. Several variants of SpCas9 have recently been engineered with relaxed PAM (such as xCas9-3.7, SpCas9-NG and ScCas9) and non-G PAM. These enzymes greatly increase the target scope for correcting human mutations. However, their performance to correct genetic mutations in preclinical animal models remains to be determined. Here, the efficacy of correcting a commonly used mouse model of DMD, *mdx*^{4cv} mice were explored, using NG-targeting base editors.

The present disclosure provides fusion proteins, systems, and compositions for editing disease-associated mutations and uses thereof for treating a disease (for example, Duchenne muscular dystrophy).

Reference will now be made in detail to the embodiments of the invention, examples of which are illustrated in the drawings and the examples. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs.

Terminology

Terms used throughout this application are to be construed with ordinary and typical meaning to those of ordinary skill in the art. However, Applicant desires that the following terms be given the particular definition as defined below.

As used herein, the article “a,” “an,” and “the” means “at least one,” unless the context in which the article is used clearly indicates otherwise.

The term “comprising” and variations thereof as used herein is used synonymously with the term “including” and variations thereof and are open, non-limiting terms. Although the terms “comprising” and “including” have been used herein to describe various embodiments, the terms “consisting essentially of” and “consisting of” can be used in place of “comprising” and “including” to provide for more specific embodiments and are also disclosed.

As used herein, the terms “may,” “optionally,” and “may optionally” are used interchangeably and are meant to include cases in which the condition occurs as well as cases in

which the condition does not occur. Thus, for example, the statement that a formulation “may include an excipient” is meant to include cases in which the formulation includes an excipient as well as cases in which the formulation does not include an excipient.

The terms “about” and “approximately” are defined as being “close to” as understood by one of ordinary skill in the art. In one non-limiting embodiment, the terms are defined to be within 10%. In another non-limiting embodiment, the terms are defined to be within 5%. In still another non-limiting embodiment, the terms are defined to be within 1 %.

The term “Cas9” or “Cas9 nuclease” refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (e.g., a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs (“sgRNA”, or simply “gRNA”) can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, e.g., Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, e.g., “Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti et al., J. J., McShan W. M., Ajdic D. J., Savic D. J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A. N., Kenton S., Lai H. S., Lin S. P., Qian Y., Jia H. G., Najar F. Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S. W., Roe B. A., McLaughlin R. E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., Eckert M. R., Vogel J., Charpentier E., *Nature* 471:602-

607(2011); and “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, “The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems” (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase.

A “composition” is intended to include a combination of active agent and another compound or composition, inert (for example, a fusion protein, nucleic acid, or virus) or active, such as an adjuvant.

Dystrophinopathies are a group of muscular dystrophies resulting from mutations in the dystrophin gene, located on the short arm of the X chromosome in the Xp21 region (Kunkel et al. 1985; Monaco et al. 1985; Ray et al. 1985). Of these, “Duchenne muscular dystrophy” or “DMD” is the most common dystrophinopathy resulting from complete absence of the dystrophin gene product, the subsarcolemmal protein dystrophin (Hoffman et al. 1987a; Koenig et al. 1987; Hoffman et al. 1988). While dystrophin deficiency can be a primary cause of DMD, multiple secondary pathways are responsible for the progression of muscle necrosis, abnormal fibrosis and failure of regeneration that results in a progressively worsening clinical status. There is evidence supporting oxidative radical damage to myofibers (Rando 2002), inflammation (Spencer and Tidball 2001; Porter et al. 2002), abnormal calcium homeostasis (Allen 2010; Millay 2009), myonuclear apoptosis (Rando 2001b; Sandri et al. 2001; Tews 2002), abnormal fibrosis and failure of regeneration (Rando 2001b; Bernasconi 1995); (Melone 2000; Morrison 2000; Luz 2002). This body of literature has been validated by cross sectional genome- wide approaches that allow an overall analysis of multiple defective mechanisms in DMD (Chen et al. 2000; Porter 2003). The main symptom of DMD is muscle weakness associated with muscle wasting first with the voluntary muscles, e.g., the hips, pelvic area, thighs, shoulders, and calf muscles.

As used herein, the term “effective amount” refers to an amount of a composition necessary or sufficient to realize a desired biologic effect. An effective amount of the composition would be the amount that achieves a selected result, and such an amount could be determined as a matter of routine experimentation by a person skilled in the art. For example, an effective amount of the

composition could be that amount necessary for preventing, treating and/or ameliorating Duchenne muscular dystrophy in a subject. The term is also synonymous with “sufficient amount.”

The term “subject” is defined herein to include animals such as mammals, including, but not limited to, primates (e.g., humans), cows, sheep, goats, horses, dogs, cats, rabbits, rats, mice and the like. In some embodiments, the subject is a human.

“Pharmaceutically acceptable carrier” (sometimes referred to as a “carrier”) means a carrier or excipient that is useful in preparing a pharmaceutical or therapeutic composition that is generally safe and non-toxic, and includes a carrier that is acceptable for veterinary and/or human pharmaceutical or therapeutic use. The terms “carrier” or “pharmaceutically acceptable carrier” can include, but are not limited to, phosphate buffered saline solution, water, emulsions (such as an oil/water or water/oil emulsion) and/or various types of wetting agents.

As used herein, the term “carrier” encompasses any excipient, diluent, filler, salt, buffer, stabilizer, solubilizer, lipid, stabilizer, or other material well known in the art for use in pharmaceutical formulations. The choice of a carrier for use in a composition will depend upon the intended route of administration for the composition. The preparation of pharmaceutically acceptable carriers and formulations containing these materials is described in, e.g., *Remington's Pharmaceutical Sciences*, 21st Edition, ed. University of the Sciences in Philadelphia, Lippincott, Williams & Wilkins, Philadelphia, PA, 2005. Examples of physiologically acceptable carriers include saline, glycerol, DMSO, buffers such as phosphate buffers, citrate buffer, and buffers with other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEENTM (ICI, Inc.; Bridgewater, New Jersey), polyethylene glycol (PEG), and PLURONICSTM (BASF; Florham Park, NJ). To provide for the administration of such dosages for the desired therapeutic treatment, compositions disclosed herein can advantageously comprise between about 0.1% and 99% by weight of the total of one or more of the subject compounds based on the weight of the total composition including carrier or diluent.

A gRNA is a component of the CRISPR/Cas system. A “gRNA” (guide ribonucleic acid) herein refers to a fusion of a CRISPR-targeting RNA (crRNA) and a trans-activation crRNA (tracrRNA), providing both targeting specificity and scaffolding/binding ability for Cas9 nuclease. A “crRNA” is a bacterial RNA that confers target specificity and requires tracrRNA to bind to

Cas9. A “tracrRNA” is a bacterial RNA that links the crRNA to the Cas9 nuclease and typically can bind any crRNA. The sequence specificity of a Cas DNA-binding protein is determined by gRNAs, which have nucleotide base-pairing complementarity to target DNA sequences. The native gRNA comprises a Specificity Determining Sequence (SDS), which specifies the DNA sequence to be targeted. At least a portion of the target DNA sequence is complementary to the SDS of the gRNA. For Cas9 to successfully bind to the DNA target sequence, a region of the target sequence is complementary to the SDS of the gRNA sequence and is immediately followed by the correct protospacer adjacent motif (PAM) sequence (e.g., NGG or NG for Cas9 used herein). In some embodiments, an SDS is 100% complementary to its target sequence. In some embodiments, the SDS sequence is less than 100% complementary to its target sequence and is, thus, considered to be partially complementary to its target sequence. For example, a targeting sequence may be 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, or 90% complementary to its target sequence.

The term “linker,” as used herein, refers to a chemical group or a molecule linking two molecules or moieties, e.g., two domains of a fusion protein, such as, for example, a Cas9 domain and a nucleic acid editing domain (e.g., a deaminase domain). In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic acid editing protein. In some embodiments, a linker joins a Cas9 and a nucleic acid editing protein. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated.

The term “mutation,” as used herein, refers to a substitution of a residue within a sequence, e.g., a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

The term “nickase” as used herein, refers to a nuclease that cleaves only a single DNA strand, either due to its natural function or because it has been engineered to cleave only a single DNA strand. Jinek et al., “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity” *Science* 337(6096):816-821 (2012) and Cong et al. Multiplex genome engineering using CRISPR/Cas systems” *Science* 339(6121):819-823 (2013).

The term “nucleic acid editing domain,” as used herein refers to a protein or enzyme capable of making one or more modifications (e.g., deamination of a cytidine residue) to a nucleic acid (e.g., DNA or RNA). Exemplary nucleic acid editing domains include, but are not limited to a deaminase, a nuclease, a nickase, a recombinase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain. In some embodiments the nucleic acid editing domain comprises a deaminase (e.g., a cytidine deaminase or an adenine deaminase).

An “adenine deaminase” is an enzyme involved in purine metabolism. It is needed for the breakdown of adenosine from food and for the turnover of nucleic acids in tissues. Its primary function in humans is the development and maintenance of the immune system. An adenine deaminase catalyzes hydrolytic deamination of adenosine (forming inosine, which base pairs as G) in the context of DNA.

The term “recombinant” as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

As used herein, the terms “treating” or “treatment” of a subject includes the administration of a drug to a subject with the purpose of curing, healing, alleviating, relieving, altering, remedying, ameliorating, improving, stabilizing or affecting a disease or disorder, or a symptom of a disease or disorder. The terms “treating” and “treatment” can also refer to reduction in severity and/or frequency of symptoms, elimination of symptoms and/or underlying cause, and improvement or remediation of damage. For example, Duchenne muscular dystrophy, may result in e.g., a slowing of muscle degeneration, decreased fatigue, increased muscle strength, reduced blood levels of creatine kinase (CK), decreased difficulty with motor skills, decreased muscle fiber deformities, decreased inflammation or fibrotic tissue infiltration in the muscle, stabilization of the progression of the disease (e.g., by halting progressive muscle weakness) etc.

“Therapeutically effective amount” or “therapeutically effective dose” of a composition (e.g. a fusion protein, a nucleic acid, or virus) refers to an amount that is effective to achieve a desired therapeutic result. In some embodiments, a desired therapeutic result is the prevention of Duchenne muscular dystrophy. In some embodiments, a desired therapeutic result is the treatment of Duchenne muscular dystrophy. Therapeutically effective amounts of a given therapeutic agent will typically vary with respect to factors such as the type and severity of the disorder or disease being treated and the age, gender, and weight of the subject. The term can also refer to an amount of a therapeutic agent, or a rate of delivery of a therapeutic agent (e.g., amount over time), effective to facilitate a desired therapeutic effect, such as coughing relief. The precise desired therapeutic effect will vary according to the condition to be treated, the tolerance of the subject, the agent and/or agent formulation to be administered (e.g., the potency of the therapeutic agent, the concentration of agent in the formulation, and the like), and a variety of other factors that are appreciated by those of ordinary skill in the art. In some instances, a desired biological or medical response is achieved following administration of multiple dosages of the composition to the subject over a period of days, weeks, or years.

A “vector” is a composition of matter which comprises an isolated nucleic acid and which can be used to deliver the isolated nucleic acid to the interior of a cell. Numerous vectors are known in the art including, but not limited to, linear polynucleotides, polynucleotides associated with ionic or amphiphilic compounds, plasmids, and viruses. Thus, the term “vector” includes an autonomously replicating plasmid or a virus. The term should also be construed to include non-plasmid and non-viral compounds which facilitate transfer of nucleic acid into cells, such as, for example, polylysine compounds, liposomes, and the like. Examples of viral vectors include, but are not limited to, lentiviral vectors, adenoviral vectors, adeno-associated virus vectors, retroviral vectors, and the like.

An “adeno-associated virus” or “AAV” is a virus which infects humans and some other primate species. The wild-type AAV genome is a single-stranded deoxyribonucleic acid (ssDNA), either positive- or negative-sensed. The genome comprises two inverted terminal repeats (ITRs), one at each end of the DNA strand, and two open reading frames (ORFs): rep and cap between the ITRs. The rep ORF comprises four overlapping genes encoding Rep proteins required for the AAV life cycle. The cap ORF comprises overlapping genes encoding capsid proteins: VP1, VP2 and VP3, which interact together to form the viral capsid. VP1, VP2 and VP3 are translated from one mRNA transcript, which can be spliced in two different manners: either a longer or shorter intron can be excised resulting in the formation of two isoforms of mRNAs: a ~2.3 kb- and a ~2.6 kb-long mRNA isoform. The capsid forms a supramolecular assembly of approximately 60

individual capsid protein subunits into a non-enveloped, T-1 icosahedral lattice capable of protecting the AAV genome. The mature capsid is composed of VP1, VP2, and VP3 (molecular masses of approximately 87, 73, and 62 kDa respectively) in a ratio of about 1:1:10.

The term "nucleic acid" as used herein means a polymer composed of nucleotides, e.g. deoxyribonucleotides or ribonucleotides.

The terms "ribonucleic acid" and "RNA" as used herein mean a polymer composed of ribonucleotides.

The terms "deoxyribonucleic acid" and "DNA" as used herein mean a polymer composed of deoxyribonucleotides.

The term "oligonucleotide" denotes single- or double-stranded nucleotide multimers of from about 2 to up to about 100 nucleotides in length. Suitable oligonucleotides may be prepared by the phosphoramidite method described by Beaucage and Carruthers, *Tetrahedron Lett.*, 22: 1859-1862 (1981), or by the triester method according to Matteucci, et al., *J. Am. Chem. Soc.*, 103:3185 (1981), both incorporated herein by reference, or by other chemical methods using either a commercial automated oligonucleotide synthesizer or VLSIPSTM technology. When oligonucleotides are referred to as "double-stranded," it is understood by those of skill in the art that a pair of oligonucleotides exist in a hydrogen-bonded, helical array typically associated with, for example, DNA. In addition to the 100% complementary form of double-stranded oligonucleotides, the term "double-stranded," as used herein is also meant to refer to those forms which include such structural features as bulges and loops, described more fully in such biochemistry texts as Stryer, *Biochemistry*, Third Ed., (1988), incorporated herein by reference for all purposes.

The term "polynucleotide" refers to a single or double stranded polymer composed of nucleotide monomers.

The term "polypeptide" refers to a compound made up of a single chain of D- or L-amino acids or a mixture of D- and L-amino acids joined by peptide bonds.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher identity over a specified region when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default

parameters described below, or by manual alignment and visual inspection (see, e.g., NCBI web site or the like). Such sequences are then said to be “substantially identical.” This definition also refers to, or may be applied to, the complement of a test sequence. The definition also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the preferred algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 10 amino acids or 20 nucleotides in length, or more preferably over a region that is 10-50 amino acids or 20-50 nucleotides in length. As used herein, percent (%) nucleotide sequence identity is defined as the percentage of amino acids in a candidate sequence that are identical to the nucleotides in a reference sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN, ALIGN-2 or Megalign (DNASTAR) software. Appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full-length of the sequences being compared can be determined by known methods.

For sequence comparisons, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide

sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01.

The term “increased” or “increase” as used herein generally means an increase by a statically significant amount; for the avoidance of any doubt, “increased” means an increase of at least 10% as compared to a reference level, for example an increase of at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% increase or any increase between 10-100% as compared to a reference level, or at least about a 2-fold, or at least about a 3-fold, or at least about a 4-fold, or at least about a 5-fold or at least about a 10-fold increase, or any increase between 2-fold and 10-fold or greater as compared to a reference level.

The term “reduced”, “reduce”, “reduction”, or “decrease” as used herein generally means a decrease by a statistically significant amount. However, for avoidance of doubt, “reduced” means a decrease by at least 10% as compared to a reference level, for example a decrease by at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and

including a 100% decrease (i.e. absent level as compared to a reference sample), or any decrease between 10-100% as compared to a reference level.

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this pertains. The references disclosed are also individually and specifically incorporated by reference herein for the material contained in them that is discussed in the sentence in which the reference is relied upon.

Compositions

In some aspects, disclosed herein is a fusion protein comprising a Cas9 nickase and a nucleotide deaminase, wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11. In some embodiments, the Cas9 nickase and the nucleotide deaminase are operably linked.

In some embodiments, the first amino acid substitution is selected from the group consisting of A262T, R324L, S409I, E480K, E543D, M694I, and E1219V when compared to SEQ ID NO: 11. In some embodiments, the second amino acid substitution is selected from the group consisting of L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, R1335Q, R1335E, and T1337R when compared to SEQ ID NO: 11.

In some embodiments, the fusion protein comprises a Cas9 nickase fused to a deaminase and further fused to a Uracil DNA glycosylase inhibitor (UGI) domain.

In some embodiments, the Cas9 nickase comprises an amino acid sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20. In some embodiments, the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

In some embodiments, the Cas9 nickase when in conjunction with a bound guide RNA (gRNA) specifically binds to a target nucleic acid sequence.

In some embodiments, the Cas9 nickase recognizes a NG protospacer adjacent motif (PAM) sequence.

In some embodiments, the nucleotide deaminase is a cytidine deaminase or an adenine deaminase. In some embodiments, the nucleotide deaminase is a cytidine deaminase. In some

embodiments, the nucleotide deaminase is an adenine deaminase. In some embodiments, the adenine deaminase is a dimeric adenine deaminase or a monomeric adenine deaminase domain. In some embodiments, the dimeric adenine deaminase comprises an amino acid sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NOs: 22. In some embodiments, the monomeric adenine deaminase comprises an amino acid sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NOs: 23. In some embodiments, the monomeric adenine deaminase comprises amino acid substitutions A56G and V82G when compared to SEQ ID NO: 47. In some embodiments, the adenine deaminase comprises the amino acid sequence set forth in SEQ ID NO: 22 or 23.

In some embodiments, the adenine deaminase is encoded by a nucleotide sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NO: 41 or 42. In some embodiments, the adenine deaminase is encoded by the nucleotide sequence as set forth in SEQ ID NO: 41 or 42.

In some embodiments, the Cas9 nickase comprising the monomeric adenine deaminase domain has a lower off-target RNA editing activity than a Cas9 nickase comprising a dimeric adenine deaminase domain (e.g., at least 10% lower, at least 20% lower, at least 30% lower, at least 40% lower, at least 50% lower, at least 60% lower, at least 70% lower, at least 80% lower, at least 90% lower, at least 95% lower, at least 99% lower, or at least 2 times lower, at least 3 times lower, at least 4 times lower, at least 5 times lower, at least 6 times lower, at least 7 times lower, at least 8 times lower, at least 9 times lower, at least 10 times lower, at least 20 times lower, at least 50 times lower, at least 100 times lower, at least 150 times lower, at time 200 times lower, at least 500 times lower, or at least 1000 times lower).

In some aspects, disclosed herein is a construct comprising a nucleotide sequence encoding the fusion protein of any preceding aspect. In some embodiments, the construct further comprises a promoter operably linked to the nucleotide sequence, wherein the promoter is a CMV promoter. Accordingly, in some embodiments, the nucleotide sequence of any preceding aspect is at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NO: 1 or SEQ ID NO: 2.

In some aspects, disclosed herein is a system for base editing comprising:

a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N; and

a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

wherein the first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

5 wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

10 wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.

In some embodiments, the Cas9 nickase comprises an amino acid sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20. In some embodiments, the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

15 Accordingly, in some embodiments, the Cas9 nickase is encoded by a nucleotide sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-39. In some embodiments, the Cas9 nickase is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-39.

20 In some embodiments, the N-terminal portion of the Cas9 nickase comprises the amino acid positions 2-560, 2-561, 2-562, 2-563, 2-564, 2-565, 2-566, 2-567, 2-568, 2-569, 2-570, 2-571, 2-572, 2-573, 2-574, 2-575, 2-576, 2-577, 2-578, 2-579, or 2-580 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20. In some embodiments, the C-terminal portion of the Cas9 nickase comprises the amino acid positions 574-1368 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

25 An "intein" is a segment of a protein that is able to excise itself and join the remaining portions (the exteins) with a peptide bond in a process known as protein splicing. Inteins are also referred to as "protein introns." The process of an intein excising itself and joining the remaining portions of the protein is herein termed "protein splicing" or "intein-mediated protein splicing." In some embodiments, an intein of a precursor protein (an intein containing protein prior to intein-mediated protein splicing) comes from two genes. Such intein is referred to herein as a split intein. The inteins used herein can be, for example, Npu DnaE intein, Cfa DnaE intein or pg41-1 intein.

The intein encoded by the DnaE-n gene is herein referred as “intein-N.” The intein encoded by the DnaE-c gene is herein referred as “intein-C.”

Other intein systems may also be used. For example, a synthetic intein based on the dnaE intein, the Cfa-N and Cfa-C intein pair, has been described (e.g., in Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5, incorporated herein by reference). Non-limiting examples of
5 intein pairs that may be used in accordance with the present disclosure include: Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (e.g., as described in U.S. Pat. No. 8,394,604, incorporated herein by reference).

In some embodiments, the intein-N is a Cfa intein-N or a gp41-1 intein-N. Accordingly,
10 in some embodiments, the intein-N comprises the amino acid sequence of SEQ ID NO: 24 or 26. In some embodiments, the intein-N is encoded by the nucleotide sequence of SEQ ID NO: 43 or 45. Accordingly, the first nucleotide sequence of any preceding aspects comprises SEQ ID NO: 43 or 45.

In some embodiments, the intein-C is a Cfa intein-C or a gp41-1 intein-C. Accordingly, in
15 some embodiments, the intein-C comprises the amino acid sequence of SEQ ID NO: 25 or 27. In some embodiments, the intein-C is encoded by the nucleotide sequence of SEQ ID NO: 44 or 46. Accordingly, the second nucleotide sequence of any preceding aspects comprises SEQ ID NO: 44 or 46.

In some embodiments, the first nucleotide sequence or the second nucleotide sequence
20 further comprises a nucleotide encoding a guide RNA (gRNA).

In some embodiments, the first nucleotide sequence comprises an N terminal portion of SEQ ID NO: 1 or 2. In some embodiments, the first nucleotide comprises a portion of SEQ ID NO: 1 or 2.

In some embodiments, the second nucleotide sequence comprises a C terminal portion of
25 SEQ ID NO: 1 or 2. In some embodiments, the second nucleotide comprises a portion of SEQ ID NO: 1 or 2.

In some aspects, disclosed herein is a composition comprising:

a first recombinant viral particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N; and

30 a second recombinant viral particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

wherein the first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

5 wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.

10 In some embodiments, the Cas9 nickase comprises an amino acid sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20. In some embodiments, the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

15 Accordingly, in some embodiments, the Cas9 nickase is encoded by a nucleotide sequence at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-39. In some embodiments, the Cas9 nickase is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-39.

20 In some embodiments, the N-terminal portion of the Cas9 nickase comprises the amino acid positions 2-560, 2-561, 2-562, 2-563, 2-564, 2-565, 2-566, 2-567, 2-568, 2-569, 2-570, 2-571, 2-572, 2-573, 2-574, 2-575, 2-576, 2-577, 2-578, 2-579, or 2-580 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20. In some embodiments, the C-terminal portion of the Cas9 nickase comprises the amino acid positions 574-1368 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

25 In some embodiments, the intein-N is a Cfa intein-N or a gp41-1 intein-N. Accordingly, in some embodiments, the intein-N comprises the amino acid sequence of SEQ ID NO: 24 or 26. In some embodiments, the intein-N is encoded by the nucleotide sequence of SEQ ID NO: 43 or 45. Accordingly, the first nucleotide sequence of any preceding aspects comprises SEQ ID NO: 43 or 45.

30 In some embodiments, the intein-C is a Cfa intein-C or a gp41-1 intein-C. Accordingly, in some embodiments, the intein-C comprises the amino acid sequence of SEQ ID NO: 25 or 27. In some embodiments, the intein-C is encoded by the nucleotide sequence of SEQ ID NO: 44 or 46. Accordingly, the second nucleotide sequence of any preceding aspects comprises SEQ ID NO: 44 or 46.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. A

selected gene can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either in vivo or ex vivo. A number of retroviral systems are known in the art. In some embodiments, adenovirus vectors are used. A number of adenovirus vectors are known in the art.

5 In one embodiment, lentivirus vectors are used

AAV is a preferred vector because it can infect many cell types and is nonpathogenic to humans. AAV type vectors can transport about 4 to 5 kb and wild type AAV is known to stably insert into chromosome 19. Vectors which contain this site-specific integration property are preferred. In some embodiments, the AAV vector of any preceding aspect further a herpes simplex virus thymidine kinase gene, HSV-tk, and/or a marker gene, such as the gene encoding the green fluorescent protein, GFP.

Accordingly, in some aspects, disclosed herein is a composition comprising:

a first recombinant adeno-associated virus (AAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N;
15 and

a second recombinant AAV particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

wherein the first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

20 wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to
25 SEQ ID NO: 11.

In some embodiments, the AAV contains a pair of inverted terminal repeats (ITRs) which flank at least one cassette containing a promoter which directs cell-specific expression operably linked to a heterologous gene. Heterologous in this context refers to any nucleotide sequence or gene which is not native to the AAV or B19 parvovirus. Typically the AAV and B19 coding
30 regions have been deleted, resulting in a safe, noncytotoxic vector. The AAV ITRs, or modifications thereof, confer infectivity and site-specific integration. United states Patent No. 6,261,834 is herein incorporated by reference for material related to the AAV vector.

The inserted genes in viral and retroviral usually contain promoters, and/or enhancers to help control the expression of the desired gene product. A promoter is generally a sequence or

sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A promoter contains core elements required for basic interaction of RNA polymerase and transcription factors, and may contain upstream elements and response elements. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. Depending on the promoter, it appears that individual elements can function either cooperatively or independently to activate transcription.

The term "promoter" as used herein is defined as a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a polynucleotide sequence. One example of a suitable promoter is the immediate early cytomegalovirus (CMV) promoter sequence. This promoter sequence is a strong constitutive promoter sequence capable of driving high levels of expression of any polynucleotide sequence operatively linked thereto. Another example of a suitable promoter is Elongation Growth Factor - 1 alpha (EF-1 α). However, other promoter sequences may also be used, including, but not limited to the simian virus 40 (SV40), early promoter, mouse mammary tumor virus (MMTV), human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter, MoMuLV promoter, PGK-1 promoter, an avian leukemia virus promoter, an Epstein-Barr virus immediate early promoter, a Rous sarcoma virus promoter, as well as human gene promoters such as, but not limited to, the actin promoter, the myosin promoter, the hemoglobin promoter, and the creatine kinase promoter as well as synthetic protein, such as a CAG promoter. Further, the invention should not be limited to the use of constitutive promoters, inducible promoters are also contemplated as part of the invention. The use of an inducible promoter provides a molecular switch capable of turning on expression of the polynucleotide sequence which it is operatively linked when such expression is desired, or turning off the expression when expression is not desired. In some embodiments, the AAV of any preceding aspect comprises a constitutive promoter or a muscle tissue specific promoter, e.g., a muscle-specific MHP1 promoter.

In some embodiments, the promoter is a CMV promoter. In some embodiments, the CMV promoter comprises the nucleotide sequence of SEQ ID NO: 28. Accordingly, in some embodiments, the first recombinant adeno-associated virus (AAV) particle of any preceding aspect comprises a first nucleotide sequence that is at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NO: 3 or SEQ ID NO: 5, wherein the second recombinant adeno-associated virus (AAV) particle of any preceding aspect comprises a first nucleotide sequence that is at least

80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NO: 4 or SEQ ID NO: 6.

In some embodiments, the promoter is a MHP1 promoter. In some embodiments, the MHP1 promoter comprises the nucleotide sequence of SEQ ID NO: 29. Accordingly, in some
5 embodiments, the first recombinant adeno-associated virus (AAV) particle of any preceding aspect comprises a first nucleotide sequence that is at least 80% identity (at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NO: 7, wherein the second recombinant adeno-associated virus (AAV) particle of any preceding aspect comprises a first nucleotide sequence that is at least 80% identity (at least
10 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5%) to SEQ ID NO: 8.

In some embodiments, viral vector is a lentivirus vector. Accordingly, disclosed herein is a composition comprising:

15 a first recombinant lentivirus particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N; and

 a second recombinant lentivirus particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

 wherein first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

20 wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

 wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to
25 SEQ ID NO: 11.

Methods of introducing and expressing genes into a cell are known in the art. In the context of an expression vector, the vector can be readily introduced into a host cell, e.g., mammalian, bacterial, yeast, or insect cell by any method in the art. For example, the expression vector can be transferred into a host cell by physical, chemical, or biological means. See, e.g.,
30 WO2012079000A1, incorporated by reference herein in their entireties.

Reporter genes are used for identifying potentially transfected cells and for evaluating the functionality of regulatory sequences. In general, a reporter gene is a gene that is not present in or expressed by the recipient organism or tissue and that encodes a polypeptide whose expression is manifested by some easily detectable property, e.g., enzymatic activity. Expression of the reporter

gene is assayed at a suitable time after the DNA has been introduced into the recipient cells. Suitable reporter genes may include genes encoding luciferase, beta-galactosidase, chloramphenicol acetyl transferase, secreted alkaline phosphatase, or the green fluorescent protein gene (e.g., Ui-Tei et al., 2000 FEBS Letters 479: 79-82). Suitable expression systems are well known and may be prepared using known techniques or obtained commercially. In general, the construct with the minimal 5' flanking region showing the highest level of expression of reporter gene is identified as the promoter. Such promoter regions may be linked to a reporter gene and used to evaluate agents for the ability to modulate promoter-driven transcription. In some embodiments, the recombinant nucleic acid of any preceding aspect further comprises a reporter gene. In some embodiments, the reporter gene in invention is GFP.

In some embodiments, the fusion protein, nucleotide, system, or composition of any preceding aspect can be further formulated in a pharmaceutically acceptable carrier.

Genetic Diseases and Methods of Treatment

It is estimated that over 10,000 human diseases are caused by genetic disorders, which are abnormalities in genes or chromosomes. See, e.g., McClellan, J. and M. C. King, Genetic heterogeneity in human disease. Cell. 141(2): p. 210-7; Leachman, S. A., et al., Therapeutic siRNAs for dominant genetic skin disorders including pachyonychia congenita. J Dermatol Sci, 2008. 51(3): p. 151-7. The compositions disclosed herein can be used to treat a number of these genetic disorders.

In some aspects, disclosed herein is a method of treating a genetic disease in a subject, comprising administering to the subject a therapeutically effective amount of the fusion protein, system, or composition of any preceding aspect.

In some embodiments, the genetic disease is a muscular dystrophy. Muscular dystrophies are a group of muscle diseases caused by mutations in a person's genes. Over time, muscle weakness decreases mobility, making everyday tasks difficult. The methods and compositions disclosed herein can be used for treating a muscular dystrophy, including, for example, Duchenne muscular dystrophy, Becker muscular dystrophy, myotonic dystrophy, limb-girdle muscular dystrophy, facioscapulohumeral dystrophy, congenital muscular dystrophy, distal muscular dystrophy, oculopharyngeal muscular dystrophy, or Emery-Dreifuss muscular dystrophy.

In some embodiments, the genetic disease or disorder is Duchenne muscular dystrophy, dysferlinopathy, or spinal muscular atrophy. In some embodiments, the methods and compositions disclosed herein can be used for treating Duchenne muscular dystrophy. In some embodiments, the genetic disease is spinal muscular atrophy.

In some embodiments, the genetic disease or disorder is hypercholesterolemia. For treating hypercholesterolemia, iABE-NGA is used to create a protective silent mutation in functional genes, for example ANGPTL3, APOC3, PCSK9, or ASGR1, to lower plasma cholesterol levels.

5 In some embodiments, the genetic disease or disorder is due to a point mutation in a causative gene from a G:C pair to an A:T pair.

In some aspects, disclosed herein is a method of treating a genetic disease in a subject, comprising administering to the subject a therapeutically effective amount of the fusion protein, system, and composition of any preceding aspect.

10 The disclosed methods can be performed any time prior to the onset of a genetic disease, even prior to the apparent of any symptom. In one aspect, the disclosed methods can be employed 60, 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 years; 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 months; 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, or 3 days; 60, 48, 36, 30, 24, 18, 15, 12, 10, 9, 8, 7, 6, 5, 4, 3, or 2 hours prior to the onset of the genetic disease or any symptom thereof, or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 18, 24, 30, 36, 48, 60 hours; 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, 45, 60, 90 or more days; 4, 5, 6, 7, 8, 9, 10, 11, 12 or more months; 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, 45, or 60 years after the onset of the genetic disease or any symptom thereof.

20 Dosing frequency for the composition of any preceding aspects, includes, but is not limited to, at least once every year, once every two years, once every three years, once every four years, once every five years, once every six years, once every seven years, once every eight years, once every nine years, once every ten year, at least once every two months, once every three months, 25 once every four months, once every five months, once every six months, once every seven months, once every eight months, once every nine months, once every ten months, once every eleven months, at least once every month, once every three weeks, once every two weeks, once a week, twice a week, three times a week, four times a week, five times a week, six times a week, daily, two times per day, three times per day, four times per day, five times per day, six times per day, 30 eight times per day, nine times per day, ten times per day, eleven times per day, twelve times per day, once every 12 hours, once every 10 hours, once every 8 hours, once every 6 hours, once every 5 hours, once every 4 hours, once every 3 hours, once every 2 hours, once every hour, once every 40 min, once every 30 min, once every 20 min, once every 10 min, once every 5 min, or once per

min. Administration can also be continuous and adjusted to maintaining a level of the compound within any desired and specified range.

The compositions of the present invention can be administered to the appropriate subject in any manner known in the art, e.g., orally, intramuscularly, intravenously, sublingual mucosal, intraarterially, intrathecally, intradermally, intraperitoneally, intranasally, intrapulmonarily, intraocularly, intravaginally, intrarectally, subcutaneously, or through by inhalation. Parenteral administration, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions.

EXAMPLES

The following examples are set forth below to illustrate the compounds, systems, methods, and results according to the disclosed subject matter. These examples are not intended to be inclusive of all aspects of the subject matter disclosed herein, but rather to illustrate representative methods and results. These examples are not intended to exclude equivalents and variations of the present invention which are apparent to one skilled in the art.

Example 1. *In vitro* reporter assay demonstrates the feasibility to correct the *mdx*^{4cv} mutation using ABE-NG

The *mdx*^{4cv} mouse carries a premature stop codon (CAA-to-TAA) in the exon 53 of *Dmd* gene, which disrupts the expression of dystrophin and leads to the development of muscular dystrophy. Targeting the noncoding strand with ABEs can correct this nonsense mutation. However, in the noncoding strand, there is a lack of 5'-NGG sequence at the downstream of this mutation within the suitable editing window, but a 5'-TGT PAM is present with the mutated A located at position 4 in the guide RNA (gRNA) (Fig. 1a), making it feasible to correct the stop codon with the NG-targeting base editors in this widely used mouse model of DMD. A reporter plasmid with the targeting sequence from the *mdx*^{4cv} mice was first constructed (Fig. 1b). The nonsense mutation in the *mdx*^{4cv} targeting sequence disrupts the expression of downstream EGFP and successful editing of the nonsense mutation is indicated by the restoration of EGFP expression. As shown in Fig. 1c, transfection with the reporter alone resulted in minimal background fluorescence. Similarly, co-transfection with the reporter, *mdx*^{4cv}-gRNA and ABEmax failed to restore EGFP expression. However, ABE-NG (based on SpCas9-NG) successfully restored EGFP expression in this reporter assay. In contrast, ABE-x (based on xCas9-3.7) was found to be less efficient in restoring EGFP expression even though xCas9-3.7 was also engineered to target 5'-

NG PAM, consistent with previous reports that xCas9-3.7 is generally less efficient than SpCas9-NG. FACS analysis showed that ABE-x and ABE-NG restored EGFP expression in 10% and 20% cells, respectively (Figs. 1d, 1e). These *in vitro* studies showed that ABE-NG corrects the nonsense *mdx^{dcv}* mutation.

5

Example 2. Improvements in the editing efficiency and specificity of ABE-NG.

The relative low efficiency of ABE-NG, together with the recently reported off-target RNA editing activity, prompted the re-design of ABE-NG in order to improve the editing efficiency and specificity. First, the targeting efficiency of ABE-NG at the sites with 5'-NG PAM can be improved by optimizing the PAM-interacting domain. The targeting property of ABE-NG can be modified by combining the mutations in SpCas9-NG (R1335V/L1111R/D1135V/G1218R/E1219F/A1322R/T1337R) with other mutations designed to target different PAM sequences such as those in xCas9(3.7) (A262T/R324L/S409I/E480K/E543D/M694I/E1219V), VQR (D1135V/R1335Q/T1337R), VRER (D1135V/G1218R/R1335E/T1337R) and the loop sequence in ScCas9 (amino acids 367-376). Seven new ABE variants were generated with different combinations of the aforementioned variants (Table 1) and compared their base editing activities at six different loci with those of ABE-NG and ABEmaxSC. While all variants except ABE-NGC (containing all NG mutations plus R1335E) performed similarly at the NGG site (Fig. 2a), it was observed that ABE-NGA (carrying all NG mutations plus R1335Q) had a small improvement in editing the NGH sites as compared to ABE-NG (Figs. 2b-2e). ABE-NGA and ABE-NGX-NGC (carrying the xCas9(3.7) mutations, NG mutations and R1335E) worked equally well at the NGC site (Fig. 2c). The ABE-NG and ABE-NGA also edited the site containing a 5'-GAT PAM with high efficiency (Fig. 2f), with similar efficiency as compared to the ABE-NGX variant carrying both the xCas9(3.7) mutations and ABE-NG mutations, which was previously reported to have the broadest targeting scope active in plant. However, the efficiency of ABE-NGX at the NGH sites was lower than ABE-NGA (Fig. 2b). Since ABE-NGA is generally superior to other variants tested at NGH sites, ABE-NGA was chosen for further *in vitro* and *in vivo* studies.

Previous studies showed that the deaminase domain in the ABEs can elicit transcriptome-wide RNA off-target editing activity, and that the off-target RNA editing activity can be substantially reduced by removing the WT ecTadA domain and mutating the evolved ecTadA domain. The dimeric adenine deaminase domain (ecTadA-ecTadA*) in ABE-NG was replaced with the originally evolved ecTadA* monomer or its high-fidelity version (ecTadA*-V82G) (Fig. 3a) in order to minimize the off-target RNA editing activity. The miniABE-NG (the monomeric

TadA* fused with SpCas9-NG nickase) performed slightly better at the *mdx^{4cv}* target site as compared to ABE7.10-NG (Fig. 3b). However, the on-target DNA editing activity of miniABE(V82G)-NG was remarkably reduced by over 50% when compared to ABE7.10-NG (Fig. 3b). Then the next experiment attempted to improve the on-target DNA editing efficiency of the high-fidelity miniABE(V82G)-NG without compromising its low off-target RNA editing activity. The V82G is one of the 26 amino acid residue positions in ecTadA that reside near the enzymatic pocket around the substrate tRNA, inferred from the *S. aureus* TadA-tRNA co-crystal structure. The V82G mutation does not only affect the non-specific affinity to RNA substrates, but can also reduce its affinity to the DNA substrates. It was noticed that the A56G mutation (which is also inferred to lie near the enzymatic pocket around the tRNA substrate) had higher on-target DNA editing activity without affecting the off-target RNA editing activity as compared to miniABEmax. Installing the A56G mutation into miniABE(V82G)-NG can improve its on-target DNA editing activity without compromising its off-target RNA editing profile. Indeed, it was observed that adding the A56G mutation into miniABE(V82G)-NG (named miniABE(GG)-NG) completely restored its on-target DNA editing activity (Fig. 3b).

RNA-seq was used to compare the transcriptome-wide off-target RNA editing activities of miniABE(GG)-NG to other ABE variants in mouse Neuro2a cells. These studies were performed in triplicate. Edited RNA adenines were identified from RNA-seq experiments by filtering out background editing observed with read-count-matched controls. MiniABE-NG and miniABE(A56G)-NG induced much higher numbers of adenine editing as compared to miniABE(V82G)-NG or miniABE(GG)-NG (Fig. 3c). There was no statistical difference between miniABE(V82G)-NG and miniABE(GG)-NG. To further verify the off-target RNA editing activity of miniABE(GG)-NG, four RT-PCR amplicons were amplified and sequenced, which were shown to be highly modified by ABEmax in human cells. Transfection of HEK293 cells with miniABE-NG induced high levels of A-to-I RNA editing in all these transcripts (Fig. 3d); however, such A-to-I RNA editing was essentially eliminated in cells transfected with miniABE(GG)-NG (Fig. 3d). Taken together, these results showed that miniABE(GG)-NG does not only have increased DNA editing activity, but also inherits the high fidelity of miniABE(V82G)-NG. Hereafter, the improved ABE-NG carrying the miniABE(GG) domain and Cas9-NGA nickase was referred to as iABE-NGA.

A new generation of ABEs were developed through directed evolution, namely, ABE8s (such as ABE8.17 and ABE8.20) and ABE8e. To directly compare miniABE(GG) with ABE8.17, ABE8.20 and ABE8e, each of them was fused with SpCas9-NG and tested their activities for editing the *mdx^{4cv}* target site using the reporter assay in Neuro2A cells. All these editors showed

above 60% editing efficiency with the ABE8e-NG exhibiting the highest activity (Fig. 4a). ABE8e can have also increased bystander activity than miniABE(GG). Since the *mdx^{4cv}* target site has no extra adenine within the editing window, the bystander editing activity of ABE8e and miniABE(GG) was compared by testing their performance to edit a nonsense mutation in human

5 *DYSF* gene (encoding dysferlin) that causes limb girdle muscular dystrophy. The target A⁶ was edited with ~58% and 88% efficiency by iABE-NG and ABE8e-NG, respectively (Fig. 4b). The two bystander adenines at positions 8 and 11 were also edited at substantially higher (quantitative numbers) rates by ABE8e-NG than miniABE(GG)-NG. Even the ABE8e with V106W mutation still displayed very high bystander editing activity at position 8. Thus, for *in vivo* applications that

10 require not only efficiency but also precision, the use of miniABE(GG) was recommended.

Example 3. Intein-split allows efficient assembly of full-length ABE-NG and editing.

The large size of the ABE-NG and other base editors poses a major challenge for viral packaging and *in vivo* delivery. A dual trans-splicing adeno-associated virus (AAV) approach was

15 used to deliver ABE and a dual protein trans-splicing (PTS) approach using the split-intein moiety from *Nostoc punctiforme* (*Npu*) was used to deliver CBE. The PTS approach was adopted to deliver ABE. The ABE was split between the ecTad-ecTadA* and the Cas9 nickase with *Npu* intein moieties, and this split renders low editing efficiency (Fig. 5). To improve the editing efficiency of the split ABE, the amino acid position 573 and 574 of the Cas9 nickase were chose

20 as the splitting site because previous studies showed that 573/574 split Cas9 exhibited near the full-length Cas9 activity. Moreover, split at this site would produce a roughly equal size of the two halves for AAV packaging (Fig. 4a). the split ABE can be further improved by using inteins with fast rate of PTS. Two inteins were selected with the remarkably fast rate of PTS: Cfa ($t_{1/2} = 20$ s at 30 °C) and Gp41-1 ($t_{1/2} = 5$ s at 37 °C), which are ~2.5-fold and ~10-fold faster than the

25 rate reported for the *Npu* DnaE intein ($t_{1/2} = 50$ s at 37 °C), respectively. Transfection of both split versions into HEK293 cells resulted in robust expression of full-length ABEs as detected by the anti-Cas9 antibody (Fig. 6b), although the expression level was generally lower than the ABEmax but higher than the original ABE7.10. Co-transfection with the split ABE-NG, *mdx^{4cv}*-gRNA and the *mdx^{4cv}* reporter restored EGFP expression to a similar level as the full-length ABE-NG (Fig.

30 6c). There was no significant difference between the Cfa and Gp41-1 intein splits (Fig. 6d). The Gp41-1 version was chosen for further studies.

It was shown that the *Npu* intein split of ABE worked well *in vitro* and *in vivo*. The Gp41-1 split and *Npu* split was directly compared. While both the Gp41-1 split and *Npu* split allowed the assembly of full-length iABE-NGA, Western blotting analysis showed that the Gp41-1 split

rendered significantly more full-length iABE-NGA protein as compared to the Npu split (Figs. 6e, 6f). The assembly efficiency (as measured by the percentage of the full-length band) of the Gp41-1 split was about 70%, while the Npu split resulted in only 21% (Fig. 6g). To further compare the editing efficiency of the Gp41-1 split and Npu split, the T-to-C conversion of the *mdx*^{4cv} stop codon was quantified in Neuro2A cells using the reporter assay. As compared to the full-length iABE-NGA, the Gp41-1 split and Npu split retained about 85.5% and 78.8% of its activity, respectively (Fig. 6h). The difference in the editing efficiency between the Gp41-1 split and Npu split was small but statistically significant ($p = 0.009$) (Fig. 6h). In the Gp41-1 split and Npu split shown herein, each half carries a U6-gRNA expression cassette, while only the C-terminal half of the Npu split reported by the Liu group carries the U6-gRNA expression cassette. To test if the double U6-gRNA cassette has higher editing activity than a single U6-gRNA cassette, the U6-gRNA cassette was removed from the N-terminal construct of the Npu split used herein, which resulted in greatly reduced editing (Fig. 6h), indicating that the gRNA dosage is a limiting factor for efficient Cas9-mediated editing. Moreover, to test the specificity of intein-mediated assembly of iABE-NGA, the N and C-terminal fragments of the Gp41-1 and Npu splits were swapped, and observed no editing (Fig. 6h), indicating that the intein-mediated protein splicing and assembly of full-length iABE-NGA are required for efficient editing.

Example 4. Systematic delivery of AAV9-iNG leads to widespread dystrophin restoration.

The two Gp41-1 intein split halves of the iABE-NGA were packaged into AAV9 (hereafter referred to as AAV9-iNG) and tested if *in vivo* delivery of iABE-NGA could correct the mutation in *mdx*^{4cv} mice. A truncated MHCK7 promoter was used to drive the expression of two halves of iABE-NGA. A preliminary testing of two dosages (a total of 5×10^{13} or 1×10^{14} vg/kg, 1:1 of the N and C-terminal half) showed that the higher dose appeared to increase the dystrophin-positive myocytes in the *mdx*^{4cv} mouse heart (Fig. 7). The higher dose (1×10^{14} vg/kg, 1:1 of the N and C-terminal half) was chosen for the rest of the study. In addition, the study also showed that injection of AAV9-iNG carrying a non-targeting gRNA failed to induce dystrophin rescue (Fig. 7).

A cohort of nine *mdx*^{4cv} mice were treated with AAV9-iNG (a total of 1×10^{14} vg/kg, 1:1 of the N and C-terminal half) through a single tail vein injection at 5 weeks of age. A subset of the mice was sacrificed at 5 weeks after AAV9-iNG administration. Dystrophin was found to be widely rescued in *mdx*^{4cv} heart (Fig. 8a and Figs. 9-15). Quantification of the entire heart sections showed that $41.9 \pm 10.5\%$ cardiomyocytes of *mdx*^{4cv} mice became dystrophin positive at 10 weeks of age after systematic AAV9-iNG treatment (N=5) while the control *mdx*^{4cv} hearts were essentially dystrophin negative ($0.03 \pm 0.02\%$, N=4) (Fig. 8b). Dystrophin was also rescued in

skeletal muscles (diaphragm and gastrocnemius) of *mdx^{4cv}* mice treated with AAV9-iNG, albeit the recovery was less efficient as compared to that in the heart (Figs. 8c, 8d, and Fig. 16). Western blot analysis showed that dystrophin was rescued in *mdx^{4cv}* mouse heart to 45.9 ± 11.7 % of the WT level following systemic AAV9-iNG treatment (Figs. 8e, 8f). Consistent with the immunofluorescence data, Western blot showed dystrophin was restored to about 8.0 ± 2.6 % of the WT level in the gastrocnemius muscle of *mdx^{4cv}* mice (Fig. 17).

A group of *mdx^{4cv}* mice treated with intravenous administration of AAV9-iNG at 5 weeks of age were kept for 10 months to study the long-term impact of systemic ABE editing therapy. A near complete dystrophin restoration was observed in the hearts of all four treated *mdx^{4cv}* mice at 10 months of age (Figs. 8a, 8b and Figs. 18-23). Dystrophin was also rescued in the skeletal muscles of these older animals with a similar percentage of dystrophin-positive muscle fibers as analyzed at the 10 weeks of age (Figs. 5c, 5d and Fig. 24). Western blot analysis showed near WT levels of dystrophin expression in the hearts of the 10-month-old *mdx^{4cv}* mice treated with AAV9-iNG (Figs. 8g, 8h).

The heart and muscle tissues contain many different types of cells, which makes it challenging to precisely determine the DNA editing efficiency in myocytes. To estimate the editing efficiency of the *Dmd* gene, the total RNA was extracted from the heart tissues treated with or without AAV9-iNG, amplified the target region by RT-PCR, and analyzed the resulting amplicons by Sanger sequencing and BEAT program. The AAV9-iNG treated *mdx^{4cv}* hearts showed an average 32.6 ± 2.0 % T-to-C editing at 10 weeks of age (Figs. 8i, 8j) and 84.6 ± 2.6 % at 10 months of age (Figs. 8i, 8k).

Repeated cycles of muscle degeneration and regeneration in muscular dystrophy result in muscle fibrosis. To examine if systemic AAV9-iNG delivery can improve the histopathology of *mdx^{4cv}* mice, Trichrome staining was performed in 10-month-old mice. As compared to WT mice, the *mdx^{4cv}* mice showed significantly elevated fibrosis in both diaphragm and gastrocnemius muscles and the fibrotic areas in these muscles were significantly reduced in the *mdx^{4cv}* mice treated with AAV9-iNG (Figs. 25a-25c). Consistent with previous studies that the *mdx* mice do not develop overt cardiomyopathy before one year old, there were no significant changes in cardiac fibrosis in *mdx^{4cv}* mice with or without AAV9-iNG treatment at 10 months of age as compared to the WT controls (Fig. 25d). AAV9-iNG treatment also significantly reduced the percentage of centrally nucleated fibers (CNF) in both diaphragm and gastrocnemius muscles at 10 weeks of age (Figs. 25e, 25g). By 10 months of age, the effects of AAV9-iNG treatments on CNF were blunted (Figs. 25f, 25h). Although a significant difference in cross-sectional area (CSA) of muscle fibers was not observe following AAV9-iNG treatment (Figs. 25i, 25j, and Figs. 26a,

26b), the AAV9-iNG treatment appeared to shift the fiber size distribution towards those of the WT muscles (Figs. 25k, 25l, and Figs. 26c, 26d), particularly in gastrocnemius muscles at 10 weeks of age.

To test if systemic AAV9-iNG treatment can improve the muscle function, the muscle contractility was measured using an *in vivo* muscle test system. Maximum plantarflexion tetanic torque was measured during supramaximal electric stimulation of the tibial nerve at 150 Hz. While the *mdx*^{4cv} mice produced significantly reduced torque as compared to the WT controls, systemic delivery of AAV9-iNG significantly increased the tetanic torque in *mdx*^{4cv} mice (Fig. 25m).

Example 5. The safety profile and off-target activity of AAV9-iNG treatment.

Previous studies showed that AAV-mediated delivery of CRISPR/Cas9 into neonatal mice resulted in humoral immune responses to AAV capsid but not Cas9. In contrast, AAV-mediated delivery of CRISPR/Cas9 into adult mice evoked robust anti-Cas9 immunity. Serum samples were collected to analyze the host immune responses to the AAV9 capsid and the base editor iABE-NGA. Intramuscular injection of AAV9-iNG into 5-6 weeks old *mdx*^{4cv} mice produced robust anti-AAV9 capsid (Fig. 27a) and anti-Cas9 antibodies (Fig. 27b) at 2 weeks after injection. The anti-AAV9 titers were similar at different time points from 2 to 7 weeks post intramuscular injection and from 7 to 9 weeks post intravenous injection (Fig. 27a). The anti-Cas9 antibody titers showed a large variation among mice at 2 weeks after intramuscular injection, but all increased to peak by 4 weeks (Fig. 27b).

The liver toxicity of AAV9-iNG treatment was examined by measuring serum aspartate aminotransferase (AST) and alanine aminotransferase (ALT), and kidney toxicity by measuring blood urine nitrogen (BUN). As compared to WT mice, the *mdx*^{4cv} mice showed elevated AST (Fig. 27c) and ALT (Fig. 27d). However, treatment of *mdx*^{4cv} mice with AAV9-iNG did not further increase the serum levels of AST and ALT at either 8 weeks or 10 months of age. Measurement of BUN did not find significant changes in the treated or untreated *mdx*^{4cv} mice (Fig. 27e).

One concern with ABE-mediated gene correction is the off-target activities such as gRNA mismatch tolerance, bystander editing, and off-target RNA editing. Previous studies showed that ABE can tolerate 1-2 mismatches between the gRNA and its target sites. Prediction by Cas-OFFinder showed that one site on chromosome 16 (Chr16_OT) has only one mismatch, two other sites have two mismatches and 55 sites have three mismatches (Fig. 27f). The Chr16_OT differs from the *mdx*^{4cv} target sequence by only one C at position 12. Neuro-2a cells were transfected with ABE-NG or iABE-NG plus the gRNA, amplified the Chr16_OT by PCR and subjected the amplicon to next generation sequencing (NGS). As shown in Fig. 27g, no significant editing of

the A4 in either ABE-NG or iABE-NG transfected cells was observed. Similarly, the off-target site on chromosome 1 (Chr1_OT) was analyzed, which differs from the *mdx*^{4cv} target sequence by an A at position 2 and a G at position 20. Again, it was found that ABE-NG or iABE-NG did not edit the A4 at Chr1_OT (Fig. 27h).

Next, the bystander editing at the on-target *mdx*^{4cv} locus was analyzed in the mice treated with AAV9-iNG. Since the 10-month treated mouse hearts showed a high level of dystrophin rescue, the on-target editing efficiency was first determined in these mouse hearts by NGS. As mouse hearts contain multiple different cell types, analysis of the genomic DNA PCR products can significantly underestimate the editing efficiency. To verify this, NGS of the genomic DNA PCR products was performed from two mouse hearts receiving AAV9-iNG and exhibiting high dystrophin rescue, and an up to 11% edits at A4 was detected. Thus, the RT-PCR products were sequenced to estimate the editing efficiency at the on-target *mdx*^{4cv} locus. The A at position 4 (corresponding to the T within the premature stop codon in the coding strand) was converted to G with high efficiency from all four mouse hearts (Fig. 28). On average, 86.2 ± 2.4 % A-to-G conversion was measured (Fig. 27i). At the *mdx*^{4cv} target site, there was only one A within the editing window of 4-8, disallowing us to analyze the bystander A-to-G editing at this particular site. Another type of undesired ABE-mediated genome edits at an on-target locus is ABE-dependent cytosine-to-uracil conversion resulting in C•G to T•A mutation at that site. It was found that C6 at the *mdx*^{4cv} target site was edited above background with an average efficiency of 1.6 ± 0.1 % (Fig. 27j).

Finally, RNA-seq was performed to characterize the transcriptome-wide RNA off-target RNA editing induced by AAV9-iNG in the *mdx*^{4cv} mouse heart samples. After filtering the confident variants from control *mdx*^{4cv} heart samples, a few hundred RNA editing events were found in the three AAV9-iNG treated heart samples with only 32 shared by all of them (Fig. 29).

The present study has improved the split ABE-NG for AAV-mediated *in vivo* delivery by engineering a new NG PAM-interacting domain variant, a new adenine deaminase domain with higher on-target DNA editing efficiency without compromising the high fidelity of ABE-V82G, and a Gp41-1 intein split that mediates higher efficiency of protein splicing and editing. Together, these improvements allowed to achieve widespread dystrophin rescue and functional improvement in dystrophic mice. The editing efficiency in the heart was extraordinarily high in *mdx*^{4cv} mice following systemic delivery of AAV9-iNG and over 90% of cardiomyocytes were corrected to express dystrophin in *mdx*^{4cv} hearts at 10 months of age after a single intravenous administration of AAV9-iNG at 5 weeks old. There was no obvious toxicity detected following

AAV9-iNG treatment, despite the host immune response to the AAV9 capsid and ABE. This has tremendous implication for base correction of genetic cardiomyopathies.

It was shown herein that the ecTadA* domain with the V82G mutation had significantly reduced on-target DNA editing activity as compared to the original ecTadA*. By adding the A56G mutation into the V82G variant of ecTadA*, the on-target DNA editing activity was dramatically improved without compromising the high fidelity of the V82G variant in terms of its low off-target RNA editing activity. Although the A56G_V82G variant was less efficient than the recently reported ABE8e, ABE8e had also significantly higher bystander editing activity than the A56G_V82G variant. It is essential to consider both the on-target DNA editing efficiency and the bystander DNA editing as well as off-target RNA editing activity for *in vivo* applications. The editors have high on-target DNA editing but induce minimal bystander DNA editing and off-target RNA editing events. The A56G_V82G variant offers a balance between the editing efficiency and the editing precision.

The iABE-NG and iABE-NGA can be broadly applied to correct DMD mutations and many other disease-causing mutations. Analysis of the ClinVar database showed that over 100 of the 174 total G>A or T>C point mutations for DMD can be targeted for repair by at least one of the ABEs (iABE-NGA). Describe all other cases including targeting the splicing sites for exon skipping and other diseases.

The recent advances in engineering Cas9 variants with non-G PAM further increases targeting capacity. Moreover, the ABE editing can be designed to induce skipping of mutant exons via targeting the canonical splicing donor or acceptor, thus further broadening the applicability of ABE editing therapy for a larger population of DMD.

The mice at ten months after AAV9-iNG delivery showed significantly higher dystrophin rescue than the mice at 10 weeks after the treatment. One explanation is that the DMD cardiomyocytes with restored dystrophin expression can gain advantage for selective survival and regeneration during the development stages after delivery of AAV9-iNG. Additionally, transduced cardiomyocyte-derived extracellular vesicles can deliver genetic materials such as transcripts encoding iABE-NG into proximal un-transduced cardiomyocytes and confer base editing in those cells.

This study has also shown that systemic delivery of AAV9-iNG resulted in dystrophin restoration in skeletal muscles and functional improvement. As compared to cardiomyocytes, the editing efficiency in skeletal muscles was substantially lower. This can be attributed to the observation that AAV9 has higher tropism towards cardiomyocytes than skeletal muscles. However, other mechanisms can also be responsible for the lower editing efficiency in skeletal

muscles. For example, the dystrophic and inflammatory microenvironment in skeletal muscles can pose further constraints on AAV9 delivery and base editing. In addition, targeting muscle satellite cells can be required to improve the overall editing outcomes in skeletal muscle as they are constantly activated to replace injured skeletal muscle in DMD. Although AAV9 has been shown to transduce muscle satellite cells, the efficiency is relatively low. Moreover, the use of a muscle-specific promoter can further reduce the base editing in muscle satellite cells in the present study.

Improvements in these examples show exceptionally high editing efficiency in comparison to other approaches. First, the intein used in the current study (Gp41-1) has superfast kinetics, which allows more efficient assembly of full-length ABE (Fig. 6). Second, each half of the AAVs carries a gRNA-expressing cassette in the present study, while in the other study, gRNA is present in only the C-terminal half of the Npu intein split constructs. The data (Fig. 6h) showed that the gRNA dosage affects the editing efficiency. Third, the promoters used in these studies were also different, which can drive different expression levels of ABE in heart tissues. Finally, the intrinsic difference in the gRNAs and ABE variants can have impacts on the overall editing outcomes. Thus, the exceptionally high editing efficiency achieved in adult dystrophic mice indicates that the AAV9-iNG vectors disclosed herein are useful for clinical applications.

Example 6. Methods.

Mice. Mice (C57BL/6J and B6Ros.Cg-*Dmd*^{*mdx*^{4Cv}}/J) were purchased from the Jackson Laboratory and maintained at The Ohio State University Laboratory Animal Resources in accordance with animal use guidelines. All the experimental procedures were approved by the Animal Care, Use, and Review Committee of the Ohio State University.

Plasmid construction. The pCMV-ABE7.10, pCMV-ABE-xCas9(3.7) and pCMV-ABEmax were obtained from Addgene. The NG mutations were introduced by fusion PCR of pCMV-ABEmax and subcloned into pCMV-ABEmax to make pCMV-ABEmaxNG. The A56G and V82G mutations were introduced into TadA* domain by fusion PCR and cloned into pCMV-ABEmaxNG to generate pCMV-iABEmaxNG. The CfaN minigene was synthesized by IDTdna and fused at the amino acid 573 of SpCas9-max through PCR amplification. The TadA-TadA*-SpCas9max(2-573)-CfaN fragment was PCR amplified and subcloned into pAAV under the control of meCMV promoter to generate pAAV-ABEmaxN-temp. The hU6 promoter with *mdx*^{*4cv*}-targeting gRNA was PCR amplified and cloned into pAAV-ABEmaxN-temp to make pAAV-ABEmaxN. The CfaC fused with SpCas9max(574-end) was generated by PCR and cloned into pAAV-ABEmaxN-temp to make pAAV-ABEmaxC. Similarly, pAAV-ABEmaxN2 and pAAV-

ABEmaxC2NG with the Gp41-1 intein, and pAAV-ABEmaxN3 and pAAV-ABEmaxC3NG with the Npu intein were constructed. The *mdx^{4cv}* gRNA and other gRNA oligos (listed in Table 2) were annealed and ligated into pLenti-ogRNA. The *mdx^{4cv}* reporter oligos were annealed and ligated into pLKO-puro-2A-*mdx^{4cv}*-EGFP. All plasmids used in this study are listed in Table 3.

5 *Generation of AAV particles.* AAV vectors were produced at the viral vector core of the Nationwide Children's Hospital as previously described. The Gp41-1 intein split of iABE-NGA and the gRNA targeting *mdx^{4cv}* mutation (GTTaTCTCCTGTTCTGCAGC TGT (SEQ ID NO: 621); note: the underlined PAM sequences were not included in the gRNA) or a non-targeting gRNA (GTTTaTGTCACCAGAGTAAC (SEQ ID NO: 571), the different nucleotides are
10 highlighted in blue) expression cassettes were packaged into AAV9 capsid using the standard triple transfection protocol. A quantitative PCR-based titration method was used to determine an encapsulated vector genome titer utilizing a Prism 7500 Fast Taqman detector system (PE Applied Biosystems Grand Island, NY USA). The following primers/probes were used: 5'-GGATTTC AAGTCTCCACCC-3' (SEQ ID NO: 630) and 5'-TCCCACCGTACACGCCTAC-
15 3' (SEQ ID NO: 631) for titrating AAV9-NG, and AAV9-iNG was titrated using digital droplet PCR. Titers are expressed as DNase resistant particles per ml (DRP/ml) and rAAV titers used for injection in mice were 8.9×10^{12} DRP/ml (AAV9-NG) and 3.0×10^{13} DRP/ml (AAV9-iNG).

20 *Cell culture and transfection.* HEK293 cells were cultured in Dulbecco's modified eagle's medium (DMEM) (Corning, Manassas, VA) containing 10% fetal bovine serum (FBS) and 1% 100x penicillin-streptomycin (10,000 U/ml, Invitrogen). Cells were plated in 6-well plates and transfected with the 2 μ g plasmids (0.5 μ g reporter, 0.75 μ g gRNA and 0.75 μ g ABE) per well unless specified otherwise by polyethylenimine (PEI) as previously described.

25 *Flow cytometry.* At 72 hour post transfection, HEK293 cells transfected with ABE plasmids were collected from 6-well plate and analyzed on Becton Dickinson LSR II (BD Biosciences) to determine GFP-positive cells. A total of 100,000 cell events were collected and data analysis was performed using the FlowJo software (Tree Star, Ashland, OR, USA).

30 *Intramuscular and intravenous administration of AAV9 particles.* AAV9-iNG viral particles (2×10^{11} vg, 25 μ l) were injected into the right gastrocnemius compartment of the male *mdx^{4cv}* mice at 5-6 weeks of age or day 3. For systematic delivery, the male *mdx^{4cv}* mice at 5-6 weeks of age were administered with AAV9-NG, AAV9-iNG or AAV9-GFP viral particles (1×10^{14} vg/kg) via tail vein injection.

Serological analysis. Blood samples were collected at various time points after intramuscular or intravenous injection. The blood samples were allowed to clot for 15 min to 30 min and centrifuged at 5000 rpm for 10 min in room temperature. The supernatant was collected

as serum and stored at -80 °C for the biochemical assays. Measurement of ALT (BioVision Incorporated), AST (BioVision Incorporated), BUN (Arbor Assays, Michigan, USA) and cardiac Troponin I (Life Diagnostics, Inc) were performed according to the manufacturer's protocols.

Antibody ELISA. Antibodies against AAV9 and SpCas9 were detected by adapting previously published protocols. In brief, recombinant AAV9 (2×10^9 vg/well) and SpCas9 protein (0.27 µg/well) were diluted in 1x Coating Buffer A (BioLegend) and used to coat a 96-well Nunc MaxiSorp plate. Proteins were incubated overnight at 4 °C to adsorb to the plate. Plates were washed four times 5 min each with PBS plus 0.05% Tween-20 and then blocked with 1x Assay Diluent A (BioLegend) for 1 h at room temperature. The anti-AAV2 (A20, cat. # 03-65155, American Research Products, Inc) and anti-SpCas9 antibody (Diagenode C15310258) was used as positive control for detection of anti-AAV9 and anti-SpCas9 antibodies, respectively. Serum samples were added in 1:50 dilution and plates were incubated for 2 h at room temperature with shaking. Plates were washed four times 5 min each and 100 µl of blocking solution containing goat anti-mouse IgG (Sigma 1:3,000) was added to each well and incubated at 1 h at room temperature. Plates were washed four times 5 min each, 100 µl of freshly mixed TMB Substrate Solution (BioLegend) was added to each well, and incubated in the dark for 20 min. The reaction was stopped by adding 100 µl 2N H₂SO₄ Stop Solution. Optical density at 450 nm was measured with a plate reader.

Muscle contractility measurements. At 5 weeks after intramuscular AAV9-NG or intravenous AAV9-iNG injection, muscle contractility was measured using an *in vivo* muscle test system (AuroraScientific Inc). Mice were anesthetized with 3% (w/v) isoflurane and anesthesia was maintained by 1.5% isoflurane (w/v) during muscle contractility measurement. Maximum plantarflexion tetanic torque was measured during a train of supramaximal electric stimulations of the tibial nerve (pulse frequency 150 Hz, pulse duration 0.2 ms).

Histopathological assessment of tissues. Mice were sacrificed at various time points, and tissues (heart, lung, diaphragm, spleen, kidney, liver, quadriceps and gastrocnemius) were harvested for histological, histochemical, biochemical and molecular analyses. For immunohistological examinations, tissues were embedded in optimal cutting temperature (OCT, Sakura Finetek, Netherlands) compound and snap-frozen in cold isopentane for cryosectioning. The tissues were stored at -80°C and processed for biochemical analysis and histology assessment. Frozen cryosections (7 µm) were fixed with 4% paraformaldehyde for 15 minutes at room temperature. After washing with PBS, the slides were blocked with 3% BSA for 1 hour. The slides were incubated with primary antibodies against dystrophin (ab15277, 1:100, Abcam) and laminin-α2 (ALX-804-190-C100, 1:100, Enzo) at 4 °C for 1 hour. After that, the slides were washed

extensively with PBS and incubated with secondary antibodies (Alexa Fluor 488 goat anti-rat IgG, Invitrogen, Carlsbad, CA or Alexa Fluor 568 donkey anti-rabbit IgG, Invitrogen) for 1 hour at room temperature. The slides were sealed with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratory, Burlingame, CA). All images were taken under a Nikon Ti-E
 5 fluorescence microscope (magnification 200x) (Nikon, Melville, NY). Laminin- α 2-positive and dystrophin-positive muscle fibers were counted using NIS-Elements AR version 4.50 (Nikon, Melville, NY). The amount of dystrophin positive muscle fibers is represented as a percentage of total laminin- α 2-positive muscle fibers.

For trichrome staining, Masson's 2000 Trichrome Kit was used (American MasterTech,
 10 Lodi, CA). The muscle and heart sections were fixed with 4% paraformaldehyde for 1 hour at room temperature. After washing with PBS, the tissue sections were stained with Masson's trichrome reagent following the manufacturer's instruction.

Western blot analysis. Mouse tissues from *mdx*^{4cv} mice treated with or without AAV9-NG or AAV9-iNG were lysed with cold RIPA buffer supplemented with protease inhibitors and
 15 extracted protein samples were separated by SDS-PAGE (BioRad, 4-15%) and transferred onto Nitrocellulose membranes (0.45 μ m). The rabbit polyclonal anti-dystrophin (E2660, 1:500, Spring Bioscience, Pleasanton, CA), rabbit polyclonal anti-Cas9 (C15310258-100, 1:1000, Diagenode, Denville, NJ) and rabbit monoclonal anti-Gapdh (#2118, 1:2000, Cell Signaling Technology, Danvers, MA) antibodies were used for immunoblotting analysis. HRP conjugated
 20 goat anti-mouse (1:4000) and goat anti-rabbit (1:4000) secondary antibodies were obtained from Cell Signaling Technology, Danvers, MA. The membranes were developed using ECL western blotting substrate (Pierce Biotechnology, Rockford, IL) and scanned by ChemiDoc XRS+ system (BioRad, Hercules, CA). Western blots were quantified using Image Lab 6.0.1 software (Bio-Rad Laboratories, Hercules, CA) according to the manufacturer's instruction.

Extraction of genomic DNA and total RNA, PCR and Sanger sequencing. Genomic DNA from mouse tissues and cultured HEK293 cells were extracted using DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD). Total RNA was extracted from mouse tissues and HEK293 cells using Quick-RNA MiniPrep Kit (ZYMO Research, Irvine, CA). Five μ g of treated RNA was used as template for first-strand cDNA synthesis by using RevertAid RT Reverse Transcription Kit
 30 (Life Technologies, Carlsbad, CA). Aliquots of the RT product were used for RT-PCR analysis of dystrophin editing. PCR reactions were carried out with 100 ng genomic DNA or cDNA in the GoTaq Master Mix (Promega) according to the manufacturer's instruction. The primers used for RT-PCR of the reporter genes and PCR of endogenous loci were listed in Table 2. The PCR products were purified using the Wizard SV Gel and PCR Clean-up System (Promega). Purified

genomic DNA and RT PCR products (100 ng) were subjected to Sanger sequencing at the Genomics Shared Resource of the Ohio State University Comprehensive Cancer Center. The sequencing data were analyzed by BEAT program.

Targeted deep sequencing. The on-target and off-target loci were first amplified by genomic DNA PCR and/or RT-PCR using gene-specific primers with Illumina adapters (primers are provided in Table 4). The first PCR products were purified using a commercial purification kit (Promega, Madison, WI, USA), diluted, pooled, and subjected to a second round PCR with primers including the index sequences. The final PCR products were electrophoresed on an agarose gel, showing a single sharp peak. The quality and quantity were assayed using an Agilent Bioanalyzer 2100 (Genomics Shared Resource, Ohio State University Comprehensive Cancer Center). The purified amplicons were pooled and sent for sequencing using a MiSeq nano-scale flow cell (Paired-end 300 base-pair reads) at The Genomics Services Laboratory of Nationwide Children's Hospital. The FASTQ files were analyzed using CRISPResso2 with default parameters.

RNA-seq experiments. RNA library preparation was performed using NEBNext® Ultra™ II Directional (stranded) RNA Kit for Illumina (NEB #E7760L New England Biolabs) with an initial input of 100ng ng extracted RNA per sample, measured using Qubit RNA HS reagents (#Q32852 Invitrogen) for fragmentation, cDNA synthesis and amplification. Depletion of ribosomal RNA (rRNA) was carried out with NEBNext rRNA Depletion Kit (human, mouse, rat) from New England Biolabs (#E6310X). NEBNext Multiplex oligos indexes kits (E7335L, E7500L and E7710L) from New England Biolabs were used to barcode each library following the manufacturer protocol. RNA-seq libraries were examined using an Agilent 2100 Bioanalyzer and a High Sensitivity DNA kit (Agilent Technologies, Inc). RNA-seq libraries were sequenced on Novaseq SP Paired-End 150bp format at The Genomics Services Laboratory of Nationwide Children's Hospital.

RNA sequence variant calling and variant filtering. Illumina paired-end fastq sequencing reads were processed according to GATK Best Practices for RNA-seq variant calling. In brief, reads were aligned to the mouse mm10 reference genome using STAR version 1.5.2 in two-pass mode with the parameters implemented by the ENCODE project. Picard tools (version 2.19.0) was then applied to sort and mark duplicates of the mapped BAM files. The refined BAM files were subject to split reads that spanned splice junctions, local realignment, base recalibration and variant calling with SplitNCigarReads, IndelRealigner, BaseRecalibrator and HaplotypeCaller tools from GATK (version 4.1.2.0), respectively. Known variants in dbSNP version 142 were used during base quality recalibration. From all called variants, downstream analyses focused solely on single-nucleotide variants (SNVs) on canonical (1–22, X, Y and M) chromosomes. To

identify variants with high confidence, clusters of at least five SNVs were filtered that were within a window of 35 bases and variants with Fisher strand values >30.0, qual by depth values <2.0 and sequencing depth <10. Base edits labelled as A-to-I comprise A-to-I edits called on the positive strand as well as T-to-C edits sourced from the negative strand, since the RNAs were converted into cDNA before sequencing, both the nucleotide and its complementary base can be sequenced. Results obtained with this pipeline can underestimate the actual number of RNA edits occurring in cells because of the high stringency of the variant calling pipeline and potential under-representation of intronic and intergenic RNA in our experiments.

Any confident variants found in wild-type Neuro2a cells were considered to be SNPs and were filtered out from the base-editor-transfected groups for off-target analysis. Similarly, any confident variants found in control mdx^{dcv} heart samples were filtered out from the AAV9-iNG group for off-target analysis. The editing rate was calculated as the number of mutated reads divided by the sequencing depth for each site.

ClinVar database analysis. The ClinVar data was converted into a tab-delimited flat file. A python script (clinvar.py) was written to process the tab-delimited flat file of ClinVar data.

Statistical analysis. The data were expressed as mean $\pm$ S.E.M. and analyzed with GraphPad Prism 8.0.1 software (San Diego, USA). Statistical significance was determined using one-way ANOVA followed by Bonferroni post hoc-tests for multiple groups or student's *t*-test for two groups. A *P* value of less than 0.05 is regarded as significant.

Data availability. The sequencing data have been deposited in the NCBI SRA under project accession numbers (PRJNA673243).

Example 7. Use of the system for treatment of spinal muscular atrophy.

The iABE-NGA system can be broadly used to rescue dystrophin expression in DMD and treat other conditions, in addition to correcting point mutations in DMD and dysferlinopathy. For example, by targeting the splice sites, iABE-NGA can mediate exon skipping and rescue dystrophin expression in a human induced pluripotent stem cell-derived cardiomyocytes with a large deletion from exon 48 through 54 in DMD gene (Fig. 30).

Mutations in the survival motor neuron 1 (*SMN1*) gene cause spinal muscular atrophy (SMA). There is a paralogous gene *SMN2* in human, present in almost all SMA patients. The *SMN2* is different from *SMN1* in exon 7, with position 6 converted from C to T in *SMN2*, which results in the skipping of exon 7 in *SMN2* mRNA and non-functional *SMN2* protein. Using iABE-NGA-mediated editing of A36 showed that the exon 7 of *SMN2* was spliced in (Fig. 31), and thus this can be a therapy for SMA.

Moreover, by targeting the splice sites, one can generate loss-of-functions in the therapeutic targets such as those involved in the regulation of plasma cholesterol levels. Previous studies identified a number of protective loss-of-functions in genes such as ANGPTL3, APOC3 and ASGR1, which reduce the plasma low cholesterol levels and renders cardioprotection. iABE-NGA can install such loss-of-function mutations in the aforementioned genes, which can be used as a one-time cure for dyslipidemia (Fig. 32).

Table 1. List of ABE variants engineered in this study.

Name	Description
ABE-NG	ABEmax with SpCas9-NG mutations R1335V/L1111R/D1135V/G1218R/E1219F/A1322R/T1337R
ABE-NGA	ABE-NG with R1335Q mutation
ABE-NGC	ABE-NG with R1335E mutation
ABE-NG-loop	ABE-NG with the loop sequence from ScCas9 (amino acids 367-376) inserted
ABE-NGX	ABE-NG with A262T/R324L/S409I/E480K/E543D/M694I mutations
ABE-NGX-NGA	ABE-NGX with R1335Q
ABE-NGX-NGC	ABE-NGX with R1335E
ABE-NGX-loop	ABE-NGX with the loop sequence from ScCas9 (amino acids 367-376) inserted
ABEmaxSc	ABEmax with SpCas9 nickase replaced with ScCas9 nickase
ABE-NGm	ABE-NG with the dimeric TadA-TadA* replaced with monomeric TadA* containing two additional mutations A56G and V82G
iABE-NGA	ABE-NGA with the dimeric TadA-TadA* replaced with monomeric TadA* containing two additional mutations A56G and V82G

Table 2. List of gRNA target sequences and primers for PCR in this study.

Name	Sequence	SEQ ID NOs
Mdx4cv-gRNA	GTTATCTCCTGTTCTGCAGC	SEQ ID NO: 570
NT-gRNA	GTTTATGTCACCAGAGTAAC	SEQ ID NO: 571
mDMD-i52-F	GAGGTAATAGAGCCAAGCCCT	SEQ ID NO: 572
mDMD-i53-R	GCAAGAATTCCACTTTTCACTTCCT	SEQ ID NO: 573
mDMD-E51-F	CTGTCATCTCCAAACTAGAAATGC	SEQ ID NO: 574
mDMD-E55-R	GCAGCCTCTTGCTCACTTACTC	SEQ ID NO: 575
S1-gRNA	GATGACAGGCAGGGGCACCG	SEQ ID NO: 576
S1-F	TTCCAGTGGTTCAATGGTCA	SEQ ID NO: 577
S1-R	CTTTCAACCCGAACGGAGAC	SEQ ID NO: 578
VEGFA-S5-gRNA	GAGCGAGCAGCGTCTTCGAG	SEQ ID NO: 579
VEGFA-S12-gRNA	GCAGACGGCAGTCACTAGGG	SEQ ID NO: 580
VEGFA-S14-gRNA	GGGAAGCTGGGTGAATGGAG	SEQ ID NO: 581
VEGFA-F	AGCTGTTTGGGAGGTCAGAA	SEQ ID NO: 582
VEGFA-R	AGGGAGCAGGAAAGTGAGGT	SEQ ID NO: 583
Site13-gRNA	GTCGCAGGACAGCTTTTCCT	SEQ ID NO: 584
Site13-F	TGTAGCTACGCCTGTGATGG	SEQ ID NO: 585
Site13-R	TGCCCTGAGATCTTTTCCTC	SEQ ID NO: 586
FANCF-gRNA	GATCCAGGTGCTGCAGAAGG	SEQ ID NO: 587
FANCF-F	CTCTTGCCCTCCACTGGTTGT	SEQ ID NO: 588
FANCF-R	TCGGTAGGATGCCCTACATC	SEQ ID NO: 589
Q623X-gRNA	ATCCTACAGCATGGTGGCTG	SEQ ID NO: 590
Puro-F	AGTGGTCTCCGGAAACCTCCGCGCCCC GCAAC	SEQ ID NO: 591
GFP-R	TCCTTGAAGAAGATGGTGCG	SEQ ID NO: 592

Table 3. List of plasmids used in this study.

ID	Name	Description
pXL-0570	pCMV_ABEmax (Addgene # 112095)	Expressing ABEmax; used in Fig. 1c-e; Fig. 4b,d; Suppl. Fig. S2
pXL-0550	pCMV_xCas9_3.7_-ABE_7.10 (Addgene #108382)	Expressing ABE-x; used in Fig. 1c-e; Fig. 6b; Fig. 5.
pXL-0645	pCMV_ABEmaxNG	Expressing ABE-NG; used in Fig. 1c-e; Fig. 2a-f; Fig. 3b; Fig. 6b-6d;
pXL-0752	pLKO-puro-2A-mdx4cv-GFP	<i>mdx</i> ^{4cv} reporter; used in Fig. 1c-e; Fig. 6c, 6d, h; Fig. 4.
pXL-0631	pLenti-puro-OgRNA_mdxE53	gRNA targeting <i>mdx</i> ^{4cv} mutation; used in Fig. 1c-e; Fig. 3b-d; Fig. 6c-6h; Fig. 27g,27h; Fig. 4.
pXL-0858	pCMV_ABEmaxNG-NGA	Expressing ABE-NGA; used in Fig. 2a-f.
pXL-0869	pCMV_ABE-NGC	Expressing ABE-NGC; used in Fig. 2a-f.
pXL-0872	pCMV_ABE-NG-loop	Expression ABE-NG-loop; used in Fig. 2a-f.
pXL-0712	pCMV-ABEmaxNGX	Expressing ABE-NGX; used in Fig. 2a-f.
pXL-0868	pCMV-ABEmaxNGX-NGA	Expressing ABE-NGX-NGA; used in Fig. 2a-f.
pXL-0875	pCMV-ABEmaxNGX-NGC	Expressing ABE-NGX-NGC; used in Fig. 2a-f.
pXL-0877	pCMV-ABEmaxNGX-loop	Expressing ABE-NGX-loop; used in Fig. 2a-f.
pXL-0723	pCMV-ABEmaxSC	Expressing ABEmaxSc; used in Fig. 2a-f.
pXL-0627	pLenti-puro-S1OgRNA	S1 targeting gRNA; used in Fig. 2a.
pZC0009	pLenti-Q2440X-ogRNA	Non-targeting gRNA; used in Fig. 7.
pXL-0796	pLenti-VEGFA-S5	gRNA targeting VEGFA Site5; used in Fig. 2b.
pXL-0797	pLenti-VEGFA-S12	gRNA targeting VEGFA Site12; used in Fig. 2c.
pXL-0798	pLenti-VEGFA-S14	gRNA targeting VEGFA Site14; used in Fig. 2d.
pXL-0800	pLenti-Site13	gRNA targeting Site13; used in Fig. 2e.
pXL-0801	pLenti-FANCF-ogRNA	gRNA targeting FANCF; used in Fig. 2f.
pZC0103	pCMV_miniABE_NG(V82G)	Expressing miniABE(82G)-NG; used in Fig. 3b,3c.
pZC0104	pCMV_miniABE-NG	Expressing miniABE-NG; used in Fig. 3b-3d.
pXL-0853	pCMV_miniABEmax-NG(A56G)	Expressing miniABE(A56G)-NG; used in Fig. 3b,3c.
pXL-0854	pCMV_miniABEmaxNG(GG)	Expressing miniABE(GG)-NG; used in Fig. 3b-3d; Fig. 4.

pXL-0420	pCMV_ABE7.10 (Addgene #102919)	Expressing ABE7.10; used in Fig. 6b, Fig. 5
pXL-0670	pBac-rAAV-ABEmaxN-E53 OgRNA	Expressing Cfa Split_N of ABE and mdx4cv gRNA; used in Fig. 6b-6d.
pXL-0671	pBac-rAAV-ABEmaxC-NG-E53 OgRNA	Expressing Cfa Split_C of ABE-NG and mdx4cv gRNA; used in Fig. 6b-6d.
pXL-0672	pBac-rAAV-ABEmaxN2-E53 OgRNA	Expressing Gp41-1 Split_N of ABE and mdx4cv gRNA; used in Fig. 6b-6d.
pXL-0673	pBac-rAAV-ABEmaxC2-NG-E53 OgRNA	Expressing Gp41-1 Split_C of ABE-NG and mdx4cv gRNA; used in Fig. 6b-6d.
pZC0117	pX601-mhCMV-ABEmaxNGA-C3-E53ogRNA	Expressing Npu Split_C of iABE-NGA and mdx4cv gRNA; used in Fig. 6e-6h.
pZC0118	pX601-mhCMV-miniABEmax-N3-E53ogRNA	Expressing Npu Split_N of iABE-NGA and mdx4cv gRNA; used in Fig. 6e-6h.
pZC0119	pX601-mhCMV-miniABEmax-N3-zeo	Expressing Npu Split_N of iABE-NGA; used in Fig. 6e-6h.
pZC0031	pX601-mhCMV-miniABEmax-N2-E53ogRNA	Expressing Gp41-1 Split_N of iABE-NGA and mdx4cv gRNA; used in Fig. 6e-6h.
pZC0033	pX601-mhCMV-ABEmaxNGA-C2-E53ogRNA	Expressing Gp41-1 Split_C of iABE-NGA and mdx4cv gRNA; used in Fig. 6e-6h.
pXL-0855	pX601-MHP1-miniABEmaxNG-N2-(GG)	Expressing Gp41-1 Split_N of iABE-NGA and mdx4cv gRNA, MHP1 promoter; used for AAV9 production in Fig. 8, 27.
pXL-0706	pX601-MHP1-ABEmaxC2-NG-E53 ogRNA	Expressing Gp41-1 Split_C of iABE-NGA and mdx4cv gRNA, MHP1 promoter; used for AAV9 production in Fig. 8, 27.
pYZ1059	pCMV_NG-ABE8e	Expressing ABE8e-NG; used in Fig. 4
pZC0105	pCMV_NG-ABE8.17	Expressing ABE8.17-NG; used in Fig. 4
pZC0106	pCMV_NG-ABE8.20	Expressing ABE8.20-NG; used in Fig. 4
pZC0111	pCMV_SpG-ABE8e(V106W)	Expressing ABE8e-SpG(V106W); used in Fig. 4
pPW-0007	pLKO-puro-2A-Q623X-GFP	Q623X reporter; used in Fig. 4b.
pPW-0010	pLenti-Q623X-ogRNA	Q623X targeting gRNA; used in Fig. 4b.
pXL-0419	pCMV-ABE7.9 (Addgene #102918)	Expressing ABE7.9; used in Fig. 8.
pXL-0419	pCMV-ABE7.9 (Addgene #102918)	Expressing ABE7.9; used in Fig. 8.
pYG9010	pBac-rAAV-IntC-SpCas9n	Expressing Npu_N-SpCas9 nickase; used in Fig. 8.
pYG9011	pBac-rAAV-ABE7.10v3	Expressing GFP-TadA-TadA*-Npu_C; used in Fig. 8.

Table 4. List of primers used for NGS in this study.

Name	Sequence	
Mdx4cv-E52-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAACTCAT TACTGCTGCCCAGA	SEQ ID NO: 593
Mdx4cv-E53-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGACCTGTT CGGCTTCTTCCTTA	SEQ ID NO: 594
Mdx4cv-i52-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAATTTCC ACTGTCTTCTCTTGAGT	SEQ ID NO: 595
Mdx4cv-i53-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGCTTGCCT CTGACCTGTCCTAT	SEQ ID NO: 596
mChr16 OT-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGACTAG GGGCAAAGCAAGAT	SEQ ID NO: 597
mChr16 OT-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCTTCCAAA CTTCTGCCCATT	SEQ ID NO: 598
mChr10 T-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACACAGC GTGCTCTTTCCTTAC	SEQ ID NO: 599
mChr10 T-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGTTCAGAA GAACATCCCGTTGAC	SEQ ID NO: 600
NGS-final-F	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACA CGAC	SEQ ID NO: 601
NGS-final-R1	CAAGCAGAAGACGGCATAACGAGATCTTGTAGTGACTGGAG TTCAGACGT	SEQ ID NO: 602
NGS-final-R2	CAAGCAGAAGACGGCATAACGAGATCAGATCGTGACTGGA GTTTCAGACGT	SEQ ID NO: 603
NGS-final-R3	CAAGCAGAAGACGGCATAACGAGATCCGTCCGTGACTGGAG TTCAGACGT	SEQ ID NO: 604
NGS-final-R4	CAAGCAGAAGACGGCATAACGAGATATGTCAGTGACTGGAG TTCAGACGT	SEQ ID NO: 605
NGS-final-R5	CAAGCAGAAGACGGCATAACGAGAT GTCCGC GTGACTGGAGTTCAGACGT	SEQ ID NO: 606
NGS-final-R6	CAAGCAGAAGACGGCATAACGAGAT TTAGGC GTGACTGGAGTTCAGACGT	SEQ ID NO: 607
NGS-final-R7	CAAGCAGAAGACGGCATAACGAGAT CGATGT GTGACTGGAGTTCAGACGT	SEQ ID NO: 608
NGS-final-R8	CAAGCAGAAGACGGCATAACGAGAT TGACCA GTGACTGGAGTTCAGACGT	SEQ ID NO: 609
NGS-final-R9	CAAGCAGAAGACGGCATAACGAGAT AGTCAA GTGACTGGAGTTCAGACGT	SEQ ID NO: 610
NGS-final-R10	CAAGCAGAAGACGGCATAACGAGAT AGTTCC GTGACTGGAGTTCAGACGT	SEQ ID NO: 611

NGS-final-R11	CAAGCAGAAGACGGCATAACGAGAT GATCAG GTGACTGGAGTTCAGACGT	SEQ ID NO: 612
NGS-final-R12	CAAGCAGAAGACGGCATAACGAGAT ACAGTG GTGACTGGAGTTCAGACGT	SEQ ID NO: 613
NGS-final-R13	CAAGCAGAAGACGGCATAACGAGAT TATACT GTGACTGGAGTTCAGACGT	SEQ ID NO: 614
NGS-final-R14	CAAGCAGAAGACGGCATAACGAGAT CAACAA GTGACTGGAGTTCAGACGT	SEQ ID NO: 615
NGS-final-R15	CAAGCAGAAGACGGCATAACGAGAT GTTGTT GTGACTGGAGTTCAGACGT	SEQ ID NO: 616
NGS-final-R16	CAAGCAGAAGACGGCATAACGAGAT TCGGTT GTGACTGGAGTTCAGACGT	SEQ ID NO: 617
NGS-final-R17	CAAGCAGAAGACGGCATAACGAGAT AGTATT GTGACTGGAGTTCAGACGT	SEQ ID NO: 618
NGS-final-R18	CAAGCAGAAGACGGCATAACGAGAT TCTTGT GTGACTGGAGTTCAGACGT	SEQ ID NO: 619

Table 5

chr	pos	hgvs_c	hgvs_p	all_traits	variants	variant type	repair editor	reference_seq	sgRNA_repair
X	3117 3588	NM_0 04020. 3:c.284 3+5081 C>T	NP_003 997.1:p. Gln3427 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duc henne muscular dystrophy	G> A	intron variant	ABE	SEQ ID NO: 48	SEQ ID NO: 222; SEQ ID NO: 223
X	3246 4674	NM_0 04006. 2:c.318 8G>A	NP_003 997.1:p. Trp1063 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	C>T	nonsense	ABE	SEQ ID NO: 49	SEQ ID NO: 224; SEQ ID NO: 225; SEQ ID NO: 226; SEQ ID NO: 227
X	3221 7037	NM_0 04006. 2:c.631 7G>A	NP_003 997.1:p. Trp2106 Ter	not provided; Not Provided	C>T	nonsense	ABE	SEQ ID NO: 50	SEQ ID NO: 228; SEQ ID NO: 229
X	3236 2826	NM_0 04006. 2:c.528 7C>T	NP_003 997.1:p. Arg176 3Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 51	SEQ ID NO: 230; SEQ ID NO: 231; SEQ ID NO: 232
X	3117 8668	NM_0 04006. 2:c.102 23+1G >A	NP_004 010.1:p. Thr340=	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	synonym ous variant	ABE	SEQ ID NO: 52	SEQ ID NO: 233; SEQ ID NO: 234; SEQ ID NO: 235; SEQ ID NO: 236
X	3118 0423	NM_0 04006. 2:c.100 33C>T	NP_003 997.1:p. Arg334 5Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne	G> A	nonsense	ABE	SEQ ID NO: 53	SEQ ID NO: 237; SEQ ID NO: 238

				muscular dystrophy; not provided					
X	3147 8163	NM_0 04006. 2:c.888 0G>A	NP_003 997.1:p. Trp2960 Ter	not specified	C>T	nonsense	ABE	SEQ ID NO: 54	SEQ ID NO: 239
X	3147 9043	NM_0 04006. 2:c.860 8C>T	NP_003 997.1:p. Arg287 0Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 55	SEQ ID NO: 240
X	3269 7947	NM_0 04006. 2:c.883 C>T	NP_003 997.1:p. Arg295 Ter	Becker muscular dystrophy; Duchenne muscular dystrophy; Dilated cardiomyopathy 3B	G>A	nonsense	ABE	SEQ ID NO: 56	SEQ ID NO: 241; SEQ ID NO: 242; SEQ ID NO: 243; SEQ ID NO: 244; SEQ ID NO: 245
X	3257 3766	NM_0 04006. 2:c.168 3G>A	NP_003 997.1:p. Trp561 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 57	SEQ ID NO: 246; SEQ ID NO: 247; SEQ ID NO: 248; SEQ ID NO: 249
X	3248 4918	NM_0 04006. 2:c.280 3+1G>A		Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	C>T	splice donor variant	ABE	SEQ ID NO: 58	SEQ ID NO: 250
X	3257 3744	NM_0 04006. 2:c.170 4+1G>A		Becker muscular dystrophy; Duchenne muscular dystrophy; not specified	C>T	splice donor variant	ABE	SEQ ID NO: 59	SEQ ID NO: 251; SEQ ID NO: 252; SEQ ID NO: 253; SEQ ID NO: 254

X	3281 6541	NM_0 04006. 2:c.457 C>T	NP_003 997.1:p. Gln153 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 60	SEQ ID NO: 255; SEQ ID NO: 256; SEQ ID NO: 257; SEQ ID NO: 258
X	3234 8501	NM_0 04006. 2:c.535 3C>T	NP_003 997.1:p. Gln1785 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 61	SEQ ID NO: 259; SEQ ID NO: 260; SEQ ID NO: 261
X	3241 1811	NM_0 04006. 2:c.417 4C>T	NP_003 997.1:p. Gln1392 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 62	SEQ ID NO: 262; SEQ ID NO: 263; SEQ ID NO: 264; SEQ ID NO: 265; SEQ ID NO: 266
X	3118 2861	NM_0 04006. 2:c.985 1G>A	NP_003 997.1:p. Trp3284 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	C>T	nonsense	ABE	SEQ ID NO: 63	SEQ ID NO: 267; SEQ ID NO: 268
X	3280 9577	NM_0 04006. 2:c.565 C>T	NP_003 997.1:p. Gln189 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 64	SEQ ID NO: 269
X	3221 6981	NM_0 04006. 2:c.637 3C>T	NP_003 997.1:p. Gln2125 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 65	SEQ ID NO: 270

X	3257 3812	NM_0 04006. 2:c.163 7G>A	NP_003 997.1:p. Trp546 Ter	Motor delay; Muscle weakness; Muscle cramps; EMG abnormality; EMG: myopathic abnormalities; Calf muscle hypertrophy	C>T	nonsense	ABE	SEQ ID NO: 66	SEQ ID NO: 271; SEQ ID NO: 272; SEQ ID NO: 273; SEQ ID NO: 274; SEQ ID NO: 275
X	3150 7314	NM_0 04006. 2:c.835 7G>A	NP_003 997.1:p. Trp2786 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 67	SEQ ID NO: 276; SEQ ID NO: 277; SEQ ID NO: 278; SEQ ID NO: 279; SEQ ID NO: 280
X	3246 3458	NM_0 04006. 2:c.341 3G>A	NP_003 997.1:p. Trp1138 Ter	Becker muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 68	SEQ ID NO: 281; SEQ ID NO: 282; SEQ ID NO: 283
X	3122 3071	NM_0 04006. 2:c.933 7C>T	NP_003 997.1:p. Arg311 3Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy; not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 69	SEQ ID NO: 284; SEQ ID NO: 285
X	3192 9602	NM_0 04006. 2:c.690 6G>A	NP_003 997.1:p. Trp2302 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	5 prime UTR variant	ABE	SEQ ID NO: 70	SEQ ID NO: 286
X	3259 5765	NM_0 04006.	NP_003 997.1:p.	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 71	SEQ ID NO: 287; SEQ ID

		2:c.159 4C>T	Gln532 Ter						NO: 288; SEQ ID NO: 289; SEQ ID NO: 290
X	3257 3834	NM_0 04006. 2:c.161 5C>T	NP_003 997.1:p. Arg539 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 72	SEQ ID NO: 291; SEQ ID NO: 292; SEQ ID NO: 293; SEQ ID NO: 294; SEQ ID NO: 295
X	3167 9565	NM_0 04006. 2:c.768 2G>A	NP_003 997.1:p. Trp2561 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 73	SEQ ID NO: 296; SEQ ID NO: 297
X	3167 9564	NM_0 04006. 2:c.768 3G>A	NP_003 997.1:p. Trp2561 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 74	SEQ ID NO: 298
X	3248 5057	NM_0 04006. 2:c.266 5C>T	NP_003 997.1:p. Arg889 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 75	SEQ ID NO: 299; SEQ ID NO: 300; SEQ ID NO: 301; SEQ ID NO: 302; SEQ ID NO: 303
X	3284 9781	NM_0 04006. 2:c.133 C>T	NP_003 997.1:p. Gln45T er	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	5 prime UTR variant	ABE	SEQ ID NO: 76	SEQ ID NO: 304; SEQ ID NO: 305; SEQ ID NO: 306;

									SEQ ID NO: 307
X	3250 1767	NM_0 04006. 2:c.236 8C>T	NP_003 997.1:p. Gln790 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 77	SEQ ID NO: 308; SEQ ID NO: 309; SEQ ID NO: 310
X	3254 5250	NM_0 04006. 2:c.207 7C>T	NP_003 997.1:p. Gln693 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 78	SEQ ID NO: 311; SEQ ID NO: 312; SEQ ID NO: 313
X	3167 9429	NM_0 04006. 2:c.781 8G>A	NP_003 997.1:p. Trp2606 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	C>T	nonsense	ABE	SEQ ID NO: 79	SEQ ID NO: 314; SEQ ID NO: 315
X	3167 9492	NM_0 04006. 2:c.775 5G>A	NP_003 997.1:p. Trp2585 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 80	SEQ ID NO: 316; SEQ ID NO: 317; SEQ ID NO: 318; SEQ ID NO: 319; SEQ ID NO: 320
X	3187 5331	NM_0 04006. 2:c.695 5C>T	NP_003 997.1:p. Gln2319 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G> A	5 prime UTR variant	ABE	SEQ ID NO: 81	SEQ ID NO: 321
X	3234 2105	NM_0 04006. 2:c.591 7C>T	NP_003 997.1:p. Gln1973 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 82	SEQ ID NO: 322
X	3177 4193	NM_0 04006. 2:c.731 0- 1G>A		Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice acceptor variant	ABE	SEQ ID NO: 83	SEQ ID NO: 323; SEQ ID NO: 324

X	3117 7932	NM_0 04006. 2:c.102 62C>T	NP_003 997.1:p. Ala3421 Val	Becker muscular dystrophy; BECKER MUSCULAR DYSTROPHY; not specified	G> A	500B downstre am variant	ABE	SEQ ID NO: 84	SEQ ID NO: 325; SEQ ID NO: 326
X	3165 8123	NM_0 04006. 2:c.789 4C>T	NP_003 997.1:p. Gln2632 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 85	SEQ ID NO: 327
X	3246 8683	NM_0 04006. 2:c.297 7C>T	NP_003 997.1:p. Gln993 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 86	SEQ ID NO: 328; SEQ ID NO: 329
X	3269 9111	NM_0 04006. 2:c.831 +1G>A		Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 87	SEQ ID NO: 330; SEQ ID NO: 331; SEQ ID NO: 332
X	3241 1772	NM_0 04006. 2:c.421 3C>T	NP_003 997.1:p. Gln1405 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G> A	nonsense	ABE	SEQ ID NO: 88	SEQ ID NO: 333; SEQ ID NO: 334
X	3117 8721	NM_0 04006. 2:c.101 71C>T	NP_003 997.1:p. Arg339 lTer	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy; not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 89	SEQ ID NO: 335; SEQ ID NO: 336
X	3147 8983	NM_0 04006. 2:c.866 8G>A	NP_003 997.1:p. Glu2890 Lys	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	C>T	missense variant	ABE	SEQ ID NO: 90	SEQ ID NO: 337; SEQ ID NO: 338

X	3118 0369	NM_0 04006. 2:c.100 86+1G >A		Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 91	SEQ ID NO: 339; SEQ ID NO: 340; SEQ ID NO: 341
X	3236 4647	NM_0 04006. 2:c.508 9C>T	NP_003 997.1:p. Gln1697 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 92	SEQ ID NO: 342
X	3149 6892	NM_0 04006. 2:c.844 3C>T	NP_003 997.1:p. Gln2815 Ter	Dilated cardiomyopath y 3B; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 93	SEQ ID NO: 343; SEQ ID NO: 344; SEQ ID NO: 345; SEQ ID NO: 346
X	3241 1868	NM_0 04006. 2:c.411 7C>T	NP_003 997.1:p. Gln1373 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 94	SEQ ID NO: 347; SEQ ID NO: 348
X	3181 9974	NM_0 04006. 2:c.730 9+1G> A		Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 95	SEQ ID NO: 349
X	3144 4621	NM_0 04006. 2:c.894 4C>T	NP_003 997.1:p. Arg298 2Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; Becker	G> A	nonsense	ABE	SEQ ID NO: 96	SEQ ID NO: 350; SEQ ID NO: 351; SEQ ID NO: 352; SEQ ID NO: 353

				muscular dystrophy					
X	31627852	NM_004006.2:c.8038C>T	NP_003997.1:p.Arg2680Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 97	SEQ ID NO: 354; SEQ ID NO: 355; SEQ ID NO: 356; SEQ ID NO: 357
X	32501833	NM_004006.2:c.2302C>T	NP_003997.1:p.Arg768Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopathy 3B	G>A	nonsense	ABE	SEQ ID NO: 98	SEQ ID NO: 358; SEQ ID NO: 359; SEQ ID NO: 360
X	32365049	NM_004006.2:c.4996C>T	NP_003997.1:p.Arg1666Ter	Dilated cardiomyopathy 3B; not provided; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 99	SEQ ID NO: 361; SEQ ID NO: 362
X	32345999	NM_004006.2:c.5530C>T	NP_003997.1:p.Arg1844Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 100	SEQ ID NO: 363; SEQ ID NO: 364; SEQ ID NO: 365; SEQ ID NO: 366
X	32468573	NM_004006.2:c.3087G>A	NP_003997.1:p.Trp1029Ter	Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	C>T	nonsense	ABE	SEQ ID NO: 101	SEQ ID NO: 367; SEQ ID NO: 368
X	32518098	NM_004006.2:c.2202G>A	NP_003997.1:p.Trp734Ter	Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 102	SEQ ID NO: 369; SEQ ID NO: 370; SEQ ID NO: 371; SEQ ID NO: 372

X	3182 0055	NM_0 04006. 2:c.722 9G>A	NP_003 997.1:p. Trp2410 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	C>T	5 prime UTR variant	ABE	SEQ ID NO: 103	SEQ ID NO: 373; SEQ ID NO: 374
X	3265 0983	NM_0 04006. 2:c.961 - 5831C >T		Dilated cardiomyopath y 3B; Duchenne muscular dystrophy	G> A	intron variant	ABE	SEQ ID NO: 104	SEQ ID NO: 375; SEQ ID NO: 376; SEQ ID NO: 377; SEQ ID NO: 378; SEQ ID NO: 379
X	3147 8129	NM_0 04006. 2:c.891 4C>T	NP_003 997.1:p. Gln2972 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 105	SEQ ID NO: 380; SEQ ID NO: 381; SEQ ID NO: 382; SEQ ID NO: 383; SEQ ID NO: 384; SEQ ID NO: 385
X	3238 9605	NM_0 04006. 2:c.441 4C>T	NP_003 997.1:p. Gln1472 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G> A	nonsense	ABE	SEQ ID NO: 106	SEQ ID NO: 386; SEQ ID NO: 387; SEQ ID NO: 388; SEQ ID NO: 389
X	3259 5855	NM_0 04006. 2:c.150 4C>T	NP_003 997.1:p. Gln502 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 107	SEQ ID NO: 390
X	3269 9219	NM_0 04006. 2:c.724 C>T	NP_003 997.1:p. Gln242 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G> A	nonsense	ABE	SEQ ID NO: 108	SEQ ID NO: 391; SEQ ID NO: 392; SEQ ID NO: 393; SEQ ID NO: 394

X	3321 1304	NM_0 04006. 2:c.9G >A	NP_003 997.1:p. Trp3Ter	Becker muscular dystrophy; BECKER MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; Duchenne muscular dystrophy; not provided; Not Provided	C>T	nonsense	ABE	SEQ ID NO: 109	SEQ ID NO: 395; SEQ ID NO: 396; SEQ ID NO: 397
X	3245 4833	NM_0 04006. 2:c.343 3- 1G>A		Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	C>T	splice acceptor variant	ABE	SEQ ID NO: 110	SEQ ID NO: 398; SEQ ID NO: 399; SEQ ID NO: 400; SEQ ID NO: 401; SEQ ID NO: 402
X	3247 2252	NM_0 04006. 2:c.286 1G>A	NP_003 997.1:p. Trp954 Ter	Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 111	SEQ ID NO: 403; SEQ ID NO: 404; SEQ ID NO: 405
X	3238 9614	NM_0 04006. 2:c.440 5C>T	NP_003 997.1:p. Gln1469 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 112	SEQ ID NO: 406; SEQ ID NO: 407; SEQ ID NO: 408
X	3243 8240	NM_0 04006. 2:c.407 1+1G> A		Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 113	SEQ ID NO: 409; SEQ ID NO: 410; SEQ ID NO: 411; SEQ ID NO: 412; SEQ ID NO: 413; SEQ ID NO: 414

X	3249 1492	NM_0 04006. 2:c.240 7C>T	NP_003 997.1:p. Gln803 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 114	SEQ ID NO: 415
X	3120 6581	NM_0 04006. 2:c.964 9+1G> A		Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 115	SEQ ID NO: 416; SEQ ID NO: 417; SEQ ID NO: 418
X	3239 0175	NM_0 04006. 2:c.424 0C>T	NP_003 997.1:p. Gln1414 Ter	Duchenne muscular dystrophy; not specified	G> A	nonsense	ABE	SEQ ID NO: 116	SEQ ID NO: 419; SEQ ID NO: 420; SEQ ID NO: 421; SEQ ID NO: 422; SEQ ID NO: 423
X	3249 1387	NM_0 04006. 2:c.251 2C>T	NP_003 997.1:p. Gln838 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 117	SEQ ID NO: 424; SEQ ID NO: 425; SEQ ID NO: 426; SEQ ID NO: 427
X	3234 2154	NM_0 04006. 2:c.586 8G>A	NP_003 997.1:p. Trp1956 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 118	SEQ ID NO: 428; SEQ ID NO: 429; SEQ ID NO: 430; SEQ ID NO: 431
X	3281 6509	NM_0 04006. 2:c.489 G>A	NP_003 997.1:p. Trp163 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	C>T	nonsense	ABE	SEQ ID NO: 119	SEQ ID NO: 432; SEQ ID NO: 433
X	3245 4685	NM_0 04006. 2:c.358 0C>T	NP_003 997.1:p. Gln1194 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy;	G> A	nonsense	ABE	SEQ ID NO: 120	SEQ ID NO: 434; SEQ ID NO: 435

				Duchenne muscular dystrophy					
X	3257 3529	NM_004006.2:c.1812+1G>A		Becker muscular dystrophy; Muscular dystrophy, Becker; Duchenne muscular dystrophy; not provided; Not Provided	C>T	splice donor variant	ABE	SEQ ID NO: 121	SEQ ID NO: 436
X	3122 3046	NM_004006.2:c.9361+1G>A		Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 122	SEQ ID NO: 437
X	3254 5310	NM_004006.2:c.2017C>T	NP_003997.1:p.Gln673Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G>A	nonsense	ABE	SEQ ID NO: 123	SEQ ID NO: 438; SEQ ID NO: 439
X	3234 3232	NM_004006.2:c.5641C>T	NP_003997.1:p.Gln1881Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 124	SEQ ID NO: 440; SEQ ID NO: 441; SEQ ID NO: 442; SEQ ID NO: 443
X	3167 9430	NM_004006.2:c.7817G>A	NP_003997.1:p.Trp2606Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 125	SEQ ID NO: 444
X	3280 9559	NM_004006.2:c.583C>T	NP_003997.1:p.Arg195Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne	G>A	nonsense	ABE	SEQ ID NO: 126	SEQ ID NO: 445; SEQ ID NO: 446; SEQ ID NO: 447;

				muscular dystrophy; not provided; Not Provided					SEQ ID NO: 448
X	3257 3786	NM_004006.2:c.1663C>T	NP_003997.1:p.Gln555Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 127	SEQ ID NO: 449; SEQ ID NO: 450; SEQ ID NO: 451; SEQ ID NO: 452
X	3150 7313	NM_004006.2:c.8358G>A	NP_003997.1:p.Trp2786Ter	Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	C>T	nonsense	ABE	SEQ ID NO: 128	SEQ ID NO: 453; SEQ ID NO: 454; SEQ ID NO: 455; SEQ ID NO: 456; SEQ ID NO: 457; SEQ ID NO: 458
X	3249 1463	NM_004006.2:c.2436G>A	NP_003997.1:p.Trp812Ter	Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	C>T	nonsense	ABE	SEQ ID NO: 129	SEQ ID NO: 459; SEQ ID NO: 460; SEQ ID NO: 461
X	3165 8118	NM_004006.2:c.7899G>A	NP_003997.1:p.Trp2633Ter	not specified	C>T	nonsense	ABE	SEQ ID NO: 130	SEQ ID NO: 462; SEQ ID NO: 463; SEQ ID NO: 464; SEQ ID NO: 465; SEQ ID NO: 466
X	3244 8495	NM_004006.2:c.3747G>A	NP_003997.1:p.Trp1249Ter	Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 131	SEQ ID NO: 467; SEQ ID NO: 468; SEQ ID NO: 469; SEQ ID NO: 470;

									SEQ ID NO: 471
X	3134 8571	NM_0 04006. 2:c.914 8C>T	NP_003 997.1:p. Gln3050 Ter	not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 132	SEQ ID NO: 472; SEQ ID NO: 473; SEQ ID NO: 474; SEQ ID NO: 475
X	3248 5072	NM_0 04006. 2:c.265 0C>T	NP_003 997.1:p. Gln884 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy; not provided	G> A	nonsense	ABE	SEQ ID NO: 133	SEQ ID NO: 476; SEQ ID NO: 477; SEQ ID NO: 478; SEQ ID NO: 479
X	3246 3444	NM_0 04006. 2:c.342 7C>T	NP_003 997.1:p. Gln1143 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 134	SEQ ID NO: 480; SEQ ID NO: 481; SEQ ID NO: 482
X	3256 5742	NM_0 04006. 2:c.195 2G>A	NP_003 997.1:p. Trp651 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B	C>T	nonsense	ABE	SEQ ID NO: 135	SEQ ID NO: 483; SEQ ID NO: 484; SEQ ID NO: 485; SEQ ID NO: 486
X	3114 7421	NM_0 04006. 2:c.106 51C>T	NP_003 997.1:p. Gln3551 Ter	not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 136	
X	3117 8700	NM_0 04006. 2:c.101 92C>T	NP_003 997.1:p. Gln3398 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 137	

X	3117 8751	NM_0 04006. 2:c.101 41C>T	NP_003 997.1:p. Arg338 1Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 138	
X	3118 0437	NM_0 04006. 2:c.100 19G>A	NP_003 997.1:p. Cys334 0Tyr	Duchenne muscular dystrophy, mental retardation, and absence of erg b-wave; DUCHENNE MUSCULAR DYSTROPHY, MENTAL RETARDATI ON, AND ABSENCE OF ERG B-WAVE	C>T	missense variant	ABE	SEQ ID NO: 139	
X	3118 2784	NM_0 04006. 2:c.992 8C>T	NP_003 997.1:p. Gln3310 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 140	
X	3120 9497	NM_0 04006. 2:c.956 3+1G> A		Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 141	
X	3122 3062	NM_0 04006. 2:c.934 6C>T	NP_003 997.1:p. Gln3116 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 142	
X	3147 8330	NM_0 04006. 2:c.871 3C>T	NP_003 997.1:p. Arg290 5Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath	G> A	nonsense	ABE	SEQ ID NO: 143	

				y 3B; Becker muscular dystrophy					
X	31478995	NM_004006.2:c.8656C>T	NP_003997.1:p.Gln2886Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 144	
X	31627681	NM_004006.2:c.8209C>T	NP_003997.1:p.Gln2737Ter	Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 145	
X	31679575	NM_004006.2:c.7672C>T	NP_003997.1:p.Gln2558Ter	not provided; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 146	
X	31836729	NM_004006.2:c.7189C>T	NP_003997.1:p.Gln2397Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G>A	5 prime UTR variant	ABE	SEQ ID NO: 147	
X	31929718	NM_004006.2:c.6790C>T	NP_003997.1:p.Gln2264Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G>A	5 prime UTR variant	ABE	SEQ ID NO: 148	
X	32287536	NM_004006.2:c.6283C>T	NP_003997.1:p.Arg2095Ter	Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	G>A	nonsense	ABE	SEQ ID NO: 149	
X	32287596	NM_004006.2:c.6223C>T	NP_003997.1:p.Gln2075Ter	Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 150	
X	32287680	NM_004006.2:c.6139C>T	NP_003997.1:p.Gln2047Ter	Elevated serum creatine phosphokinase	G>A	nonsense	ABE	SEQ ID NO: 151	

X	3228 7702	NM_0 04006. 2:c.611 8- 1G>A		Duchenne muscular dystrophy	C>T	splice acceptor variant	ABE	SEQ ID NO: 152	
X	3234 2171	NM_0 04006. 2:c.585 1C>T	NP_003 997.1:p. Gln1951 Ter	not provided; not provided	G> A	nonsense	ABE	SEQ ID NO: 153	
X	3234 5975	NM_0 04006. 2:c.555 4C>T	NP_003 997.1:p. Gln1852 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 154	
X	3234 5978	NM_0 04006. 2:c.555 1C>T	NP_003 997.1:p. Gln1851 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 155	
X	3234 6023	NM_0 04006. 2:c.550 6C>T	NP_003 997.1:p. Gln1836 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 156	
X	3236 4602	NM_0 04006. 2:c.513 4C>T	NP_003 997.1:p. Gln1712 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 157	
X	3236 4605	NM_0 04006. 2:c.513 1C>T	NP_003 997.1:p. Gln1711 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 158	
X	3236 5175	NM_0 04006. 2:c.487 0C>T	NP_003 997.1:p. Gln1624 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 159	
X	3238 9536	NM_0 04006. 2:c.448 3C>T	NP_003 997.1:p. Gln1495 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 160	

X	3238 9644	NM_0 04006. 2:c.437 5C>T	NP_003 997.1:p. Arg145 9Ter	Dilated cardiomyopath y 3B; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 161	
X	3245 4778	NM_0 04006. 2:c.348 7C>T	NP_003 997.1:p. Gln1163 Ter	not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 162	
X	3246 3438	NM_0 04006. 2:c.343 2+1G> A		Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 163	
X	3246 4585	NM_0 04006. 2:c.327 6+1G> A		Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 164	
X	3246 4603	NM_0 04006. 2:c.325 9C>T	NP_003 997.1:p. Gln1087 Ter	not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 165	
X	3246 8509	NM_0 04006. 2:c.315 1C>T	NP_003 997.1:p. Arg105 1Ter	Dilated cardiomyopath y 3B; Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 166	
X	3246 8704	NM_0 04006. 2:c.295 6C>T	NP_003 997.1:p. Gln986 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy; Dilated cardiomyopath y 3B; not	G> A	nonsense	ABE	SEQ ID NO: 167	

				provided; Not Provided					
X	3247 2310	NM_0 04006. 2:c.280 4- 1G>A		Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice acceptor variant	ABE	SEQ ID NO: 168	
X	3248 4925	NM_0 04006. 2:c.279 7C>T	NP_003 997.1:p. Gln933 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 169	
X	3248 4964	NM_0 04006. 2:c.275 8C>T	NP_003 997.1:p. Gln920 Ter	Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	G>A	nonsense	ABE	SEQ ID NO: 170	
X	3249 1276	NM_0 04006. 2:c.262 2+1G>A		Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	C>T	splice donor variant	ABE	SEQ ID NO: 171	
X	3249 1344	NM_0 04006. 2:c.255 5G>A	NP_003 997.1:p. Trp852 Ter	not provided; Not Provided	C>T	nonsense	ABE	SEQ ID NO: 172	
X	3249 1480	NM_0 04006. 2:c.241 9C>T	NP_003 997.1:p. Gln807 Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 173	
X	3250 1803	NM_0 04006. 2:c.233 2C>T	NP_003 997.1:p. Gln778 Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne	G>A	nonsense	ABE	SEQ ID NO: 174	

				muscular dystrophy					
X	3254 5158	NM_0 04006. 2:c.216 8+1G> A		not provided; Not Provided; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 175	
X	3254 5190	NM_0 04006. 2:c.213 7C>T	NP_003 997.1:p. Gln713 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 176	
X	3256 5704	NM_0 04006. 2:c.199 0C>T	NP_003 997.1:p. Gln664 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 177	
X	3256 5782	NM_0 04006. 2:c.191 2C>T	NP_003 997.1:p. Gln638 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 178	
X	3257 3796	NM_0 04006. 2:c.165 3G>A	NP_003 997.1:p. Trp551 Ter	Duchenne muscular dystrophy; not provided	C>T	nonsense	ABE	SEQ ID NO: 179	
X	3259 5870	NM_0 04006. 2:c.148 9C>T	NP_003 997.1:p. Gln497 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 180	
X	3261 4320	NM_0 04006. 2:c.146 5C>T	NP_003 997.1:p. Gln489 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 181	

X	3261 4397	NM_0 04006. 2:c.138 8G>A	NP_003 997.1:p. Trp463 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 182	
X	3264 4131	NM_0 04006. 2:c.133 1+1G> A		Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 183	
X	3264 4139	NM_0 04006. 2:c.132 4C>T	NP_003 997.1:p. Gln442 Ter	Duchenne muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 184	
X	3264 4202	NM_0 04006. 2:c.126 1C>T	NP_003 997.1:p. Gln421 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 185	
X	3264 5020	NM_0 04006. 2:c.109 3C>T	NP_003 997.1:p. Gln365 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 186	
	3280 9492	NM_0 04006. 2:c.649 +1G>A		Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 187	
X	3281 6641	NM_0 04006. 2:c.358 -1G>A		Becker muscular dystrophy	C>T	splice acceptor variant	ABE	SEQ ID NO: 188	
X	3282 3297	NM_0 04006. 2:c.355 C>T	NP_003 997.1:p. Gln119 Ter	Duchenne muscular dystrophy; Duchenne	G> A	5 prime UTR variant	ABE	SEQ ID NO: 189	

				muscular dystrophy					
X	3284 9727	NM_004006.2:c.186+1G>A		Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 190	
X	3284 9736	NM_004006.2:c.178C>T	NP_003997.1:p.Gln60Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; not provided	G>A	5 prime UTR variant	ABE	SEQ ID NO: 191	
X	3302 0138	NM_004006.2:c.93+1G>A		Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice donor variant	ABE	SEQ ID NO: 192	
X	3259 5756	NM_004006.2:c.1602+1G>A		not provided; Not Provided	C>T	splice donor variant	ABE	SEQ ID NO: 193	SEQ ID NO: 487
X	3120 6668	NM_004006.2:c.9564-1G>A		Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	splice acceptor variant	ABE	SEQ ID NO: 194	SEQ ID NO: 488; SEQ ID NO: 489; SEQ ID NO: 490; SEQ ID NO: 491
X	3221 7062	NM_004006.2:c.6292C>T	NP_003997.1:p.Arg2098Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopathy 3B	G>A	nonsense	ABE	SEQ ID NO: 195	SEQ ID NO: 492
X	3264 4314	NM_004006.2:c.1150-1G>A		Duchenne muscular dystrophy; not provided; Not Provided	C>T	splice acceptor variant	ABE	SEQ ID NO: 196	SEQ ID NO: 493; SEQ ID NO: 494; SEQ ID NO: 495

X	3284 9821	NM_0 04006. 2:c.94- 1G>A		Duchenne muscular dystrophy; Duchenne muscular dystrophy	C>T	splice acceptor variant	ABE	SEQ ID NO: 197	SEQ ID NO: 496
X	3149 6871	NM_0 04006. 2:c.846 4C>T	NP_003 997.1:p. Gln2822 Ter	Dilated cardiomyopath y 3B; Dilated cardiomyopath y 3B	G> A	nonsense	ABE	SEQ ID NO: 198	SEQ ID NO: 497; SEQ ID NO: 498; SEQ ID NO: 499
X	3238 0626	NM_0 04006. 2:c.472 9C>T	NP_003 997.1:p. Arg157 7Ter	Duchenne muscular dystrophy; not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 199	SEQ ID NO: 500; SEQ ID NO: 501
X	3264 5052	NM_0 04006. 2:c.106 1G>A	NP_003 997.1:p. Trp354 Ter	Duchenne muscular dystrophy	C>T	nonsense	ABE	SEQ ID NO: 200	SEQ ID NO: 502; SEQ ID NO: 503; SEQ ID NO: 504; SEQ ID NO: 505
X	3249 1414	NM_0 04006. 2:c.248 5C>T	NP_003 997.1:p. Gln829 Ter	not specified	G> A	nonsense	ABE	SEQ ID NO: 201	SEQ ID NO: 506; SEQ ID NO: 507; SEQ ID NO: 508; SEQ ID NO: 509; SEQ ID NO: 510
X	3116 9519	NM_0 04020. 3:c.284 3+9150 C>T	NP_003 997.1:p. Gln3493 Ter	Duchenne muscular dystrophy; Becker muscular dystrophy	G> A	intron variant	ABE	SEQ ID NO: 202	SEQ ID NO: 511; SEQ ID NO: 512; SEQ ID NO: 513; SEQ ID NO: 514
X	3236 4704	NM_0 04006. 2:c.503 2C>T	NP_003 997.1:p. Gln1678 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 203	SEQ ID NO: 515; SEQ ID NO: 516

X	3254 5295	NM_0 04006. 2:c.203 2C>T	NP_003 997.1:p. Gln678 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 204	SEQ ID NO: 517; SEQ ID NO: 518; SEQ ID NO: 519; SEQ ID NO: 520; SEQ ID NO: 521; SEQ ID NO: 522; SEQ ID NO: 523
X	3117 8784	NM_0 04006. 2:c.101 08C>T	NP_003 997.1:p. Arg337 0Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 205	SEQ ID NO: 524; SEQ ID NO: 525; SEQ ID NO: 526; SEQ ID NO: 527
X	3234 2123	NM_0 04006. 2:c.589 9C>T	NP_003 997.1:p. Arg196 7Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 206	SEQ ID NO: 528; SEQ ID NO: 529
X	3172 9736	NM_0 04006. 2:c.755 5G>A	NP_003 997.1:p. Asp251 9Asn	not provided; Not Provided; Duchenne muscular dystrophy	C>T	missense variant	ABE	SEQ ID NO: 207	SEQ ID NO: 530; SEQ ID NO: 531; SEQ ID NO: 532
X	3317 4335	NM_0 04006. 2:c.31+ 36947 G>A		Duchenne muscular dystrophy; Becker muscular dystrophy;	C>T	intron variant	ABE	SEQ ID NO: 208	SEQ ID NO: 533; SEQ ID NO: 534

				Dilated cardiomyopathy 3B					
X	3247 2247	NM_004006.2:c.2866C>T	NP_003997.1:p.Gln956Ter	Dilated cardiomyopathy 3B; Dilated cardiomyopathy 3B	G>A	nonsense	ABE	SEQ ID NO: 209	SEQ ID NO: 535; SEQ ID NO: 536; SEQ ID NO: 537; SEQ ID NO: 538
X	3238 9496	NM_004006.2:c.4518+5G>A		Dilated cardiomyopathy 3B; Duchenne muscular dystrophy	C>T	intron variant	ABE	SEQ ID NO: 210	SEQ ID NO: 539; SEQ ID NO: 540; SEQ ID NO: 541
X	3243 8372	NM_004006.2:c.3940C>T	NP_003997.1:p.Arg1314Ter	Becker muscular dystrophy; BECKER MUSCULAR DYSTROPHY; not provided; Not Provided; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 211	SEQ ID NO: 542; SEQ ID NO: 543
X	3120 6663	NM_004006.2:c.9568C>T	NP_003997.1:p.Arg3190Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopathy 3B; Becker muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 212	SEQ ID NO: 544; SEQ ID NO: 545; SEQ ID NO: 546; SEQ ID NO: 547
X	3172 9634	NM_004006.2:c.7657C>T	NP_003997.1:p.Arg2553Ter	Dilated cardiomyopathy 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G>A	nonsense	ABE	SEQ ID NO: 213	SEQ ID NO: 548

X	3150 7280	NM_0 04006. 2:c.839 0+1G> A		not provided; Not Provided	C>T	splice donor variant	ABE	SEQ ID NO: 214	SEQ ID NO: 549; SEQ ID NO: 550; SEQ ID NO: 551; SEQ ID NO: 552
X	3282 3316	NM_0 04006. 2:c.336 G>A	NP_003 997.1:p. Trp112 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	C>T	5 prime UTR variant	ABE	SEQ ID NO: 215	SEQ ID NO: 553; SEQ ID NO: 554; SEQ ID NO: 555
X	3281 6565	NM_0 04006. 2:c.433 C>T	NP_003 997.1:p. Arg145 Ter	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B; not provided; Not Provided; Becker muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 216	SEQ ID NO: 556; SEQ ID NO: 557; SEQ ID NO: 558; SEQ ID NO: 559
X	3246 3576	NM_0 04006. 2:c.329 5C>T	NP_003 997.1:p. Gln1099 Ter	Dilated cardiomyopath y 3B; Becker muscular dystrophy; Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 217	SEQ ID NO: 560; SEQ ID NO: 561
X	3284 4794	NM_0 04006. 2:c.253 C>T	NP_003 997.1:p. Gln85T er	Duchenne muscular dystrophy; DUCHENNE MUSCULAR DYSTROPHY	G> A	5 prime UTR variant	ABE	SEQ ID NO: 218	SEQ ID NO: 562
X	3246 8539	NM_0 04006. 2:c.312 1C>T	NP_003 997.1:p. Gln1041 Ter	Duchenne muscular dystrophy; DUCHENNE	G> A	nonsense	ABE	SEQ ID NO: 219	SEQ ID NO: 563; SEQ ID NO: 564;

				MUSCULAR DYSTROPHY; Dilated cardiomyopath y 3B					SEQ ID NO: 565; SEQ ID NO: 566
X	3239 0103	NM_0 04006. 2:c.431 2C>T	NP_003 997.1:p. Gln1438 Ter	Duchenne muscular dystrophy	G> A	nonsense	ABE	SEQ ID NO: 220	SEQ ID NO: 567; SEQ ID NO: 568
X	3234 2264	NM_0 04006. 2:c.575 8C>T	NP_003 997.1:p. Gln1920 Ter	not provided; Not Provided	G> A	nonsense	ABE	SEQ ID NO: 221	SEQ ID NO: 569

SEQUENCES

SEQ ID NO: 1 (pCMV-ABEmaxNG)

5 ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTG
 GCATTATGCCCAGTACATGACCTTATGGGACTTTTCCTACTTGGCAGTACATCTACGT
 ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTG
 ATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAATGGGAG
 TTTGTTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCC
 10 ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTG
 GTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTAT
 AGGGAGAGCCGCCACCATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCA
 AAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCA
 CGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCC
 15 GTGCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCC
 GCCACGACCCTACCGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTC
 ATGCAGAATTACCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGT
 GATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCA
 CGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGG
 20 CATGAACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCC
 CTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGC
 ACAGAGCTCCACCGACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACAC
 CAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCC
 TCTGAGGTGGAGTTTTCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAA
 25 GAGGGCACGCGATGAGAGGGAGGTGCTGTGGGAGCCGTGCTGGTGTGAACAAT
 AGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCCC
 ATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTG
 ATTGACGCCACCCTGTACGTGACATTTCGAGCCTTGCGTGATGTGCGCCGGCGCCATG
 ATCCACTCTAGGATCGGCCGCGTGGTGTGTTGGCGTGAGGAACGCAAAAACCGGCGC
 30 CGCAGGCTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAA
 TTACCGAGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTTGGA
 TGCTTAGACAGGTGTTCAATGCTCAGAAGAAGGCCCAGAGCTCCACCGACTCCGGA
 GGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGAGAGCGCAAC
 ACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGC
 35 CTGGCCATCGGCACCAACTCTGTGGGCTGGGCGGTGATCACCGACGAGTACAAGGT
 GCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAG
 AACCTGATCGGAGCCCTGCTGTTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCT
 GAAGAGAACC GCCAGAAGAAGATACACCAGACGGAAGAACC GGATCTGCTATCTG
 CAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACT
 40 GGAAGAGTCCTTCCTGGTGGAAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG
 GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTG
 AGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGC
 CCTGGCCACATGATCAAGTTCCGGGGGCCACTTCCTGATCGAGGGCGACCTGAACC
 CCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAG
 45 CTGTTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTC
 TGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCCGGCG
 AGAAGAAGAATGGCCTGTTTCGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCC
 AACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGA
 CACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCCGACCAGTACGCCG
 50 ACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA

GAGTGAACACCGAGATCACCAAGGCCCCCTGAGCGCCTCTATGATCAAGAGATAC
GACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCC
TGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACA
TTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAA
5 AAGATGGACGGCACCAGGAAGTCTGCTGTAAGCTGAACAGAGAGGACCTGCTGC
GGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAG
CTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCTGAAGGACAACCG
GGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGGCCCTCTGG
CCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCAC
10 CCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCG
AGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCAC
AGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGT
GACCGAGGGAATGAGAAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATC
GTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGG
15 ACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGAT
CGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAA
GGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCC
TGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCAC
CTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGG
20 GCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGAC
AATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGAT
CCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCC
AGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAG
AAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCC
25 GGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCA
GAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGGCATCAA
AGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAG
AACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCA
GGAACCTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCTCAGA
30 GCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAAC
CGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTGAAGAAGATGAAGAACT
ACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGA
GACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCC
35 CGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGA
TCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAG
TGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCCGTCTGTG
GGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGA
CTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGC
40 AAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAG
ATTACCTTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGA
AACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGC
TGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTC
AGCAAAGAGTCTATCCgGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGA
45 AGGACTGGGACCCTAAGAAAGTACGGCGGCTTCGtCAGCCCCACCGTGGCCTATTCTG
TGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGAACTGAAGAGTGTGAA
AGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCG
ACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTG
CCTAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGC
50 CcGCtttCTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACCTTCTG
TACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAA

ACAGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCA
GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCC
GCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCC
ACCTGTTTACCCTGACCAATCTGGGAGCCCCCtggGCCTTCAAGTACTTTGACACCAC
5 CATCGACCGGAAGgtGTACcggAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCA
CCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTG
ACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAA
GAGGAAAGTCTAACC GGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATC
AGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCT
10 TCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTCTCTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGC
TCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAG
CTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCT
15 CACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTAGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG
GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGT
TTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCTTC
GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAACGGTTATCCACAGAA
20 TCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGA
ACCGTAAAAAGGCCGCGTGTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGC
ATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
ATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
25 CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCGTTTACGCCCAGCGCTGCGCCTTATCCGGTAAGTATCGTCTTGA
GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAAC
TACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
30 TTCGGA AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG
TGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAG
ATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAACCTCACGTTAAG
GGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTTAA
AATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACC
35 AATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTTCATCCATAG
TTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCC
CCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCA
ATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAAGTGGTCCTGCAACTTTATCCGC
CTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTTCGCCAGTTAA
40 TAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTT
TGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCC
CATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTAA
GTTGGCCGCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT
CATGCCATCCGTAAGATGCTTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTG
45 AGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATA
CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGG
CGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGT
GCACCCAACCTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA
50 TACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA

CATTTCCTCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGAT
 CCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT
 CTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCT
 ACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGGTTAGGCG
 5 TTTTGCGCTGCTTCGCGATGTACGGGGCCAGATATACGCGTTGACATTGATTATTGAC
 TAGTTATTAATAGTAATCAATTACGGGGGTCATTAGTTTCATAGCCCATATATGGAGTT
 CCGCGTTACATAACTTACGGTAAATGGCCCCGCCTGGCTGACCGCCCCAACGACCCCC
 GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
 ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
 10 GTGTATC

SEQ ID NO: 2 (pCMV-iABEmaxNG)

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCCTG
 GCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT
 15 ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTG
 ATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG
 TTTGTTTTTGGCACCAAAATCAACGGGACTTTCCAATAATGTCGTAACAACTCCGCCCC
 ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTG
 GTTTAGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTAT
 20 AGGGAGAGCCGccaccatggctagcATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTC
 ACCAAAGAAGAAGCGGAAAGTCGGATCCTCTGAGGTGGAGTTTTCCACGAGTACT
 GGATGAGACATGCCCTGACCCTGGCCAAGAGGGGCACGCGATGAGAGGGAGGTGCC
 TGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGA
 GCCATCGGCCTGCACGACCCAACAGgCCATGCCGAAATTATGGCCCTGAGgCAGGG
 25 CGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGgGACATTGCA
 GCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTT
 TGGCGTGAGGAACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACT
 ACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGT
 GCCGCCCTGCTGTGCTATTTCTTTCCGATGCCTAGACAGGTGTTCAATGCTCAGAAG
 30 AAGGCCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGA
 GACACCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGG
 GGGTCAGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTG
 GGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCA
 ACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAGC
 35 GGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCA
 GACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAG
 GTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCTTCTGCTGGTGGAAAGAGGATAA
 GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACG
 AGAAGTACCCCAACATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
 40 GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGCGGGGCC
 ACTTCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTC
 ATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAG
 CGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG
 AAAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCTG
 45 ATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGA
 GGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGC
 TGGCCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCC
 GACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCC
 CCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGC
 50 TGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC
 CAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGT

TCTACAAGTTCATCAAGCCCATCCTGGAAAAAGATGGACGGCACCGAGGAACTGCTC
 GTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCA
 GCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGGCGGCAGGAA
 GATTTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTT
 5 CCGCATCCCCTACTACGTGGGCCCCTCTGGCCAGGGGAAACAGCAGATTTCGCCTGGA
 TGACCAGAAAAGAGCGAGGAAACCATCACCCCTGGAACTTCGAGGAAGTGGTGGGA
 CAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACC
 TGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
 TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAAGCCCGCCTT
 10 CCTGAGCGGCGAGCAGAAAAAGGCCATCGTGACCTGCTGTTCAAGACCAACCGGA
 AAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGAC
 TCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGACACATACCA
 CGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG
 ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATC
 15 GAGGAACGGCTGAAAACCTATGCCACCTGTTCGACGACAAAGTGATGAAGCAGCT
 GAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGC
 ATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTT
 CGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGG
 ACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC
 20 AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGT
 GGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCCGAGAACATCGTGATCGAA
 ATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
 ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAAC
 ACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAG
 25 AATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
 CGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACA
 AGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGA
 AGAGGTCTGTGAAGAAGATGAAGAATACTGGCGGCAGCTGCTGAACGCCAAGCTG
 ATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCG
 30 AACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAACCCGGCAGATCACA
 AAGCACGTGGCACAGATCCTGGACTCCC GGATGAACACTAAGTACGACGAGAATGA
 CAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATT
 TCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCC
 CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAA
 35 GCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGA
 TCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
 AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAA
 GCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGC
 CGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAA
 40 AAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC_gGCCCAAGAGGA
 ACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGG
 CTTCG_gCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGG
 CAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA
 AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGA
 45 AGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAA
 ACGGCCGGAAGAGAATGCTGGCCTCTGCC_gGC_{ttt}CTGCAGAAGGGAAACGAACTGGC
 CCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAA
 GGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACT
 ACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCC
 50 GACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCCAT
 CAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCC

CTcggGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGgtGTACcggAGCACCAAA
GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACG
GATCGACCTGTCTCAGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACC GCCGACG
GCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAACC GGTCATCATCACCATC
5 ACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCAT
CTGTTGTTTGCCCCCTCCCCCGTGCTTCCTTGACCCTGGAAGGTGCCACTCCC ACTGT
CCTTTCCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTAT
TCTGGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTG
10 GGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTT
TCCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATAACGAGCCGGAAGCAT
AAAGTGTAAGCCTAGGGTGCTAATGAGTGAGCTAACTCACATTAATTGCGTTGC
GCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA
15 CTGACTCGCTGCGCTCGGTCTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
GCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAG
CAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTC
CATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTG
GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
20 TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTC
GGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGT
CGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGC
CTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT
GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA
25 GAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCG
CAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTC
AGTGGAACGAAAACCTACGTTAAGGGATTTTGGTTCATGAGATTATCAAAAAGGATC
30 TTCACCTAGATCCTTTTAAATTA AAAATGAAGTTTTTAAATCAATCTAAAGTATATAT
GAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCG
ATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGA
TACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGC
TCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
35 AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC
TAGAGTAAGTAGTTTCGCCAGTTAATAGTTTGC GCAACGTTGTTGCCATTGCTACAGG
CATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCG
GTCTCCGATCGTTGTCAGAAGTAAGTTGGCCG CAGTGTTATCACTCATGGTTATGG
40 CAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTG
GTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTT
GCCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTC
ATCATTGAAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAG
ATCCAGTTCGATGTAACCCACTCGTGCACCCA ACTGATCTTCAGCATCTTTTACTTTC
45 ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAA
TAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAA
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGAC
GGATCGGGAGATCGATCTCCCGATCCCTAGGGTCGACTCTCAGTACAATCTGCTCT
50 GATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAG
TAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCAT

GAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGGCCAGATA
 TACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTTCATT
 AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCC
 TGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCAT
 5 AGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAA
 CTGCCCCACTTGGCAGTACATCAAGTGTATC

SEQ ID NO: 3 (pAAV-meCMV-ABEmax-N)

aaagtagccgaagatgacgggtttgacatggagttggcaggatggttgattaaaaacataacaggaagaaaaatgccccgctg
 10 tgggcgacaaaatagttgggaactgggaggggtggaaatggagttttaaggattatttagggaagagtgacaaaatagatgggaactg
 ggtgtagcgtcgtagtaagtaatacgaataaataaataagacaaaatagtttgaactagatttcacttatctggttcggatctcctaggtcaagca
 gtgatcagatccagacatgataagatacattgatgagtttgacaaaccacaactagaatgcagtgaaaaaatgctttatttgtgaaatttgtg
 atgctattgctttatttgttaaccattataagctgcaataaacaagttaacaacaacaattccatgtcctgcaggcagctgcgcgctcgctcgctc
 actgagggccgccccggcgctggggcgacctttggtcgccccggcctcagtgagcgcgagcgcgcagagagggagtgcccaactccat
 15 cactaggggttcctgcgccctctagactcgagCGCGTGATGAGAGCAGCCACTACGGGTCTAGGCTGC
 CCATGTAAGGAGGCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATTAACCCA
 GACATGTGGCTGCCCCCCCCCCCCCAACACCTGCTGCCTGCTAAAAATAACCTGTC
 CCTGGTGccctgcatgccACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT
 GGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCGTAACAACCTCC
 20 GCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
 AGCTGGTTTAGTGAACCGTCAGATCcgccaccATGgctagcATGAAACGGACAGCCGACG
 GAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGC
 CACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAA
 GAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACAACAATAGAGTGATCGGAGAGGG
 25 ATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCACACGCAGAGATCATGGCAC
 TGAGGCAGGGAGGCCTGGTTCATGCAGAATTACCGCCTGATCGATGCCACCCTGTAT
 GTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGG
 AAGAGTGGTGTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATG
 GATGTGCTGCACCACCCCGGCATGAACCACCGGGTGGAGATCACAGAGGGAATCCT
 30 GGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGA
 TCAAGGCCCCAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGGAGG
 ATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCCGCCACACCAGAGAGCTCCG
 GCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGTTTTCCACGAGTACTGGATGAGA
 CATGCCCTGACCCTGGCCAAGAGGGGCACGCGATGAGAGGGAGGTGCCTGTGGGAG
 35 CCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGC
 CTGCACGACCCAACAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGT
 CATGCAGAACTACAGACTGATTGACGCCACCCTGTACGTGACATTCGAGCCTTGCGT
 GATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGTGTTGGCGTGA
 GGAACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACTACCCCGGC
 40 ATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGTGCCGCCCT
 GCTGTGCTATTTCTTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGGCCCA
 GAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTG
 GCACAAGCGAGAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGGGGGTGAGA
 CAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGA
 45 TCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGAC
 CGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAGCGGCGAAAC
 AGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAG
 AACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGA
 CAGCTTCTTCCACAGACTGGAAGAGTCTTTCCTGGTGGAAAGAGGATAAGAAGCACG
 50 AGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTAC
 CCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGGCCGACCT

GCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGGCCACTTCCTGAT
 CGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATCCAGCTGG
 TGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCGGGCTGGAC
 GCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGAT
 5 CGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCTGATTGCCCTGA
 GCCTGGGCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAA
 CTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGAT
 CGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCT
 GCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAGGGCCCCCTGAGCGCCT
 10 CTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTC
 GTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
 CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCA
 TCAAGCCCATCCTGGAAAAGATGGACGGCACCCGAGGAACTGCTCGTGAAGCTGAAC
 AGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCACCA
 15 GATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCAT
 TCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTAC
 TACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTTCGCTGGATGACCAGAAAGAG
 CGAGGAAACCATCACCCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCTTCCG
 CCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAG
 20 GTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGAC
 CAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAG
 CAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAA
 GCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGTCTCAGTTATGA_cACCGAAA
 TCCTGACAGTCGAGTATGGAT_tCTGCCGATCGGCAAGATTGTGGAG_gAGAGAATTG
 25 AATGTACGGTCTATA_cGTGACAA_gAATGGT_tCgTCTACACCCAACCA_aTTGCTCAA
 TGGCATA_aTCGAGGGGAGCAGGAGGTGTTT_gGAGTATTGCCTGGAGGACGGGTCA_aTC
 ATTAGAGCTACAAAGGACCATAAGTTTATGACA_acCGATGGTCAAATGCTGCCGATA
 GATGAAATATTTCGAAAGGG_gACTGGATCTTA_aGC_aAGTCGAT_ggCCTTCCA_acTAgTAg
 aatlcctagagctcgtgatcagcctcgtgctgcttctagtgtccagccatctgtgttgcctcccccgtgcttcttgacctggaag
 30 gtgccactcccactgtcctttcctaataaaatgaggaaattgcattgcattgtctgagtaggtgtcattctattctggggggtggggtggggca
 ggacagcaagggggaggattgggaagagaatagcaggcatgtctggggaggtaccgagggcctatttccatgattccttcatttgc
 atacgatacaaggctgttagagagataattggaattaatttgactglaaacacaaagatattagtaaaaaacgtgacgtagaaagtaataat
 ttcttgggtagtgttcagttttaaattatgttttaaattggactatcatatgcttaccgttaactgaaagtatttcgatttcttggctttatatacttG
 TGGAAAGGACGAAACACCGGTTATCTCCTGTTCTGCAGCGTTT_cAGAGCTA_tgtctgGAA
 35 AcagcaTAGCAAGTT_gAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA
 GTCGGTGCTTTTTT_gcggccgcaggaaccctagtgtgatggagtggccactccctctctgcgcgtcgtcgtcactgagggc
 cgggcgaccaaaggtcgcgcgacgcccgggcttgcgcggggcgccctcagtgcgcgagcgcgcgcagctgcctgcagggggcgcc
 tgtaccgggagatgggggaggctaactgaaacacggaaggagacaataaccggaaggaaaccgcgtatgacgggcaataaaaagacag
 aataaaacgcacgggtgttgggtcgtttgttcataaacgcggggttcgggtcccagggctggcactctgtcgataccccacagagacccatt
 40 gggaccaataacgcccgcgtttcttcttttccccacccaacccccaaagttcgggtgaaggcccagggtcgcagccaacgtcggggcg
 gcaagccctgcatagccactacgggtacgttaggccaaccactagaactatagctagagtcctggggaacaaacgatgtcgccttcca
 gaaaaccgaggatgcgaaccacitcatccgggtcagcaccacgggaagcgcgcgacggccgaggtctaccgatctcctgaagcca
 gggcagatccgtgcacagcaccttgcgtagaagaacagcaaggccgccaatgcctgacgatgcgtggagaccgaaaccttgcgctcg
 ttcgccagccaggacagaaatgcctcgaacttcgtgctgcccaggttgcgggtgacgcacaccgtggaacggatgaaggcacgaa
 45 cccagttgacataagcctgttcgggtcgtaaactgtaactgcaatgacgtatgcgtcagcaactggtccagaaccttgaccgaacgcagc
 ggtgtgaacggcgagtgggcggtttcatggctgttatgactgtttttgtacagtctatgcctcgggcatccaagcagcaagegcgttacg
 ccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcaggaacgatgttacgcagcagggcagtcgccctaaaacaaagt
 taggtggctcaagtatgggcatcattgcacatgttaggtcgcgcctgaccaagtcaaatccatgcgggctgctcttgatctttcggctcgtg
 agttcggagacgtagccacctactccaacatcagccgactccgattacctcgggaactgtcctcgtagtaagacattcatcgcgctgct
 50 gccttcgaccaagaagegggtgttggcgctctcgcggcttacgttctgccaggttgagcagccgcgtagtagatctatactatgatctc
 gcagctcggcgagcaccggaggcagggttcaccaccgcgtcatcaatctcctcaagcatgaggccaacgcgcttggtgcttatgtg

atctacgtgcaagcagattacgggtgacgatcccgagtggtctctatatacaaagtgggcatacgggaagaagtgatgcactttgataatga
 cccaagtaccgcccacctaacaattcggtcaagccgagatcggtctcccgccgaggagttgttcggttaaattgtcacaacgccgcgaatat
 agtctttaccatgcccttggccacgcccctctttaatacagacgggcaatttgcacttcagaaaaatgaagagtttgccttagccataacaaaagtc
 cagtatgctttttcacagcataactggactgatttcagtttacaactattctgtctagtttaagactttattgtcatagtttagatctattttgttcagttt
 5 aagactttattgtccgcccacacccgcttacgcaggcatccatttattactcaaccgtaaccgattttgccagggttacgcgggctggtctgcgg
 tgtgaaataccgcacagatgctgaaggagaaaaataccgcacaggcgctcttccgcttcctcgctcactgactcgctgcgctcggtctgctg
 gctgcggcgagcggtatcagctcactcaaaaggcggttaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaa
 aaggccagcaaaaggccaggaaccgtaaaaaaggccgctgtgctggcgttttccataggtcgcgccccctgacgagcatcacaataat
 cgacgctcaagtgcagaggtggcgaaacccgacaggactataaagataccaggcggttccccctggaagctccctcgctgcgctcctctgttc
 10 cgacctgcccgttaccggatacctgtccgcttctcccttcgggaagcggtggcgcttctcaatgctcagctgtaggtatctcagttcggt
 gtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgcttatccggttaactatcgtcttgagttcca
 acccggttaagacacgacttatcgccactggcagcagccactggtaacaggatagcagagcgagggtatgtagggcggtgctacagagttct
 tgaagtgggtggcctaactacgggtacactagaaggacagtatttgggtatctgcgctcgtctgaagccagttaccttcggaaaaagagtggta
 gctcttgatccggcaaaacaaaccaccgctggtagcggtgtgtttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaa
 15 gatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcagtttaagggttttggctcatgagattatcaaaaaggatcttcac
 ctgatctcttttaataaaaaatgaagtfttaataaactaaagtataatagtaaaacttggctgacagttaccaatgcttaatacagtgaggca
 cctatctcagcgatctgtctatttctgttcacatagttgctgactccccgctggtgtagataactacgatacgggaggggcttaccatctggccc
 cagtgtctgcaatgataccgcgagaccacgctcaccgggtccagatttatcagcaataaaccagccagccggaaggggccgagcgcaga
 agtggtcctgcaactttatccgctccatccagcttattaattgttgcggggaagctagagtaagtagttcgccagtttaagtttgcgcaacgt
 20 tgttgccattgctacaggcatcggtgtgtcacgctcgtcgttttggtatggcttcattcagctccgggtcccaacgatcaaggcgagttacatgat
 ccccatgttgtgcaaaaaagcggttagctccttcggctcctccgacgttgcagaaagtaagttggccgagtggttatcactcatggttatggc
 agcactgcataattcttactgtcatgccatccgtaagatgctttctgtgactggtagtactcaaccaagtcattctgagaatagtgatgcg
 gcgaccgagttgctcttgcggcgctcaatacgggataataccgcgcccacatagcagaactttaaaagtgtctcatcttggaaaaacgttcttc
 ggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaaccactcgtgcacccaactgatcttcagcatcttttactttca
 25 ccagcggttctgggtgagcaaaaaacaggaaaggcgaataaaggcgacacgggaaatgttgaaatactcactatct
 tctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaataagggttccg
 cgcacatttccccgaaaagtgccacctgaaattgtaaacgttaattttgttaaaatccggttaattttgttaaatcagctcatttttaaccaa
 taggcccgaatcggaataatccctataatcaaaagaatagaccgagataggggttagtggttccagtttggacaagaagagtcactatta
 aagaacgtggactccaacgtcaaaaggcggaataaaccgctctatcaggggcgatggccactacgtgaacatcacctaataagtttttgg
 30 ggtcgaggtgccgtaaaagcactaaatcggaacccataaaggagccccgatttagagcttgacggggaaagccggcgaaacgtggcgag
 aaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcgggtcacgctgcgcgtaaccaccacacccgccc
 cgcttaatgcgcgctacagggcgctccatcgcacattcagggtgcaataagcggtgatatcagttcaattacaacattataaacgaag
 agatgacagaaaaatttcttctgtgacagagaa

35

SEQ ID NO: 4 (pAAV-ABEmaxNG-C)

aaagtagccgaagatgacgggtgtcacatggaggtggcaggatgtttgattaaaaacataacaggaagaaaaatgccccgctg
 tggcgcgacaaaatagttgggaactgggaggggtggaaatggagttttaaggattatttagggaagagtgacaaaatagatgggaactg
 ggtgtagcgtcgtaagctaatacgaataaataaagacaaaatagtttggaaactagatttcattatctggttcggatctcctagggtcaagca
 40 gtgatcagatccagacatgataagatacattgatgagtttggacaacacacactagaatgcagtgaaaaaaatgctttatttgtgaaattgtg
 atgctattgctttatttgaaccattataagctgcaataaacaagtaacaacaacaattccatgtcctgcaggcagctgcgcgctcgtcgtc
 actgaggcgcccgggcgctcggcgacatttggctcggcgccctcagtgagcgagcgagcgcgagagagggagtgcccaactccat
 cactagggggttcctgcggcctctagactcgagCGCGTGATGAGAGCAGCCACTACGGGTCTAGGCTGC
 CCATGTAAGGAGGCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATTAACCCA
 45 GACATGTGGCTGCCCCCCCCCCCCCAACACCTGCTGCCTGCTAAAAATAACCCTGTC
 CCTGGTGGccctgcatgccACTCACGGGGATTTCGAAGTCTCCACCCCATTTGACGTCAAT
 GGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCGTAACAACCTCC
 GCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
 AGCTGGTTTAGTGAACCGTCAGATCcgccaccATGgTcAAgATTatCAgcCGCAAAATcCTTG
 50 GGGAcACAGAATGTATATGACATCGGCGTGGAaAaGGATCACAATTTTctgCTGAAGA
 ATGGTcTTgTTGCTTccAAfTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGG

TTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGA
 CTTCTGAGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
 CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTG
 TTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCA
 5 GGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAAT
 CCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCA
 CGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGG
 GCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAG
 GGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGC
 10 ACAAGCCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAA
 GGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGA
 GCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAAC
 GAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
 ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTT
 15 TCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGG
 GGCAAGAGCGACAACGTGCCCTCCGAAGAGGTTCGTGAAGAAGATGAAGAATACT
 GGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACC
 AAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGAC
 AGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGG
 20 ATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCA
 CCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGC
 GCGAGATCAACAATACTACCACACGCCACGACGCCTACCTGAACGCCGTCTGTGGGA
 ACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTA
 CAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAG
 25 GCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATT
 ACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAA
 CCGGGGAGATCGTGTGGGATAAAGGGCCGGGATTTTGGCACCGTGGCGAAAGTGCTG
 AGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCA
 GCAAAGAGTCTATCCgGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAA
 30 GGACTGGGACCCTAAGAAGTACGGCGGCTTCGtCAGCCCCACCGTGGCCTATTCTGT
 GCTGGTGGTGGCCAAAGTGGAAGAGGGCAAGTCCAAGAACTGAAGAGTGTGAAA
 GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGA
 CTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGC
 CTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCc
 35 GCtttCTGCAGAAGGGAAACGAACCTGGCCCTGCCCTCCAAATATGTGAACCTTCTGT
 CCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAAC
 AGCTGTTTGTGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGC
 GAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGC
 CTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACC
 40 TGTTTACCCTGACCAATCTGGGAGCCCCCTcggGCCTTCAAGTACTTTGACACCACCAT
 CGACCGGAAGgtGTACcggAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCA
 GAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACT
 CTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCCCAAGAAGAAGAG
 GAAAGTCTaacTAgtAgaattcctagagctcgtgatcagcctcgaactgtgcectctagtgtccagccatctgtgtgttgccectccc
 45 cegtgccttcttgacctggaagggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcatgtctgagttaggtgtcattct
 attctgggggggtgggggtggggcaggacagcaagggggaggattgggaagagaatagcaggcatgctggggagggtaccgagggccta
 ttcccatgattccttcatatttgcataatagatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaa
 atacgtgacgtagaagtaataatttctgggttagtttcagttttaaattatgttttaaattggactatcatatgcttaccgtaacttgaaagtatt
 tcgatttcttggtttatatatcttGTGGAAAGGACGAAACACCGGTTATCTCCTGTTCTGCAGCGTT
 50 TcAGAGCTAtgctgGAAAcagcaTAGCAAGTTgAAATAAGGCTAGTCCGTTATCAACTTGA
 AAAAGTGGCACCGAGTCGGTGCTTTTTTgcggecgaggaacccctagtgtgatggagtggccactccctctctg

cgcgctcgtcgtcactgagggccggcgaccaaagggtcgcccgacgcccgggctttgcccggcgccctcagtgagcgagcgagcg
 cgcagctgcctgcagggggcgctgtaccgggagatgggggagggctaactgaaacacggaaggagacaataccggaaggaaacccgcg
 ctatgacggcaataaaaagacagaataaaacgcacgggtgttgggtcgtttgtcataaacgcggggttcggtccagggttgccactctg
 tcgataccccaccgagaccccatfgggaccaatacggcggtttcttctttcccaaccccaagttcgggtgaaggcccagg
 5 gctcgcagccaacgtcggggcggaagccctgccatagccactacgggtacgttaggccaaccactagaactatagctagagtcttggc
 gaacaaacgatgctcgccttcagaaaaaccgaggtgcgaaccacttcatccgggggtcagcaccaccgggaagcgccgcgacggccg
 aggtctaccgatctcctgaagccaggggcagatccgtgcacagcaccttgcgtagaagaacagcaaggccgccaatgcctgacgatgcg
 tggagaccgaaacctgtcgtcgttcgccagccaggacagaaatgcctcgacttcgctgctgcccaaggttgcgggtgacgcacacct
 ggaaacggatgaaggcagcaacccagttgacataagcctgttcggttcgtaaactgtaatgcaagtagcgtatgcgtcacgcaactggc
 10 cagaaccttgaccgaacgcagcgggtgtaacggcgagtgccggtttcatggcttgtatgactgtttttgtacagtctatgcctcgggca
 tccaagcagcaagcgcttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagcaacgatgttacgcagcag
 ggcagtcgccctaaaaaaagttaggtggctcaagtatgggcatcttcgcacatgtaggctcggccctgaccaagtc aaatccatgcggg
 ctgctcttgatctttcggctgtgagttcggagacgtagccacctactcccaacatcagccggactccgattacctcgggaacttgcctcgtag
 taagacattcatcgccttgccttcgaccaagaagcgggtgttggcgctctcgcggcttacgttctgccaggtttgagcagccgcgta
 15 gtgagatctatctatgatctcgcagctcggcgagcaccggaggcagggcattgccaccgcgctcatcaatctctcaagcatgaggc
 caacgcgcttgggtcttatgtgatctacgtgcaagcagattacgggtgacgatcccgagtggtctctatacaaaagttgggcatacgggaag
 aagtgatgcactttgatacgaccaagttaccgccacctaaacatcggtcaagccgagatcggttcccgccgagggttgcgtgtaaa
 ttgtcacaacgcccgcaatatagtctttaccatgcccttggccacgccccctttaaatacagcgggcaatttgcacttcagaaaaatgaagagt
 tgccttagccataaaaaagtcagatgcttttccacagcataactggactgatttcagtttacaactatttctgtctagttaaagactttattgcat
 20 agtttagatctattttgtcagtttaagactttattgtccgcccacaccgcttacgcagggcacatccatttattactcaaccgtaaccgattttgcca
 gggtacgcgggtggtctgcgggtgtgaaataccgcacagatgcgtaaggagaaaaataccgcatcaggcgctcttcgcttctcgtcactg
 actcgtcgcgctcggctgttcggctcggcgagcgggtatcagctcactcaaaaggcggttaatacgggttatccacagaatcaggggataacg
 caggaaagaacatgtgagcaaaaaggccagcaaaaaggccaggaaccgtaaaaaggccggttgcgtggcgttttccataggctccgcccc
 cctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcggttccccctggaag
 25 ctccccctgtgcgctcctcgttccgacctgccgcttaccggataacctgtccgctttctccttcgggaagcgtggcgctttctcaatgctcac
 gctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagccccaccgctgcgcttacc
 ggtaactatcgtcttgagccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatg
 tagggcggtgctacagagttctgaagtgtggcctaactacggctacactagaaggacagtatttggtatctgcgctcgtgtaagccagtta
 ccttcggaaaaagagttggtagctctgatccggcaaaacaaaccaccgctggtagcggtgtttttgtttgcaagcagcagattacgcgca
 30 gaaaaaaaggatctcaagaagatcctttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggtattttggtcatga
 gattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagtttaataaactaaagtatatatgagtaaaacttggctgacagttac
 caatgcctaatacagtgaggcacctatctcagcagatcgtctatcttcgttcatccatagttgctgactccccgctgtgtagataactacgatac
 ggagggttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaacaggccagc
 cggaaggccgagcgcagaagtgtcgtcgaactttatccgctccatccagctctattaattgttgccgggaagctagagtaagtagtgcg
 35 cagtttaatagtttgcgaacgttgttccattgctacaggcatcgtgtgtcacgctcgtcgtttggtatggcttcattcagctcgggttccaa
 cgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcgggttagctccttcggctcccgatcgttgcagaagtaagttggccgca
 gtgttatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgctttctgtgactggtgagttacaaacga
 cattctgagaatagtgtatgcggcgaccgagttgctcttgcggcgctcaatacgggataataccgcgccacatagcagaactttaaagt
 ctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcagatgaaccactcgtgcacccaact
 40 gatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacagggaaggcaaaatgccgcaaaaaagggaataaggcgacacg
 gaaatgttgaatactcactcttcttttcaatatttgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaa
 aataaacaataagggttccgcgcacatttccccgaaaagtgcacctgaaattgtaaacgttaattttgttaaaatcgcgttaaattttgtt
 aaatcagctcatttttaaccaataggccgaaaatcgcaaaaatcccttataaatcaaaagaatagaccgagatagggttgagttgttccagt
 ttggaacaagagtcactattaaagaacgtggactccaacgtcaaaggcgcaaaaacgctctatcaggcgatggccactacgtgaacc
 45 ataccctaataagtttttggggtcaggggtccgtaaaagcactaaatcggaacctaaaggagccccgatttagagcttgacgggga
 aagccggcgaaacgtggcgagaaagggaagggaagaaagcgaaaggagcggcgctagggcgctggcaagtgtagcggtcacgctgc
 gcgtaacaccacaccgcgcgcttaatgcgcgctacagggcgctccattcgcattcaggctgcaataaagcgttgatattcagtc
 aattacaacattaataacgaagagatgacagaaaaatttcttctgtgacagagaa

50

SEQ ID NO: 5 (pX601-meCMV-ABEmax-N2)

cctgcaggcagctgcgcgctcgctcgctcactgagggccggcggtcgggcgaccttgggtcgccggcctcagtgagc
gagcgagcgcgcagagaggggagtgccaaactccatcactaggggttcctgcggcctctagactcgagCGCGTGATGAGAG
CAGCCACTACGGGTCTAGGCTGCCCATGTAAAGGAGGCAAGGCCTGGGGACACCCGA
5 GATGCCTGGTTATAATTAACCCAGACATGTGGCTGCCCCCCCCCCCCCAACACCTGC
TGCTTGCTAAAAATAACCCTGTCCCTGGTGGccctgcatgccACTCACGGGGATTTCCAA
GTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT
TTCCAAAATGTTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTA
CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTGTAGTGAACCGTCAGATCcgccaccATGg
10 ctagcATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGA
AAGTCTCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCACGCACTGACCCTG
GCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACA
ACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTAC
CGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGGCCTGGTTCATGCAGAATTACC
15 GCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGA
GCAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGA
CCGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATGAACCACCGG
GTGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTT
CTTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACC
20 GACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACAAGCG
AGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAG
TTTTCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGA
TGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGTCTGAACAATAGAGTGATCGGCG
AGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATG
25 GCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCT
GTACGTGACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGAT
CGGCCGCGTGGTGTCTTGGCGTGAGGAACGCAAAAACCGGCGCCGCAGGCTCCCTGA
TGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATC
CTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTTTCGGATGCCTAGACAGGTG
30 TTCAATGCTCAGAAGAAGGCCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGG
CTCCTCTGGCTCTGAGACACCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGCG
GGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGCCTGGCCATCGGCAC
CAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAAT
TCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGC
35 CCTGCTGTTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCA
GAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGC
AACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCT
GGTGGAAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGAC
GAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGT
40 GGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCATATGA
TCAAGTTCGGGGGCCACTTCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAA
CCCCATCAACGCCAGCGGCGTGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCA
AGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGG
45 CCTGTTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCA
ACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCTACGACGAC
GACCTGGACAACCTGCTGGCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCG
AGATACCAAGGCCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCAC
50 CAGGACCTGACCCTGCTGAAAGCTCTCGTGCAGCAGCTGCCTGAGAAGTACAA
AGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAG

91

gggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatTTTTgtatgctcgtcagggggcgaggaccta
tggaaaaacgccagcaacgcggcctttttacgggttctggccttttctggccttttctcacatgt

SEQ ID NO: 6 (pX601-meCMV-ABEmaxC2NG)

5 cctgcaggcagctgcgcgtcgcctcactgagggcgcccgggcgacetttggtcgccggcctcagtgagc
gagcgagcgcgcagagagggagtgcccaactccatcactaggggttctgcggcctctagactcgagCGCGTGATGAGAG
CAGCCACTACGGGTCTAGGCTGCCCATGTAAAGGAGGCAAGGCCTGGGGACACCCGA
GATGCCTGGTTATAATTAACCCAGACATGTGGCTGCCCCCCCCCCCCCAACACCTGC
10 TGCTGCTAAAAATAACCCTGTCCCTGGTGGcctgcagcccACTCACGGGGATTTCCTAA
GTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACT
TTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTA
CGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCgctagccaccAT
GATGCTCAAGAAGATCCTCAAGATTGAAGAGTTGGACGAGCGCGAGCTTATAGACA
TAGAAGTCAGTGGTAATCACCTTTTCTACGCAAATGACATTTTGACTCACAACCTCCT
15 CTTCAGACGTTTGTCTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACG
CCTCCCTGGGCACATAACCAGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTG
GACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTT
TGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACG
ACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAG
20 CCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATT
TCCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAG
CCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCC
TGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCC
25 CGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAG
AAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCA
GCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCT
GTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACA
TCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCTCAGAGCTTTCTGAAGG
30 ACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAG
CGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAGTACTGGCGGCAG
CTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG
GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACAC
35 TAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGT
CCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTACAAAGTGC GCGAGATCA
ACAACCTACCACACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTG
ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTA
CGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCC
40 AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGA
TCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCC
CAAGTGAAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGT
CTATCCgGCCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGAC
45 CCTAAGAAGTACGGCGGCTTCGtCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTG
GCCAAAGTGGAAGAGGGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGG
GGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAA
GCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTC
CCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCcGCtttCTGCAG
50 AAGGGAAACGAACCTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAG
CCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTG

TGGAACAGCACAAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCC
 AAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCTACAACAA
 GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCC
 TGACCAATCTGGGAGCCCCCtggGCCTTCAAGTACTTTGACACCACCATCGACCGGA
 5 AGgtGTACcggAGCACCAAGAGGGTGCTGGACGCCACCCTGATCCACCAGAGCATCA
 CCGGCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCTGGCGGC
 TCAAAAAGAACC GCGACGGCAGCGAATTTCGAGCCCAAGAAGAAGAGGAAAGTCTa
 acTAgtAgaattcctagagctcgtgatcagcctcgactgtgccttctagttgccagccatctgtgtttgccccctccccctgccttcttg
 acccttgaaggtgccactcccactgtcccttcttaataaaatgaggaaattgcatcgcatgtctgagtaggtgtcattctattctgggggtg
 10 ggggtggggcaggacagcaagggggaggattgggaagagaatagcaggcatgctggggagggtaccgagggcctatttcccatgattcct
 tcatatttgcatacagatacaaggctgttagagagataattggaattatgtgactgtaaacacaaagataattagtaaaaaatcgtgacgtag
 aaagtaataatttctgggtagtgttcagttttaaataattatgtttaaataaggactatcatatgcttaccgtaacttgaaagtatttcgatttcttgctt
 tatatatcttGTGGAAGGACGAAACACCGGTTATCTCCTGTTCTGCAGCGTTTcAGAGCTA
 tgctgGAAAcagcaTAGCAAGTTgAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
 15 CACCGAGTCGGTGCTTTTTTgeggccgcaggaaacccctagtgtgaggttggccactccctctctgcgcgtcgtcgc
 tctactgaggccggggcgaccaaaggctgccccgacgccccgggctttgccccggcgccctcagtgcgcgagcgcgcgcgcgtgcctgc
 agggggcgctgatgcggtatttctccttacgcctctgtgcggtatttcacaccgcatacgtcaaaagcaaccatagtagcgcgcctgtagcgg
 cgcattaagcgcggcggtgtgtgtgtacgcgcagcgtgaccgctacacttgcagcgccttagcgcgcctccttgccttcttccctt
 ccttctcgcacgttgcgggcttccccgtcaagctctaaatcgggggctccctttaggggttccgatttagtgccttaccggcacctcgaccc
 20 caaaaaacttgatttgggtgatgttcacgtagtgggcatcgccctgatagacgggttttgccttgcagttggagtcacgttctttaaata
 gtggactcttgtccaaacttgaacaacactcaaccctatctcgggctatttctttagttataagggaatttgcgatttgcgctatttggtaaa
 aaatgagctgatttaacaaaaatataacgcgaattttaaacaataatataacgtttacaattttatgggtcactctcagtacaactctgctctgatgcc
 gcatagttaagccagccccgacacccgccaacacccgctgacgcgcctgacgggcttgcctgcctccggcatccgcttacagacaagct
 gtgaccgtctccgggagctgcatgtgtcagagggtttcacccgtcatcaccgaaacgcgcgagacgaaagggcctcgtgatacgcctatttt
 25 atagggtaatgtcatgataataatggttcttagacgtcaggtggcacttttccgggaaatgtgcgcggaacccctatttgttattttctaaatac
 attcaaatatgtatccgctcatgagacaataacccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgc
 gcccttattcccttttttgcggcattttgccttctgttttgcctacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgca
 cgagtgggttacatgaactggatctcaacagcggtaagatccttgagagtttgcggcgaagaacgttttccaatgatgagcacttttaa
 gttctgctatgtggcgcggtatttaccgtattgacgcggggaagagcaactcggctgcgcgatacactattctcagaatgacttgggtgag
 30 tactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgcctgcataacctatgagtataacactgcgg
 ccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatgggggatcatgtaactgccttgatcgttggg
 aaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaaactattaactg
 ggaactacttactctagcttccgggcaacaataatagactggatggaggcggataaagttgcaggaccacttctgcgcctcgcccttccg
 gctggctggttatttgcgtataaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggttaagccctccc
 35 gtatcgtagtattctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtgagataggtgcctcactgattaagcat
 tggtaactgtcagaccaagtttactcatatatacttttagattgattttaaacttcatttttaattttaaaggatctaggtgaagatccttttgataatct
 catgacaaaaatcccttaacgtgagtttccgttccactgagcgtcagaccccgtagaaaagatcaaaagatcttcttgagatcctttttctgcg
 cgtaatctgctgcttgcacaaaaaaacacccgtaccagcgggtggttgggttgcggatcaagagctaccaactcttttccgaaggtaact
 ggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgttagttagccaccacttcaagaactctgtagcaccgcctacata
 40 cctcgtctctgataactcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttgactcaagacgatagtaccggataa
 ggcgcagcggctcgggctgaacggggggttcgtgcacacagcccagcttgagcgaacgacctacaccgaactgagatacctacagcgt
 gagctatgagaaagcgcacgcttcccgaagggaagggcgacaggtatccggtaagcggcagggctgggaacaggagagcgcac
 gagggagcttccagggggaaacgcctgggtatcttatagtcctgtcgggttccgcaacctgacttgagcgtcgaattttgtgatgctcgtca
 gggggggcgagcctatggaaaaacgccagcaacgcggcctttttacgggtccttggccttttgcctgctcactgt

45

SEQ ID NO: 7 (pX601-MHP1-iABEmaxN2)

cctgcaggcagctgcgcgtcgtcgtcactgaggccgcccgggcgacctttggtcgcggcgccctcagtgcgc
 gaggcagcgcgcagagagggagtgccaactccatcactaggggttcctgcggcctctagactcaggcCCTTCAGATTAA
 AAATAACTGAGGTAAGGGCCTGGGTAGGGGAGGTGGTGTGAGACGCTCCTGTCTCT
 50 CCTCTATCTGCCCATCGGCCCTTTGGGGAGGAGGAATGTGCCCAAGGACTAAAAAA
 AGGCCATGGAGCCAGAGGGGCGAGGGCAACAGACCTTTCATGGGCAAACCTTGGG

GCCCTGCTGactgtaGATGAGAGCAGCCACTACGGGTCTAGGCTGCCCATGTAAGGAG
 GCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTG
 CCCCCCCCCCCCCAACACCTGCTGCCTGCTAAAAATAACCCTGTCCCTGGTGgacctgc
 atgCCCTCCCTGGGGACAGCCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATAT
 5 AACCAGGGGCACAGGGGCTGCCctcattetACCACCACCTCCACAGCACAGACAGACA
 CTCAGGAGCCAGCcaccatggctagcATGAAACGGACAGCCGACGGAAGCGAGTTCGAGT
 CACCAAAGAAGAAGCGGAAAGTCGGATCCTCTGAGGTGGAGTTTTTCCCACGAGTAC
 TGGATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGGAGGTGC
 CTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGA
 10 GCCATCGGCCTGCACGACCCAACAGgCCATGCCGAAATTATGGCCCTGAGgCAGGG
 CGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCTGTACGgGACATTCTGA
 GCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTGGTGT
 TGGCGTGAGGAACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACT
 ACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGT
 15 GCGGCCCTGCTGTGCTATTTCTTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAG
 AAGGCCCAGAGCTCCACCGACTCCGGAGGATCTAGCGGAGGCTCCTCTGGCTCTGA
 GACACCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGCGGGGGCAGCAGCGGG
 GGGTCAGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTG
 GGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCA
 20 ACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAGC
 GGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACC GCCAGAAGAAGATACACCA
 GACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAG
 GTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATAA
 GAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACG
 25 AGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAA
 GGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGGCC
 ACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTC
 ATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAG
 CGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG
 30 AAAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCTG
 ATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGA
 GGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGC
 TGGCCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCC
 GACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCC
 35 CCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGC
 TGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGAC
 CAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGT
 TCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTC
 GTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCA
 40 GCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAA
 GATTTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTT
 CCGCATCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGA
 TGACCAGAAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGG
 CAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACC
 45 TGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
 TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGGAATGAGAAAAGCCCGCCTT
 CCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGA
 AAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGTTTGGAT
 CTGAAAACGCAAGTTCAAACGCCACAGGGTATGAAAGAAATATCCAATATACAGGT
 50 CGGCGATCTCGTCTTGTCTAACACTGGCTATAACGAGGTGCTGAATGTATTTCCAAA
 AAGCAAGAAAAAAAGTTACAAGATAACTCTGGAAGATGGAAAAGAAATTATCTGT

TCTGAGGAGCATCTGTTTCCGACCCAAACAGGGGAGATGAATATCAGTGGCGGTCT
 CAAAGAGGGTATGTGTTTGTATGTCAAGGaataactagtagaattcctagagctcgtgatcagcctcactgt
 gccttctagtgtccagccatctgtgttggccctccccctgccttccctgacctggaaggtgccactcccactgtccttccctaataaaatga
 ggaaattgcacgcattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagaga
 5 atagcaggcatgctggggaggtaccgagggcctatttcccatgattccttcatatttgcataacgatacaaggctgttagagagataattgg
 aattaatttgactgtaaacacaaagatattagtagacaaaatacgtgacgtagaaagtaataatttcttgggtagtgttcagttttaaattatgtttta
 aaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatacttGTGGAAAGGACGAAACACCG
 GTTATCTCCTGTTCTGCAGCGTTTcAGAGCTAtgctgGAAAcagcaTAGCAAGTTgAAATA
 AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTgcgccgca
 10 ggaaccctagtgtatggagttggccactccctctctgcgcgctcgtcgtcactgaggccggggcgaccaaaggctgcggcgacgcccgg
 gctttgcccggggcgccctcagtgagcgagcgagcgcgagctgcctgcagggcgccctgatgcggtattttctccttacgcactctgtgcg
 gtatttcacaccgcatacgtcaaaagcaaccatagtagcgccctgtagcggcgcaataagcgcgggcggtgtggtggttacgcgcagcgt
 gaccgctacacttgccagcgccctagcgcccgctccttcttcttcccttcttctcgcacgttcgcccgggttcccccgtcaagctctaa
 atcggggggtcctcttaggggtccgatttagtgccttacggcacctcgacccccaaaaaacttgatttgggtgatggttcacgtagtggccatc
 15 gccctgatagacggttttgcctttagcgttggagtccacgttcttaatagtggactcttgttccaaactggaacaacactcaacctatctc
 gggctattctttgatttataagggtatttggcgatttcggcctatttggtaaaaaatgagctgatttaacaaaaatttaacgcgaatttaacaaaa
 tattaacgtttacaattttatgggtgcactctcagtacaactctgctctgatgccgatagtaagccagccccgacaccgccaacaccgctga
 cgcgcccctgacgggcttgcctgcctcccgccatccgcttacagacaagctgtgaccgctcggggagctgcatgttcagagggtttaccgt
 catcacgaaacgcgcgagacgaaagggcctcgtgatacgcctattttataggtaattgtcatgataataatggtttcttagacgtcaggtgg
 20 cacttttcggggaaatgtgcgcggaacccctatttgtttattttctaataacattcaaatatgtatccgctcatgagacaataaccctgataaatg
 cttaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgccttattcccttttttgcggcatttgccttctgttttgcac
 ccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatc
 cttagagagtttgcggcggaagacgtttccaatgatgagcacttttaagttctgctatgtggcgcggtattatcccgtattgacgcccggca
 agagcaactcggtcgccgcatacactattctcagaatgacttgggtgagtactaccagtcacagaaaagcatcttacggatggcatgacag
 25 taagagaattatgcagtgtcgcataaccatgagtataacactgcggccaacttactctgacaacgatcggaggaccgaaggagctaac
 cgcttttttgcacaacatgggggatcatgtaacicgccttgcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgac
 accagatgcctgtagcaatggcaacaacgttgcgcaaaactaacttggcgaactacttactcttagcttcccggcaacaattaatagactgg
 atggaggcggaataaagtgcaggaccacttctgcgtcggcccttccggctggctggttatttgcgtataaatctggagccggtgagcgtg
 gaagccgcggtatcattgcagcactggggccagatggtaagccctccgctatcgtatgtatcacaacgaggggagtcaggcaactatgg
 30 atgaacgaaatagacagatcgtgagatagggtgcctcactgattaagcatttggtaactgtcagaccaagttactcatatatactttagattgat
 ttaaaacttattttaatttaaaaggatctaggtgaagatccttttgataatctcatgacaaaaatccctaacgtgagtttcttccactgagcg
 tcagaccctgtagaaaagatcaaaaggatcttcttgagatcctttttctgcgcgtlaatctgctgcttgcacacaaaaaaaccaccgctaccag
 cgggtggtttgttgcggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgagataccaaatactgtccttctagt
 gtagccgtagttaggccaccacttcaagaactctgtagcaccgctacataacctcgtctgtaactctgttaccagtggctgctgccagtgg
 35 cgataagtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgagcggtcgggctgaacgggggggtcgtgcacaca
 gccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggagaaag
 ggggacaggtatccggtaagcggcagggtcgggaacaggagagcgacagggagcttccagggggaacgcctggtatctttatagtc
 ctgtcgggttccacctctgacttgagcgtcgattttgtgatctcgtcagggggcgagcctatggaaaaacgcgacgaacgcggc
 ctttttacggttctcggccttttgcctggttgcacatgt
 40

SEQ ID NO: 8 (pX601-MHP1-iABEmaxC2NG: ITR-MHP1-Gp41-1 inteinC-
 Cas9NG574-1368-NLS-bGHPA-hU6-mdx4cv_spacer-ogRNA_scaffold-ITR)

cctgcaggcagctgcgcgctcgtcgtcactgaggccggccggggcgtcggcgaccttggctgcccggccctcagtgagc
 45 gagcgagcgcgcagagagggagtgcccaactccatcactaggggttccctgcggccttagactcaggcCCCTTCAGATTAA
 AAATAACTGAGGTAAGGGCCTGGGTAGGGGAGGTGGTGTGAGACGCTCCTGTCTCT
 CCTCTATCTGCCCATCGGCCCTTTGGGGAGGAGGAATGTGCCCAAGGACTAAAAAA
 AGGCCATGGAGCCAGAGGGGCGAGGGCAACAGACCTTTTCATGGGCAACCTTGGG
 GCCCTGCTGactgtGATGAGAGCAGCCACTACGGGTCTAGGCTGCCCATGTAAGGAG
 50 GCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTG
 CCCCCCCCCCCCCAACACCTGCTGCCTGCTAAAAATAACCCTGTCCCTGGTGGeectgc

atgCCCTCCCTGGGGACAGCCCCCTCCTGGCTAGTCACACCCTGTAGGCTCCTCTATAT
 AACCCAGGGGCACAGGGGCTGCCCTcattctACCACCACCTCCACAGCACAGACAGACA
 CTCAGGAGCCAGCtagccaccATGATGCTCAAGAAGATCCTCAAGATTGAAGAGTTGGA
 CGAGCGCGAGCTTATAGACATAGAAGTCAGTGGTAATCACCTTTTCTACGCAAATG
 5 ACATTTTGACTCACAACCTCCTCTTCAGACGTTTGCTTCGACTCCGTGGAAATCTCCG
 GCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATAACCACGATCTGCTGAAAATT
 ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATA
 TCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAA
 ACCTATGCCCACCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATA
 10 CACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG
 TCCGGCAAGACAATCCTGGATTTCTGAAAGTCCGACGGCTTCGCCAACAGAACTT
 CATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCC
 AGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGC
 CCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAA
 15 AGTGATGGGCGGCACAAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAAC
 CAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAG
 AGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC
 CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGT
 ACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATC
 20 GTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAG
 CGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTGAAGAAG
 ATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTT
 CGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGC
 TTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGAT
 25 CCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAG
 TGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
 TTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCCTACCTGAAC
 GCCGTCTGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGT
 GTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAG
 30 GAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTTC
 AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGAC
 AAACGGCGAAACCGGGGAGATCGTGTGGGATAAAGGGCCGGGATTTTGCCACCGTGC
 GGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGAC
 AGGCGGCTTCAGCAAAGAGTCTATCCgGCCCAAGAGGAACAGCGATAAGCTGATCG
 35 CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGtCAGCCCCACCGTG
 GCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGAGGGCAAGTCCAAGAACTGAA
 GAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAG
 AATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGAT
 CATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGC
 40 TGGCCTCTGCCcGCtttCTGCAGAAGGGAAACGAACCTGGCCCTGCCCTCCAAATATGT
 GAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGGCTCCCCCGAGGATA
 ATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATC
 GAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAA
 AGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAG
 45 AATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTcggGCCTTCAAGTACT
 TTGACACCACCATCGACCGGAAGgtGTACcggAGCACCAAAGAGGTGCTGGACGCCA
 CCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAG
 CTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAATTCGAGCC
 CAAGAAGAAGAGGAAAGTCTaacTAgtAgaattctagagctcgtgatcagcctcactgtgcttctagttgccag
 50 ccatctgttgttgccttccccctgccttcttgaccttggaaggtgccactcccactgtccttcttaataaaatgaggaaattgcatcgca
 ttgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagagaatagcaggcatgctg

SEQ ID NO: 9 (pLenti-mdx^{4cv}-ogRNA: U6 promoter, mdx4cv spacer, optimized gRNA scaffold)

gtcgcaggatcgggagatctcccgatcccctatgggtgcactctcagtacaatctgctctgatgccgcatalagttaagccagtatctg
40 ctccctgcttgtgtgttgagggtcgtgagtagtgccgcgagcaaaatttaagctacaacaaggcgaaggcttgaccgacaattgcatgaagaa
tctgcttaggggttagggcggtttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaataatt
acgggggtcattagttcatagcccataatgaggttccgcgttacataacttacggtaaatggcccgctggctgaccgccccaacgacccccg
cccattgacgtcaataatgacgtatgttcccatagtaacgcccaatagggactttccattgacgtcaatgggtggagtatttacggtaaacgtcc
cacttggcagttacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgctggcattatgccagttac
45 atgaccttatgggactttcctacttggcagttacatctacgtattagtcatcgtattaccatgggtgatgcgggtttggcagttacatcaatgggcgt
ggatagcgggttgactcacggggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgggactttcca
aaatgtcgtacaactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagcagcgcgttttgcctgtactg
ggctctctcgttgtagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagccccaataaagcttgcccttgagtg
ttcaagtagtgtgtgcccgtctgtgtgactctggttaactagagatccctcagacccttttagtcagttggaaaaatctctagcagttggcgcc
50 cgaacaggggacttgaagcgaagggaaccagaggagctctctgcacgcaggactcggcgttgctgaagcgcgcacgggaagaggc
gagggggccggcactggtgagttacgcaaaaattttgactagcggaggctagaaggagagagatgggtgccgagagcgtcagttattaagc

ggggggagaattagatcgcatgggaaaaaattcggttaaggccaggggggaaagaaaaataataaataacatatagtagggcaagca
 gggagctagaacgattcgagttaatcctggcctgttagaacaacagagagcctgttagacaaatactgggacagctacaacatcccttca
 gacaggatcagaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaggatagagataaaagacaccaaggaag
 5 ctttagacaagatagagggaagagcaaaaagtaagaccaccgcacagcaagcggcgtgacatctcagacctggaggaggagata
 tgaggggacaattggagaagtgaattatataataaagtagtaaaaattgaaccattaggagtagcaccaccaaggcgaagagaagagt
 ggtgcagagagaaaaaagagcagtgagggaataggagcttcttcttgggttcttgggagcagcaggaagcactatgggagcagcgtcaa
 tgacgctgacggtagcaggccagacaattattgtctggtatagtgacagcagcagaacaatttgctgaggggctattgaggcgcaacagcatct
 gttgcaactcacagctctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatt
 10 tgggggttgcctggaaaactcatttgcaccactgctgtgcttgggaatgctagtgtggagtaataaatctctggaacagatttgggaatcacacga
 cctggatggagtgaggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaaaatga
 acaagaattattggaattagataaattgggcaagtttgggaattggttaacatacaaaattggctgtggtatataaaattattcataatgatagta
 ggaggccttggtaggttgaagaatagttttgctgtactttctatagtaatagagttaggcagggatattcaccattatcggttcagaccacctc
 ccaaccccgaggggacccgacaggcccggaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtga
 15 cggatcggcactgctgtgcgcaatttgcagacaaatggcagattatccacaattttaaaagaaaaggggggattgggggttacagtgc
 aggggaaagaatagtagacataatagcaacagacatacaactaaagaattacaaaaaattacaaaaatttccgggttatta
 caggggacagcagagatccagtttgggttaattagctagcaggtacctgagggcctatttccatgattccttcataattgcatatacgatacaag
 gctgttagagagataattggaattatttgaactgtaaacacaaagatattagtacaaaatacgtgacgtagaagtaataatttcttgggtagtt
 gcagttttaaattatgttttaaattggactatcatatgcttaccgttaacttgaaagtatttcgatttcttggctttatatacttgtgaaaggacga
 aacaccgGTTATCTCCTGTTCTGCAGCGTTTcAGAGCTAtgctgGAAAcagcaTAGCAAGTTgA
 20 AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCCGGTGCTTTTTTgcg
 gccgcggtacctgcaaagatggataaagttttaaacagagaggaatcttgcagctaattggaccttctaggtcttgaaggagtggggaattg
 gctccggtgcccgtcagtgaggcagagcgcacatcgccacagtcctccgagaagttgggggggaggggtcggaattgatccggtgccta
 gagaaggtggcgcggggttaaactgggaaagtgtgtgactggctccgcttttcccaggggtgggggagaaccgtatataagtgc
 agtagtcgcccgtgaacgttcttttcgcaacgggttgcgcccagaaacacaggtaagtgcctgtgtgtggttcccgcgggcctggcctcttac
 25 ggggttatggcccttgcgtgcttgaattacttccactggctgcagtagctgattcttgatcccgagcttcgggttggaaagtgggtgggagagtt
 cgaggccttgcgcttaaggagcccccttgcctcgtgcttgaattgagggcctggcctgggctggcgctggggccgcccgtgcgaatctggtggc
 acctcgcgcctgtctcgtgcttctgataagttcttagccatttaaaattttgatgacctgctgcgacgcttttttctggaagatagcttcta
 aatgcgggccaagatctgcacactgggtatttgcgttttggggccgcccggcgacggggcccgtgctcccagcgcacatgttcggc
 gagggcggggctgcgagcgcggccaccgagaatcggacgggggtagictcaagctggcggcctgctctgggtcctggcctcgcgcc
 30 gccgtgtatcgccccgcccctggggcggaaggctggcccgtgcggcaccagttgcgtgagcggaaagatggccgcttcccggcccctgct
 gcaggagctcaaaatggaggacgcggcgctcgggagagcggggcggtgagtcaccacacaaaggaaaagggccttccgtcctca
 gccgtcgttcatgtgactccacggagtaccggcgccgtccaggcacctcgattagttctcgagcttttggagtagctgctttaggttgg
 ggggaggggtttatgcgatggagtttccccacactgagtggtgggagactgaagttaggccagcttggcacttgatgaattctccttggaa
 tttgccctttttgagtttggatcttgggttatttcaagcctcagacagtggttcaaaatttttcttccatttcaggtgctgtagtacaattggcca
 35 agttgaccagtgcggttccggtgctcaccgcgcgcgacgtgcgggagcggctcaggtctggaccgaccggctcgggttctccgggac
 ttcgtggaggacgacttccgggtgtggttccgggacgactgacctgttcatcagcgcggtccaggaccaggtgggtcgggacaacac
 cctggccttgggtgtgggtgctcggcctggacgagctgtacgccagtggtcggaggtcgtgtccacgaacttccgggacgctcgggg
 ccggccatgaccgagatcgccgagcagcgttgggggcccggagttcgccttgcgcgaccggccggcaactgcgtgcacttcgtggcc
 gagggagcaggactgagaattcgatataagcttccgttaatacaaccttggattacaaaatttggaaagattgactggtatttctaactatgtt
 40 gctccttttacgtatgtggatagcgtccttfaatgccttctgtatcatgctattgcttcccgtatggcttccatttctccttctgtataaactcctggtt
 gctgtctctttatgaggagttgtggcccgtgtcaggcaacgtggcgtggtgtgactgtgttctgacgcaacccccactgggttggggcat
 tggcaccacctgtcagctccttccgggacttgccttccccctccctatttccacggcggaactcatcgccgctgcttcccgtcgtg
 gacaggggctcggctgttgggcaactgacaattccgtgtgtgtcggggaaatcatcgttcttcttggctgctcgcctgtgttggccactg
 gattctgcgcgggacgtccttctgctacgtcccttggccctcaatccagcgacacttcttcccggcctgctgcgggctctgcggcctct
 45 tccgctcttgccttccctcagacgagtcggatctcccttggggccgctccccgcacatcgataccgtcgacctcgagacctagaaaaac
 atggagcaatcacaagtagcaatacagcagctaccaatgctgattgtgcttggctagaaagcacaagaggaggaggaggttgggtttccag
 tcacacctcaggtaccttaagaccaatgacttacaaggcagctgtagatcttagccactttttaaagaaaaggggggactggaagggtta
 attcactcccaacgaagacaagatatacttgaatctgtggtatctaccacacacaaggctacttccctgattggcagaactacacaccaggggcc
 agggatcagataccactgaccttggatgggtgctacaagctagtaccagttgagcaagagaaggtagaagaagccaatgaaggagagaa
 50 caccgccttgttacacctgtgagcctgcatgggatggatgacccggagagagaagattagagtgagggtttgacagcgcgctagcattt
 catcacatggcccagagagctgcatccggactgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactaggga

acccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtggtgcccgtctgtgtgactctggtaactagagatccctcaga
 ccccttttagtcagtggtgaaaatctctagcagggcccggtttaaaccgctgatcagcctcagctgtgcttctagttgccagccatctgtgttt
 gccccccccctgcttcccttgacctgggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtag
 gtgtcattctattctgggggggtggggggcaggacagcaaggggggaggttgggaagacaatagcagggcatgctggggatgcgggtg
 5 ggctctatggctctgagggcggaagaaccagctggggctctagggggatccccacgcgccctgtagcggcgcatlaagcgcggcg
 gtgtgggtgttacgcgcagcgtgaccgctacacttgccagcgccctagcggccgctcttctgcttcttcccttcttctgcccacgttgcg
 cggctttccccgtcaagctcctaaatcgggggtccctttagggttccgatttagtgccttacggcacctcgacccccaaaaaacttgattagggt
 gatggttcacgtagtgggcatcgcctgatagacgggttttgcgcttgacgttggagtgccacgttcttaatagtggaactctgttccaaact
 ggaacaacactcaaccctatctcggcttattctttgattataagggattttgcccatttgcgctattgggttaaaaaatgagctgatttaaaaa
 10 aatttaacgagaattaattctgtggaatgtgtgtagttaggtgtggaaagtcgccaggtccccagcagggcagaagtatgcaaagcatgc
 atctcaatttagtcagcaaccaggtgtggaagtcgccaggtccccagcagggcagaagtatgcaaagcatgcatctcaatttagtcagcaac
 catagtcgccgccccctaaactcgcgccatccccccccctaaactcgcgccaggttccgcccattctccgccccatggctgactaatttttttattatgc
 agaggccgaggccgctctgctctgagctattccagaagtagtgaggaggtttttggaggccctaggcttttgcaaaaagctcccgga
 gcttgatatccattttcggtatctgatcagcagctgtgtgacaattaatcatcgccatagatatcgccatagataatacacaaggtgaggaac
 15 taaaccatggccaaggtgaccagtgccgttccgggtgctaccgcgcgcgacgtcgccggagcggtcgagttctggaccgaccggctcgg
 gttctcccgggacttcgtggaggacgacttcgcccgtgtgttccgggacgacgtgaccctgttcatcagcgggtccaggaccaggtggt
 gccggacaacacctggcctgggtgtgggtgcccggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacgaacttcggg
 gacgctccggggccggccatgaccgagatcgccgagcagccgtggggggcgagggttccgctgcgcgacccggccgggcaactgcgt
 gcacttcgtggccgaggagcaggactgacacgtgctacgagatttcgattccaccgccgcttctatgaaaggttgggcttcggaatcggtt
 20 tccgggacgcgggtggatgatcctccagcgcggggatctcatgctggagttcttcgccacccccaaacttgttattgagcttataatgggt
 acaataaagcaatagcatcacaatttcacaaataaagcattttttcactgcattctagttgtgtgttggccaaactcatatgtatctatcat
 gtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgttctctgtgtgaaattgtatccgctcacaattccacacaac
 atacgagccggaagcataaagtgtaaagcctgggggtgcctaattgagtgagctaaactcacattaattgctgtgcctcactgcccgtttcca
 gtcgggaaacctgtcgtgccagctgcattaatgaatcgcccaacgcgcggggagaggcggttgcgtattgggcgctcttccgcttctcg
 25 ctactgactcgtcgcgtcggctcgttcggctcgcgcgagcgggtatcagctcactcaaaaggcggtatacgggtatccacagaatcagggg
 ataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaaccgtaaaaaggccggtgtcgttggctgttttccataggctc
 cgccccctgacgagcatcacaataatcagcgtcgaagtcagaggtggcgaaaaccgacaggactataaagataaccaggcggttcccc
 tggaaagctccctcgtcgcgtctctgttccgacctgcgcgttaccggatactgtccgcttctcccttcgggaagcgtggcgcttctcat
 agctcacgcgttaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccggtcagcccagccgtcgcg
 30 ccttatccggtaactatcgtctttagtccaaccggtaagacacgacttatcgccactggcagcagccactggttaacaggattagcagagc
 gaggtatgtaggcggtgtacagagttctgaagtgttggcctaactacggctacactagaagaacagttatttggtatctcgcgtctgtgaa
 gccagttaccttcggaaaaagagttggtagctcttgatccggcaaaacaaccaccgctggttagcggtgtgtttttgttgcaagcagcagat
 tacgcgcagaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggatttt
 ggtcatgagattatcaaaaaggatcttccactagatccttttaataaaaaatgaagtttaaatcaatctaaagtatataatgagtaaaacttggtct
 35 gacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcacatagttgctgactccccgctgtgtgataact
 acgatacggggagggttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaac
 cagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgctccatccagcttattatgttgcggggaagctagagtaa
 gtagttgccagttaatagtttgcgcaacgttgttgcattgtacagggcatcgtggtgtcacgctcgtcgtttggtaggttcattcagctcgg
 gttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaagcggttagctccttcggctccgaltgtgtcagaagtaagtt
 40 ggccgcagtggttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgctttctgtgactgggtgagtctc
 aaccaagtcattctgagaatagtgtagtgcggcgaccgagttgcttcttgcggcgctcaatacgggataataccgcgccacatagcagaactt
 taaaagtgctcatcattggaaaacgttcttcggggcgaaaacttcaaggatcttaccgctgttgagatccagttcagtgtaaccactcgtgc
 acccaactgatcttcagcatcttttactttcaccagcggttctgggtgagcaaaaacagggaaggcaaaatgccgcaaaaagggaataagg
 gcgacacggaaatgttgaatactcatactcttcttttcaataattattgaagcatttatcagggtattgtctcatgagcgggatacatattgaatg
 45 tatttagaaaaataaacaataaggggttccgcgcacatttccccgaaaagtgccacctgac

SEQ ID NO: 10 (pLKO-puro-2A-mdx^{4cv}-EGFP)

gggctctctgtggttagaccagatctgagcctgggagctctctggctaactagggaaaccactgcttaagcctcaataaagcttgc
 cttgagtgcttcaagtagtggtgcccgtctgtgtgactctggtaactagagatccctcagacccttttagtcagtggtgaaaatctctagc
 50 agtggcgcccgaaacagggacttgaaagcgaaagggaaccagaggagctctctcagcgcaggactcggcttgcgtgaagcgcgcacgg
 caagaggcgagggggcgactggtgagtagcgcacaaaaatttgactagcggaggctagaaggagagagatgggtgagagagcgtca

gtattaagcgggggagaattagatcgcatgggaaaaaattcggftaaggccagggggaaagaaaaatataaattaaacatatagtatg
 ggcaagcagggagctagaacgattcgagftaatcctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaacc
 atcccttcagacagggatcagaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacac
 caaggaagctttagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatttcagacctggagg
 5 aggagatatgagggacaattggagaagtgaattatataataaagtagtaaaattgaaccattaggagtagcaccaccaaggcaaaag
 agaagagtgggtgcagagagaaaaaagagcagtggggaataggagccttgttcttgggttcttgggagcagcaggaagcactatgggcgc
 agcgtcaatgacgtgacggtacaggccagacaattattgtctggtatagtcagcagcagacaacaatttctgagggctattgagggcga
 cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaggatcaacagctcc
 tggggatttgggggttgccttggaaaactcaattgcaccactgctgtgcttgggaatgctagtgtgagtaataaattctctggacagatttggat
 10 cacacgacctggatggagtgggacagagaaftaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaa
 agaatgaacaagaattattggaattagataaattgggcaagtttgggaattgggttaacataacaaattggctgtggtatataaattattcataa
 ttagtagtaggagggcttggtaggttaagaatagtttttctgtactttctatagtgaaatagagttaggcaggggatattaccattatcgtttcagac
 ccacctcccaaccccgagggggacccgacaggcccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgatt
 agtgaacgggatctcgacgggtatcgatcacgagactagcctcgagacaaatggcagttatccacaaattttaaagaaaaaggggggattg
 15 ggggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaaactaaagaattacaaaaacaaattacaaaaattcaaa
 attttcgggtttattacagggacagcagagatccactttggccgaggTAGTTATTAATAGTAATCAATTACGGGG
 TCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGC
 CCGCCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTT
 CCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACG
 20 GTAAACTGCCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTAT
 TGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGACCTTATG
 GGACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT
 GCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTC
 AAGTCTCCACCCCATTTGACGTCAATGGGAGTTTTGTTTTGGCACCAAAATCAACGGG
 25 ACTTTCCAAAATGTCGTAACAACTCCGCCCATTTGACGCAAAATGGGCGGTAGGCGT
 GTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCGCTA
 GCCACCATGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCC
 CAGGGCCGTACGCACCCTCGCCGCCGCGTTTCGCCGACTACCCCGCCACGCGCCACA
 CCGTCGATCCGGACCGCCACATCGAGCGGGTCAACGAGCTGCAAGAACTCTTCCTC
 30 ACGCGCGTTCGGGCTCGACATCGGCAAGGTGTGGGTTCGCGGACGACGGCGCCGCGGT
 GGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTTCGCCGAGATC
 GGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGA
 AGGCCTCCTGGCGCCGACCGGCCCAAGGAGCCCGCGTGGTTCTTGCCACCGTCG
 GCGTgTCGCCCCGACCACAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGA
 35 GTGGAGGCGGCCGAGCGCGCCGGGTGCCCGCCTTCCTGGAaACCTCCGCGCCCCG
 CAACCTCCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCC
 CGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCCTGTACAAGCAAT
 GTACTAACTACGCTTTGTTGAACTCGCTGGCGATGTTGAAAGTAACCCCGGTCTCTg
 aattcCAAGAACAGCTGCAGAACAGGAGATAACAGTTGgagatccGTGAGCAAGGGCGAG
 40 GAGCTGTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAAACGG
 CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
 CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGA
 CCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG
 CACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC
 45 TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
 CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
 CTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCATGGCCGA
 CAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGAC
 GGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC
 50 CGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
 CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGATC

ACTCTCGGCATGGACGAGCTGTACAAGtaaGgtaccttaagaccaatgacttacaaggcagctgtagatcttag
 ccactttttaaagaaaaggggggactggaagggtacttaactcccaacgaagacaagatctgcttttgccttgactgggtctctctggtta
 gaccagatctgagcctgggagctctctggtactagggaaaccactgcttaagcctcaataaagcttgccttgagtgttcaagtagtgtgt
 gcccgtctgtgtgactctggttaactagagatccctcagacccttttagtcagtggtgaaaatctctagcagtagtagttcatgtcatcttatta
 5 ttcagtatttataacttgcaaagaaatgaatatcagagagtggagggaaactgtttattgcagcttataatgggttacaataaagcaatagcatca
 caaatttcacaaataaagcaatttttactgcattctagttgtgtgtgttgcacaaactcatcaatgtatcttcatgtctggtcttagctatcccgcc
 cctaactccgcccattcccgcccctaactccgcccagttccgcccattctccgcccattggtcagctactaatttttttattatgcagaggccgag
 gccgctcggcctctgagctattccagaagtagtgaggaggctttttggaggcctaggggacgtaccaattccgcttatagtgagtcgtatt
 acgcgcgctcactggccgtctgtttacaacgtcgtgactgggaaaaccctggcggttaccacacttaatcgcttgcagcacatccccctttcg
 10 ccagctggcgtaatagcgaagaggcccgaccgatcccttcccaacagttggcgagcctgaatggcggaatgggacgcgcctgttag
 cggcgcatthaagcgcggcggtgtgtgtgtgtacgcgcagcgtgaccgtacacttgcagcgccctagcgcgcctctcttctgctttctt
 ccttcttctcgcacgttccgcccgttcccgctcaagcttaaaatcgggggctcccttttaggggtccgatttagtgccttacggcacctcga
 ccccaaaaaacttgattagggtgatgggtcagctagtgggccatcgccctgatagacgggttttgcgccccttgacgttggagtcacggttctta
 atagtgagctctgttccaaactggaacaacactcaaccctatctcggtctattcttttgattataagggttttgcgatttcggcctattggtta
 15 aaaaatgagctgatttaacaaaaatlaacgcgaatttlaacaaaatlaaacgcttacaatttaggtggcacttttggggaaatgtgcgcgga
 accctatttgttttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaaga
 gtatgagtattcaacattccgtgtcgccttattccctttttgcggcattttgccttctgttttgcctcaccagaaacgctggtgaaagtaaaa
 gatgctgaagatcagttgggtgcacgagtggtgtacatgaactggatcacaacagcggtgaagatccttgagagtttgcggccgaagaac
 gttttcfaatgatgagcacttttaaaagtctgtatgtggcgcggtattatcccgattgacggcggaagagcaactcggctcggcgataca
 20 ctattctcagaatgacttgggtgagtactaccagtcacagaaaageatcttacggatggcatgacagtaagagaattatgagtgctgccat
 aacctgagtgataacacttgcggccaacttacttctgacaacgatcgaggaccgaaggagctaacgccttttttgcacaacatgggggat
 catgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggca
 acaacgttgcgcaaaactattacttggcgaaactacttacttagcttcccgcaacaattaatagactggatggaggcggaataaagtgcagg
 accacttctgcgctcggcccctccggctggctgtttattgtgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcact
 25 gggggccagatggtaagccctcccgatctgtagttatctacacgacggggagtcaggcgaactatggatgaacgaaatagacagatcgcgtga
 gatagggtgccctcactgattaagcattggtaactgtcagaccaagtttactatataacttttagattgatttaaaacttatttttaaaaggat
 ctagggtgaagatccttttgataatctcatgacaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaa
 ggatcttcttgagatccttttttgcgcgtaactgtctgcttgcacaaaaaaaccaccgtaccagcgggtgtgttgcggatcaaga
 gctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagcgttagtagccaccacttc
 30 aagaactctgtagcaccgcctacatacctcgtctgctaatcctgttaccagtggtgctgctgccagtggcgataagtcgtgtcttaccgggttg
 gactcaagacgatagttaccgggataaggcgagcgggtcgggctgaacgggggtcgtgcacacagcccagcttggagcgaacgacct
 acaccgaactgagatacctacagcgtgagctatgagaaagcgccacgttcccgaaggagaaaggcgagcaggtatccggtaagcgg
 cagggtcgggaacaggagagcgcacgagggagcttcaggggggaacgcctgggtatctttatagtcctgtcgggttccgcccctctgactt
 gagcgtcgaattttgtgatgtcgtcagggggggcgagcctatggaaaaacgccagcaacgcggccttttacgggtcctggccttttgcgtg
 35 gccttttgcctcacatgttcttctcgttatccctgattctgttgataaccgtattaccgctttgagtgagctgataccgctcgcgcgagccg
 aacgaccgagcgcagcagctgagtgagcggaggaagcggaagagcgcccaatacgcacaaaccgcctctcccgcgctgtggcgattca
 ttaatgcagctggcacgacagggttcccagctggaaagcgggcagtgagcgcacgcaatlaattgtgagttagctcactcattaggcaccc
 caggctttacactttatgcttccgctcgtatgtgtgtggaattgtgagcgggataacaatttcacacaggaaacagctatgacctgattacgc
 caagcgcgcaatlaaacctcactaaagggaacaaaagctggagctgcaagcttaattgtagtcttatgcaatacttcttagcttgcacatg
 40 gtaacgatgagttagcaacatgccttacaaggagagaaaaagcaccgtgcatgcccattgggtggaagtaagggtgtacgatcgtgccttat
 taggaaggcaacagacgggtctgacatggattggacgaaccactgaattgccgcattgcagagatattgtatttaagtgcctagctcgatac
 ataaac

SEQ ID NO: 11, Wild Type SpCas9(D10A): Protein sequence:

45 MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 50 KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQOLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH

QIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
 TPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVT
 EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
 LGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 5 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ
 KAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHHPENIVIAMAREN
 QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
 QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYW
 RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
 10 YDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
 RPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDK
 LIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEK
 NPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
 15 LYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYN
 KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLY
 ETRIDLSQLGGD

SEQ ID NO: 12, SpCas9(D10A)-NG: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSIDKADLRILIYLAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 25 KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
 QIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
 TPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVT
 EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
 30 LGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ
 KAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHHPENIVIAMAREN
 QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
 QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYW
 35 RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
 YDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIK
 YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
 RPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESIRPKRNSDK
 LIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEK
 40 NPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARFLQKGNELALPSKYVNF
 LYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYN
 KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDTLIHQSIITGLY
 ETRIDLSQLGGD

SEQ ID NO: 13, xCas9(3.7)-D10A: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSIDKADLRILIYLAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 50 NGLFGNLIASLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFD

QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQ
 IHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
 PWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
 GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFRNASL
 5 GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
 KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKA
 QVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
 QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
 DINRLSDYDVHDIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKKNYWRQL
 10 LNAKLITQRKFDNLTAKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
 NDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKL
 ESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARK
 KDWDPPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSSFENPIDFL
 15 EAKGYKEVKKDLIIKLPKYSLENGRKRMLASAGVLQKGNELALPSKYVNFLYLAS
 HYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD
 KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETR ID
 LSQLGGD

SEQ ID NO: 14, SpCas9(D10A)-NGX: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLA LAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 25 NGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQ
 IHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
 PWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
 30 GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFRNASL
 GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
 KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKA
 QVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
 QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
 35 DINRLSDYDVHDIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKKNYWRQL
 LNAKLITQRKFDNLTAKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
 NDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKL
 ESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESIRPKRNSDKLIARK
 40 KDWDPPKKYGGFVSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSSFENPIDFL
 EAKGYKEVKKDLIIKLPKYSLENGRKRMLASARFLQKGNELALPSKYVNFLYLASH
 YEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD
 KPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVL DATLIHQ SITGLYETR ID
 LSQLGGD

SEQ ID NO: 15, SpCas9(D10A)-NGA: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLA LAHMIKFRGHFLIEGDL
 50 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAA

KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
 QIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGPLARGNSRFAMTRKSEETI
 TPWNFEVVDKGAASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVT
 5 EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
 LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQ
 KAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHHPENIVIAMAREN
 QTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
 10 QELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKKNYW
 RQLLNAKLITQRKFDNLTAKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTK
 YDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIK
 YPKLESEFVYGDYKVVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
 RPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESIRPKRNSDK
 15 LIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSF
 NPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASARFLQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVL SAYN
 KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKQYRSTKEVLDATLIHQSTGLY
 ETRIDLSQLGGD

SEQ ID NO: 16, SpCas9(D10A)-NGC: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL
 25 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
 QIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGPLARGNSRFAMTRKSEETI
 30 TPWNFEVVDKGAASAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVT
 EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS
 LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQ
 LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQ
 KAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHHPENIVIAMAREN
 35 QTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
 QELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKKNYW
 RQLLNAKLITQRKFDNLTAKAERGGLSELDAKAGFIKRQLVETRQITKHVAQILDSRMNTK
 YDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIK
 YPKLESEFVYGDYKVVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
 40 RPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESIRPKRNSDK
 LIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSSF
 NPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASARFLQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVL SAYN
 KHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKEYRSTKEVLDATLIHQSTGLY
 45 ETRIDLSQLGGD

SEQ ID NO: 17, SpCas9(D10A)-NGX-NGA: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 50 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK

NGLFGNLIASLSGLTPNFKSNFDLAEDTKLQLSKDTYDDDDLNDLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQ
 IHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
 5 PWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
 GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
 GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQL
 KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKA
 QVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
 10 QKGQKNSRERMKRIEELGKELGSQLKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
 DINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKKNYWRQL
 LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
 NDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKL
 ESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 15 TNGETGEIVWDKGRDFATVRKVLSPQVNVKKTETVQTGGFSKESIRPKRNSDKLIARK
 KDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFL
 EAKGYKEVKKDLIIKLPKYSLENGRKRMLASARFLQKGNELALPSKYVNFLYLASH
 YEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHARDK
 PIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKQYRSTKEVLDATLIHQSIITGLYETRIDL
 20 SQLGGD

SEQ ID NO: 18, SpCas9(D10A)-NGX-NGC: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 25 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIASLSGLTPNFKSNFDLAEDTKLQLSKDTYDDDDLNDLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQ
 30 IHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
 PWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE
 GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
 GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQL
 KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKA
 35 QVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
 QKGQKNSRERMKRIEELGKELGSQLKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
 DINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVKKMKKNYWRQL
 LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
 NDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKL
 40 ESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
 TNGETGEIVWDKGRDFATVRKVLSPQVNVKKTETVQTGGFSKESIRPKRNSDKLIARK
 KDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFL
 EAKGYKEVKKDLIIKLPKYSLENGRKRMLASARFLQKGNELALPSKYVNFLYLASH
 YEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHARDK
 45 PIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKQYRSTKEVLDATLIHQSIITGLYETRIDL
 SQLGGD

SEQ ID: 19, SpCas9(D10A)-NG-loop: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 50 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEGDL

NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIASLSGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 5 QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRF
 AWMTRKSEETITPWNFEFVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
 YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV
 EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
 AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMOI
 10 HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHK
 PENIVIEMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLY
 YLQNGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP
 EEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAY
 15 LNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF
 KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFS
 KESIRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELL
 GITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARFLQKGN
 ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
 20 ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLD
 ATLIHQISITGLYETRIDLSQLGGD

SEQ ID NO: 20, SpCas9(D10A)-NGX-loop: Protein sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
 25 GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLAALAHMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK
 NGLFGNLIASLSGLTPNFKSNFDLAEDTKLQLSKDTYDDDLNLLAQIGDQYADLFLAA
 KNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 30 QSKNGYAGYVGADKKLRKRSSKLATEEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
 QRTFDNGIIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
 WMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
 NELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
 SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
 35 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQIQLIHD
 DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPEN
 IVIEMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQ
 NGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEV
 VKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ
 40 ILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNA
 VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESI
 RPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVAKVEKGKSKKLKSVKELLGITI
 MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASARFLQKGNELA
 45 LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL
 DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDATL
 IHQSITGLYETRIDLSQLGGD

SEQ ID NO: 21, ScCas9(D10A): Protein sequence:

50 MEKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFD
 SGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFQRLEESFLVEEDKKN

ERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKL
 NAENSDVAKLFYQLIQTYNQLFEEESPLDEIEVDAKGILSARLSKSKRLEKLIIVFPNEKK
 NGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKDQYDDDLDELLGQIGDQYADLFSAA
 KNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFK
 5 DDTKNGYAGYVGIGIKHRKRTTKLATQEEFYKFIPILEKMDGAEELLAKLNRDDLRLK
 QRTFDNGSIPHQIHLKELHAILRRQEEFYFPLKENREKIEKILTFRIPYYVGPLARGNSRFA
 WLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYN
 ELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIG
 VEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLF
 10 DDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDFGFSNRNFMQLIHDDS
 LTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIIKKGILQTVKIVDELVKVMGHKPENIVIE
 MARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRD
 MYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSR
 15 MNTKRDKNKPIREVKVITLKSCLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVG
 ALIKKYPKLESEFVYGDYKVDYVRKMIKSEQEIGKATAKRFFYSNIMNFFKTEVKLAN
 GEIRKRPLIETNGETGEVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSK
 RESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKEKGKAKKLKSVKVLVGITIMEK
 GSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQH
 20 LVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSSFDEQF
 AVSDSILLSNSFVSLKYTSFGASGGFTFLDLVDKQGRRLRYQTVTEVLDTLIYQSITGL
 YETRTDLSQLGGD

SEQ ID NO: 22, TadA-TadA*: Protein sequence:

25 SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHD
 PTAHAIEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKT
 GAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAAQKKAQSSTDG
 GSSGGSSGSETPGTSESATPSSGGSSGSSSEVEFSHEYWMRHALTLAKRARDEREVPV
 GAVLVNNRVIGEGWNRRAIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCV
 30 MCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALL
 CYFFRMPRQVFNAQKKAQSSTD

SEQ ID: 23, TadA*(A56G_V82G): Protein sequence:

35 SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNRRAIGLHDP
 TGHAEIMALRQGGGLVMQNYRLIDATLYGTTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
 GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

SEQ ID NO: 24, Gp41-1-N: Protein sequence:

40 CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYKITLEDG
 KEIICSEEHLFPTQTGEMNISGGLKEGMCLYVKE

SEQ ID NO: 25, Gp41-1-C: Protein sequence:

MMLKKILKIEELDERELIDIEVSGNHLFYANDILTHNS

45 SEQ ID NO: 26, Cfa-N: Protein sequence:

CLSYDTEILTVEYGFPLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEV
 FEYCLEDGSIIRATKDHKFMTTDGMPLPIDEIFERGLDLKQVDGLPN

SEQ ID NO: 27, Cfa-C: Protein sequence:

50 MVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

SEQ ID NO: 28, meCMV:

tcgagCGCGTGATGAGAGCAGCCACTACGGGTCTAGGCTGCCCATGTAAGGAG
GCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATTAACCCAGACATGTGGCTG
CCCCCCCCCCCCAACACCTGCTGCCTGCTAAAAATAACCCTGTCCCTGGTG Gcctgc
5 atgccACTCACGGGGATTTCCAAGTCTCCACCCCATGACGTCAATGGGAGTTTGT
TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGAC
GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATC

SEQ ID NO: 29, SEMHP1:

CCCTTCAGATTAAAAATAACTGAGGTAAGGGCCTGGGTAGGGGAGGTGGTG
TGAGACGCTCCTGTCTCTCCTCTATCTGCCCATCGGCCCTTTGGGGAGGAGGAATGT
GCCCAAGGACTAAAAAAGGCCATGGAGCCAGAGGGGCGAGGGCAACAGACCTTT
CATGGGCAAACCTTGGGGCCCTGCTGactgtaGATGAGAGCAGCCACTACGGGTCTAG
15 GCTGCCCATGTAAGGAGGCAAGGCCTGGGGACACCCGAGATGCCTGGTTATAATTA
ACCCAGACATGTGGCTGCCCCCCCCCCCCAACACCTGCTGCCTGCTAAAAATAACC
CTGTCCCTGGTG GcctgcatGCCCTCCCTGGGGACAGCCCCCTCCTGGCTAGTCACACCC
TGTAGGCTCCTCTATATAACCCAGGGGCACAGGGGCTGCCCTcattctACCACCACCTCC
ACAGCACAGACAGACACTCAGGAGCCAGC

SEQ ID NO: 30, Wild Type SpCas9(D10A): Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
GGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAG
25 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGAAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
GAGAAGTACCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACA
30 AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGC
CACTTCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
GAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAAATGGCCTGTTTCGGAACCT
35 GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
GCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
CCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCT
40 GCTGAAAGCTCTCGTGCAGGACGAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
GTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACCTGC
TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
CAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
45 AAGATTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAACAGCAGATTCGCCTG
GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACTTCGAGGAAGTGGTG
GACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
50 TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
CTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC

GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
 GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
 CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
 AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
 5 ATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAGTGATGAAGCA
 GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
 GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
 CTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
 AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
 10 GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
 GGTGGACGAGCTCGTGAAAGTGATGGGCGGCACAAGCCCGAGAACATCGTGATC
 GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
 AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
 AACACCCCGTGGA AAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
 15 CAGAATGGGCGGGATATGTACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGA
 CTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
 ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
 CGAAGAGGTTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAG
 CTGATTACCCAGAGAAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
 20 CGAACTGGATAAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
 CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
 GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
 TTTCCGGAAGGATTTCCAGTTTTTACAAAGTGC GCGAGATCAACA ACTACCACACG
 CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCT
 25 AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
 GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
 GCAACATCATGA ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
 AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG
 GCCGGGATTTTGCCACCGTGC GGAAAGTGCTGAGCATGCCCAAGTGAATATCGTG
 30 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGA
 GGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGG
 CGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGA
 AAGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCAT
 GGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
 35 AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTG
 GAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACG
 AACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGA
 AGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCAC
 AAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGAT
 40 CCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATA
 AGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTG
 GGAGCCCCTGCCGCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACAC
 CAGCACCAAGAGGTGCTGGACGCCACCTGATCCACCAGAGCATCACCGGCCTGT
 ACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC
 45

SEQ ID NO: 31, SpCas9(D10A)-NG: Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCAGCAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 50 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA

GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGC
5 CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
GAAAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
10 AGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
GCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
CCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCT
GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
15 ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
GTTCTACAAGTTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCAGGAACTGC
TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
CAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
20 TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAACAGCAGATTTCGCTG
GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTG
GACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
25 CTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC
GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
GACTCCGTGGAAATCTCCGGCGTGGAAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
30 ATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAGTGATGAAGCA
GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
CTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
35 GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
GGTGGACGAGCTCGTGAAAGTGATGGGCGGCACAAGCCCGAGAACATCGTGATC
GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
AACACCCCGTGGAACACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
40 CAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGA
CTACGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
CGAAGAGGTCTGTGAAGAAGATGAAGAATACTGGCGGCAGCTGCTGAACGCCAAG
CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
45 CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
TTTCCGGAAGGATTTCCAGTTTACAAAGTGCGCGAGATCAACAATACTACCACCACG
CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCT
50 AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA

GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
 AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG
 GCCGGGATTTTGGCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG
 5 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCgGCCCAAGAG
 GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAAGTACGGC
 GGCTTCGtCAGCCCCACCGTGCCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAG
 GGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
 AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGGCTACAAA
 GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGA
 10 AAACGGCCGGAAGAGAATGCTGGCCTCTGCCcGCtttCTGCAGAAGGGGAAACGA
 GGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT
 GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAG
 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCT
 GGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCCGGGATAAGC
 15 CCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
 GCCCCCTcggGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGgtGTACcggAGCAC
 CAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
 CACGGATCGACCTGTCTCAGCTGGGAGGTGAC

20 SEQ ID NO: 32, xCas9(3.7)-D10A, Nucleotide sequence:
 ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCAGCAGGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 25 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGAAAGAGGATA
 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
 GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCCGACA
 AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGGC
 30 CACTTCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
 CATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCA
 GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
 GAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
 GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
 35 AGGATACCAAACCTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
 GCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
 CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
 CCCCTGAGCGCCTCTATGATCAAGCTGTACGACGAGCACCAACAGGACCTGACCCT
 GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
 40 ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
 GTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGC
 TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
 CATCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
 AAGATTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
 45 TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTTCGCCTG
 GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACTTCGAGAAGGTGGTG
 GACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
 CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
 TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
 50 CTTCTGAGCGGCGACCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC
 GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC

GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
 CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
 AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
 ATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAGTGATGAAGCA
 5 GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
 GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
 CTTCGCCAACAGAACTTCATCCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
 AGGACATCCAGAAAGCCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
 GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
 10 GGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATC
 GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
 AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
 AACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
 CAGAATGGGCGGGATATGTACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGA
 15 CTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
 ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
 CGAAGAGGTCTGTGAAGAAGATGAAGAAGTACTGGCGGCAGCTGCTGAACGCCAAG
 CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
 CGAACTGGATAAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
 20 CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
 GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
 TTTCCGGAAGGATTTCCAGTTTACAAAGTGCGCGAGATCAACAACCTACCACCACG
 CCCACGACGCCTACCTGAACGCCGCTCGTGGGAACCGCCCTGATCAAAAAGTACCCT
 AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
 25 GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
 GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
 AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAAG
 GCCGGGATTTTGCCACCGTGCAGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG
 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGA
 30 GGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGG
 CGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGA
 AAGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCAT
 GGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
 AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTG
 35 GAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGTGCTGCAGAAGGGGAAACG
 AACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGA
 AGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCAC
 AAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGAT
 CCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATA
 40 AGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTG
 GGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACAC
 CAGCACCAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGT
 ACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGAC

45 SEQ ID NO: 33, SpCas9(D10A)-NGX: Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 50 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATA

AGAAGCACGAGCGGCACCCCATCTTTCGGCAACATCGTGGACGAGGTGGCCTACAC
GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGC
CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
5 CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
GAAAATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATACCAAACCTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
10 GCTGGCCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
CCCCTGAGCGCCTCTATGATCAAGCTGTACGACGAGCACCAACAGGACCTGACCCT
GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
15 GTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGC
TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
CATCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
TTCCGCATCCCCTACTACGTGGGGCCCTCTGGCCAGGGGAAAACAGCAGATTTCGCTG
20 GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGAAGGTGGTG
GACAAGGGCGCTTCCGCCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
CCTGCCCAACGAGAAGGTGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
CTTCCTGAGCGGCGACCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC
25 GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
ATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAGTGATGAAGCA
30 GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
CTTCGCCAACAGAACTTCATCCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
35 GGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATC
GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
AACACCCCGTGGAACACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
CAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGA
40 CTACGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
CGAAGAGGTGCTGAAGAAGATGAAGAATACTGGCGGCAGCTGCTGAACGCCAAG
CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
45 CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
TTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCAG
CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCT
AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
50 GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG

AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG
 GCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG
 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC_gGCCCAAGAG
 GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGC
 5 GGCTTCG_tCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAG
 GGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
 AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGGCTACAAA
 GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGA
 AAACGGCCGGAAGAGAATGCTGGCCTCTGCC_cGC_{ttt}CTGCAGAAGGGGAAACGAAC
 10 GGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT
 GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG
 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCT
 GGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGC
 CCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
 15 GCCCC_{Tcgg}GCCTTCAAGTACTTTGACACCACCATCGACCGGAAG_{gt}GTAC_{cgg}AGCAC
 CAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
 CACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID NO: 34, SpCas9(D10A)-NGA: Nucleotide sequence:

20 ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCAGCAGGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 25 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGAAAGAGGATA
 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
 GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
 AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGC
 CACTTCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
 30 CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
 GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
 GAAAATCTGATCGCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
 GATTGCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCG
 AGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
 35 GCTGGCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
 CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
 CCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCT
 GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
 ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
 40 GTTCTACAAGTTTCATCAAGCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGC
 TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
 CAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
 AAGATTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
 TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTTCGCTG
 45 GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTG
 GACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
 CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
 TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
 CTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC
 50 GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
 GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA

CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
 AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
 ATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAGTGATGAAGCA
 GCTGAAGCGGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
 5 GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
 CTTCGCCAACAGAACTTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
 AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
 GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
 GGTGGACGAGCTCGTGAAAGTGATGGGCGGCACAAGCCCGAGAACATCGTGATC
 10 GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
 AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
 AACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
 CAGAAATGGGCGGGATATGTACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGA
 CTACGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
 15 ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
 CGAAGAGGTCTGTGAAGAAGATGAAGAAGTACTGGCGGCAGCTGCTGAACGCCAAG
 CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
 CGAACTGGATAAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
 CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
 20 GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
 TTTCCGGAAGGATTTCCAGTTTACAAAGTGCGCGAGATCAACAACCTACCACCACG
 CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCT
 AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
 GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
 25 GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
 AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG
 GCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG
 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCgGCCAAGAG
 GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGC
 30 GGCTTCGcCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAG
 GGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
 AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAA
 GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGA
 AAACGGCCGGAAGAGAATGCTGGCCTCTGCCcGCtttCTGCAGAAGGGGAAACGAAC
 35 GGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT
 GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG
 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCT
 GGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGC
 CCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
 40 GCCCTcggGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGcaGTACcggAGCAC
 CAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
 CACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID NO: 35, SpCas9(D10A)-NGC: Nucleotide sequence:

45 ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCAGCAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 50 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATA
 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC

GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGC
CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
5 GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
GAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
GCTGGCCCGAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
10 CCGACGCCATCCTGTCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
CCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCT
GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
GTTCTACAAGTTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCAGGAACTGC
15 TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
CAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTG
GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACTTCGAGGAAGTGGTG
20 GACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
CCTGCCCAACGAGAAGGTGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
CTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGACCTGCTGTTCAAGACCAACC
GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
25 GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
ATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAGTGATGAAGCA
GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
30 GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
CTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
GGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATC
35 GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
AACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
CAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGA
CTACGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
40 ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
CGAAGAGGTGCTGAAGAAGATGAAGAATACTGGCGGCAGCTGCTGAACGCCAAG
CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
45 GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
TTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACG
CCCACGACGCCTACCTGAACGCCGTCTGGGAACCGCCCTGATCAAAAAGTACCCT
AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
50 GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG

GCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG
 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC_gGCCCAAGAG
 GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGC
 GGCTTCG_tCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAG
 5 GGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
 AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGGCTACAAA
 GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGA
 AAACGGCCGGAAGAGAATGCTGGCCTCTGCC_cGC_{ttt}CTGCAGAAGGGGAAACGAAC
 GGCCCTGCCCTCCAAATATGTGAAC_{tt}CCTGTACCTGGCCAGCCACTATGAGAAGCT
 10 GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG
 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCT
 GGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCCGGGATAAGC
 CCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
 GCCCC_{Tcgg}GCCTTCAAGTACTTTGACACCACCATCGACCGGAAG_{ga}GTAC_{cgg}AGCAC
 15 CAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
 CACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID NO: 36, SpCas9(D10A)-NGX-NGA: Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 20 GGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCCTGGTGGAAAGAGGATA
 25 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
 GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
 AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCGGGGC
 CACTTCCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
 CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
 30 GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
 GAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
 GATTGCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCG
 AGGATACCAAACCTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
 GCTGGCCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
 35 CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
 CCCCTGAGCGCCTCTATGATCAAGCTGTACGACGAGCACCACCAGGACCTGACCCT
 GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
 ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
 GTTCTACAAGTTCATCAAGCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGC
 40 TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
 CATCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
 AAGATTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
 TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAACAGCAGATTTCGCTG
 GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTCGAGAAGGTGGTG
 45 GACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
 CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
 TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
 CTTCCTGAGCGGCGACCAAGAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACC
 GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
 50 GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
 CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG

AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
 ATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAGTGATGAAGCA
 GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
 GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
 5 CTTCGCCAACAGAACTTCATCCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
 AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
 GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGGCATCCTGCAGACAGTGAAGGT
 GGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATC
 GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
 10 AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
 AACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
 CAGAATGGGCGGGATATGTACGTGGACCAGGAAGTGGACATCAACCGGCTGTCCGA
 CTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
 ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
 15 CGAAGAGGTGCTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAG
 CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
 CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA
 CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
 GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
 20 TTTCGGGAAGGATTTCCAGTTTACAAAGTGC GCGAGATCAACA ACTACCACCACG
 CCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCT
 AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
 GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
 GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
 25 AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG
 GCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCAAGTGAATATCGTG
 AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC_gGCCCAAGAG
 GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGC
 GGCTTCGTCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAG
 30 GGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
 AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAA
 GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGA
 AAACGGCCGGAAGAGAATGCTGGCCTCTGCC_cGC_{ttt}CTGCAGAAGGGAAACGA
 GGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT
 35 GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG
 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCT
 GGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGC
 CCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
 GCCCC_{egg}GCCTTCAAGTACTTTGACACCACCATCGACCGGAAG_{Gca}GTAC_{egg}AGCAC
 40 CAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
 CACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID NO: 37, SpCas9(D10A)-NGX-NGC, Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 45 GGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGAAAGAGGATA
 50 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
 GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA

AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGC
CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
GCGGCGTGACGCCAAGGCCATCCTGTCCTGCCAGACTGAGCAAGAGCAGACGGCTG
5 GAAAATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATACCAAACCTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
GCTGGCCCGAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
10 CCCCTGAGCGCCTCTATGATCAAGCTGTACGACGAGCACCACCAGGACCTGACCCT
GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGA
GTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGC
TCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGG
15 CATCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGG
AAGATTTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACC
TTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTTCGCTG
GATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGAAGGTGGTG
GACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAA
20 CCGTGGCCAAACGAGAAGGTGCTGCCCCAAGCACAGCCTGCTGTACGAGTACTTCACCG
TGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGC
CTTCCTGAGCGGCGACCAGAAAAAGGCCATCGTGACCTGCTGTTCAAGACCAACC
GGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC
GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATA
25 CCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATG
ATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAGTGATGAAGCA
GCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC
GGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAAGTCCGACGG
30 CTTCGCCAACAGAACTTCATCCAGCTGATCCACGACGACAGCCTGACCTTTAAAG
AGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATT
GCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGT
GGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATC
GAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAG
35 AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAG
AACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTG
CAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGA
CTACGATGTGGACCATATCGTGCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
40 CGAAGAGGTTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAG
CTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAACCCGGCAGATCA
CAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
GACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGA
45 TTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCAG
CCCACGACGCCTACCTGAACGCCGTCTGGGAACCGCCCTGATCAAAAAGTACCCT
AAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGAT
GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACA
GCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGG
50 AAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG
GCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG

AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCgGCCCAAGAG
 GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGC
 GGCTTCGcCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGA
 5 GGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
 AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAA
 GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGA
 AAACGGCCGGAAGAGAATGCTGGCCTCTGCCcGCtttCTGCAGAAGGGGAAACGA
 GGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCT
 GAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG
 10 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCT
 GGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGC
 CCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
 GCCCCTcggGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGgaGTACcggAGCAC
 CAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA
 15 CACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID: 38, SpCas9(D10A)-NG-loop: Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
 20 AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATA
 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
 25 GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
 AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGC
 CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
 CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
 GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
 30 GAAAATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTGTTTCGGAAACCT
 GATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCG
 AGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
 GCTGGCCCGAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
 CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
 35 CCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCT
 GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
 ACCAGAGCAAGAACGGCTACGCCGGCTACgtgggcgcgcgacaagaagctgcgcgaagcgcagctctaaactg
 gccacagagGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCAC
 CGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACC
 40 TTCGACAACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCT
 GCGGCGGCAGGAAGATTTTTTACCCATTTCCTGAAGGACAACCGGGAAAAGATCGAGA
 AGATCCTGACCTTCGCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
 AGATTTCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGA
 GGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACT
 45 TCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAG
 TACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGGAATGAG
 AAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCA
 AGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAAT
 CGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCT
 50 GGGCACATAACCAGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
 AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGAC

AGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTCGACGACAAAGT
 GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAG
 CTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAA
 GTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGA
 5 CCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCAC
 GAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGAC
 AGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCGGGCACAAGCCCCGAGAAC
 ATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACA
 GCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGAT
 10 CCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
 ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACCTGGACATCAACCGG
 CTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCC
 ATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACG
 TGCCCTCCGAAGAGGTGCTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAAC
 15 GCCAAGCTGATTACCCAGAGAAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGGCG
 GCCTGAGCGAACTGGATAAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGG
 CAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
 CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGG
 TGTCGGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACC
 20 ACCACGCCACGACGCCTACCTGAACGCCGTGCTGGGAACCGCCCTGATCAAAAAG
 TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCG
 GAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCT
 TCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAG
 ATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGG
 25 ATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCAAGTGAAT
 ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC_gGCC
 CAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAG
 TACGGCGGCTTCG_hCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTG
 GAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCA
 30 TCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGC
 TACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGA
 GCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCC_cGCT_{tt}CTGCAGAAGGGAAA
 CGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGA
 GAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGC
 35 ACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
 ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGA
 TAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATC
 TGGGAGCCCCCT_{cgg}GCCTTCAAGTACTTTGACACCACCATCGACCGGAAG_{gt}GTAC_{cgg}
 AGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCTGT
 40 CGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID NO: 39, SpCas9(D10A)-NGX-loop: Nucleotide sequence:

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCT
 GGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGC
 45 AACACCGACCGGCACAGCATCAAGAAGAACCCTGATCGGAGCCCTGCTGTTCGACAG
 CGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
 GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATA
 AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCAC
 50 GAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACA
 AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCGGGGC

CACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCA
GCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
5 GAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCT
GATTGCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCG
AGGATACCAAACCTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
GCTGGCCCAGATCGGCGACCAAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGT
CCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCC
CCCCTGAGCGCCTCTATGATCAAGCTGTACGACGAGCACCACCAGGACCTGACCCT
10 GCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACgtgggcccgcgacaagaagctgcgcaagcgcagctctaaactg
gccacagagGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCAC
CGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACC
TTCGACAACGGCATCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCT
15 GCGGCGGCAGGAAGATTTTTACCCATTTCCTGAAGGACAACCGGGAAAAGATCGAGA
AGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
AGATTGCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGA
GAAGGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAG
20 TACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAG
AAAGCCCGCCTTCCTGAGCGGCGACCAGAAAAAGGCCATCGTGGACCTGCTGTTCA
AGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAAT
CGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCT
GGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATG
25 AGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGAC
AGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTTCGACGACAAAGT
GATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAG
CTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCTGAA
GTCCGACGGCTTCGCCAACAGAACTTCATCCAGCTGATCCACGACGACAGCCTGA
30 CCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCAC
GAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGAC
AGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAAC
ATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACA
GCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGAT
35 CCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGT
ACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGG
CTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCC
ATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACG
TGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAAC
40 GCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCG
GCCTGAGCGAACTGGATAAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGG
CAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGA
CGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGTCCAAGCTGG
TGTCCGATTTCCGGAAGGATTTCCAGTTTTTACAAAGTGCGCGAGATCAACAACCTACC
45 ACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
TACCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCG
GAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCT
TCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAG
ATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGG
50 ATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCAAGTGAAT
ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCgGCC

CAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAG
TACGGCGGCTTCGtCAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTG
GAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCA
TCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGC
5 TACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCA
GCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCcGCtttCTGCAGAAGGGAAA
CGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGA
GAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGC
ACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
10 ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGA
TAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATC
TGGGAGCCCCCTcggGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGgtGTACcgg
AGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTA
CGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

SEQ ID NO: 40, ScCas9(D10A): Nucleotide sequence:

ATGGAGAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACAGCGTGGGC
TGGGCCGTGATCACCGACGACTACAAGGTGCCAGCAAGAAGTTCAAGGTGCTGGG
CAACACCAACCGCAAGAGCATCAAGAAGAACCTGATGGGCGCCCTGCTGTTTCGACA
20 GCGGCGAGACCGCCGAGGCCACCCGCCTGAAGCGCACCGCCCGCCGCGCTACACC
CGCCGCAAGAACCGCATCCGCTACCTGCAGGAGATATTCGCCAACGAGATGGCCAA
GCTGGACGACAGCTTCTTCCAGCGCCTGGAGGAGAGCTTCCTGGTGGAGGAGGACA
AGAAGAACGAGCGCCACCCCATCTTCGGCAACCTGGCCGACGAGGTGGCCTACCAC
CGCAACTACCCCAACCATCTACCACCTGCGCAAGAAGCTGGCCGACAGCCCCGAGAA
25 GGCCGACCTGCGCCTGATCTACCTGGCCCTGGCCACATCATCAAGTTCGCGGCCA
CTTCTGATCGAGGGCAAGCTGAACGCCGAGAACAGCGACGTGGCCAAAGCTGTTCT
ACCAGCTGATCCAGACCTACAACCAGCTGTTTCGAGGAGAGCCCCCTGGACGAGATC
GAGGTGGACGCCAAGGGCATCCTGAGCGCCCGCCTGAGCAAGAGCAAGCGCCTGG
AGAAGCTGATCGCCGTGTTCCCCAACGAGAAGAAGAACGGCCTGTTTCGGCAACATC
30 ATCGCCCTGGCCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGACCGA
GGACGCCAAGCTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACGAGCTG
CTGGGCCAGATCGGCGACCAGTACGCCGACCTGTTACGCGCCGCCAAGAACCTGAG
CGACGCCATCCTGCTGAGCGACATCCTGCGCAGCAACAGCGAGGTGACCAAGGCC
CCCTGAGCGCCAGCATGGTGAAGCGCTACGACGAGCACCAACAGGACCTGGCCCTG
35 CTGAAGACCCTGGTGCGCCAGCAGTTCCCCGAGAAGTACGCCGAGATATTCAAGGA
CGACACCAAGAACGGCTACGCCGGCTACGTGGGCATCGGCATCAAGCACCGCAAGC
GCACCACCAAGCTGGCCACCCAGGAGGAGTTCTACAAGTTCATCAAGCCCATCTTG
GAGAAGATGGACGGCGCCGAGGAGCTGCTGGCCAAGCTGAACCGCGACGACCTGC
TGCGCAAGCAGCGCACCTTCGACAACGGCAGCATCCCCACCAGATCCACCTGAAG
40 GAGCTGCACGCCATCCTGCGCCGCCAGGAGGAGTTCTACCCCTTCCTGAAGGAGAA
CCGCGAGAAGATCGAGAAGATCCTGACCTTCCGCATCCCCCTACTACGTGGGCCCCC
TGGCCCCGCGCAACAGCCGCTTCGCCTGGCTGACCCGCAAGAGCGAGGAGGCCATC
ACCCCTTGGAACCTTCGAGGAGGTGGTGGACAAGGGCGCCAGCGCCAGAGCTTCAT
CGAGCGCATGACCAACTTCGACGAGCAGCTGCCCAACAAGAAGGTGCTGCCCAAGC
45 ACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAGGTGAAGTAC
GTGACCGAGCGCATGCGCAAGCCCGAGTTCTGAGCGGCGAGCAGAAGAAGGCCA
TCGTGGACCTGCTGTTCAAGACCAACCGCAAGGTGACCGTGAAGCAGCTGAAGGAG
GACTACTTCAAGAAGATCGAGTGCTTCGACAGCGTGGAGATCATCGGCGTGGAGGA
CCGCTTCAACGCCAGCCTGGGCACCTACCACGACCTGCTGAAGATCATCAAGGACA
50 AGGACTTCTGAGACAACGAGGAGAACGAGGACATCCTGGAGGACATCGTGCTGACC
CTGACCCTGTTTCGAGGACCGCGAGATGATCGAGGAGCGCCTGAAGACCTACGCCCA

CCTGTTTCGACGACAAGGTGATGAAGCAGCTGAAGCGCCGCCACTACACCGGCTGGG
 GCCGCCTGAGCCGCAAGATGATCAACGGCATCCGCGACAAGCAGAGCGGCAAGAC
 CATCCTGGACTTCCTGAAGAGCGACGGCTTCAGCAACCGCAACTTCATGCAGCTGA
 TCCACGACGACAGCCTGACCTTCAAGGAGGAGATCGAGAAGGCCCAGGTGAGCGG
 5 CCAGGGCGACAGCCTGCACGAGCAGATCGCCGACCTGGCCGGCAGCCCCGCCATCA
 AGAAGGGCATCCTGCAGACCGTGAAGATCGTGGACGAGCTGGTGAAGGTGATGGG
 CCACAAGCCCCGAGAACATCGTGATCGAGATGGCCCCGCGAGAACCAGACCACCACC
 AAGGGCCTGCAGCAGAGCCGCGAGCGCAAGAAGCGCATCGAGGAGGGGCATCAAGG
 AGCTGGAGAGCCAGATCCTGAAGGAGAACCCCGTGGAGAACACCCAGCTGCAGAA
 10 CGAGAAGCTGTACCTGTACTACCTGCAGAACGGCCGCGACATGTACGTGGACCAGG
 AGCTGGACATCAACCGCCTGAGCGACTACGACGTGGACCACATCGTGCCCCAGAGC
 TTCATCAAGGACGACAGCATCGACAACAAGGTGCTGACCCGCGAGCGTGGAGAACCG
 CGGCAAGAGCGACAACGTGCCCAGCGAGGAGGTGGTGAAGAAGATGAAGAACTAC
 TGGCGCCAGCTGCTGAACGCCAAGCTGATCACCCAGCGCAAGTTCGACAACCTGAC
 15 CAAGGCCGAGCGCGGCGGCTGAGCGAGGCGGACAAGGCCGGCTTCATCAAGCGC
 CAGCTGGTGGAGACCCGCCAGATCACCAGCACGTGGCCCGCATCCTGGACAGCCG
 CATGAACACCAAGCGCGACAAGAACGACAAGCCCATCCGCGAGGTGAAGGTGATC
 ACCCTGAAGAGCAAGCTGGTGAAGCGACTTCCGCAAGGACTTCCAGCTGTACAAGGT
 GCGCGACATCAACAACCTACCACCACGCCACGACGCCTACCTGAACGCCGTGGTGG
 20 GCACCGCCCTGATCAAGAAGTACCCCAAGCTGGAGAGCGAGTTCGTGTACGGCGAC
 TACAAGGTGTACGACGTGCGCAAGATGATCGCCAAGAGCGAGCAGGAGATCGGCA
 AGGCCACCGCCAAGCGCTTCTTCTACAGCAACATCATGAACTTCTTCAAGACCGAG
 GTGAAGCTGGCCAACGGCGAGATCCGCAAGCGCCCCCTGATCGAGACCAACGGCG
 AGACCGGCGAGGTGGTGTGGAACAAGGAGAAGGACTTCGCCACCGTGCGCAAGGT
 25 GCTGGCCATGCCCCAGGTGAACATCGTGAAGAAGACCGAGGTGCAGACCGGCGGCT
 TCAGCAAGGAGAGCATCCTGAGCAAGCGCGAGAGCGCCAAGCTGATCCCCCGCAA
 GAAGGGCTGGGACACCCGCAAGTACGGCGGCTTCGGCAGCCCCACCGTGGCCTACA
 GCATCCTGGTGGTGGCCAAGGTGGAGAAGGGCAAGGCCAAGAAGCTGAAGAGCGT
 GAAGGTGCTGGTGGGCATCACCATCATGGAGAAGGGCAGCTACGAGAAGGACCCC
 30 ATCGGCTTCCTGGAGGCCAAGGGCTACAAGGACATCAAGAAGGAGCTGATCTTCAA
 GCTGCCCAAGTACAGCCTGTTTCGAGCTGGAGAACGGCCGCCGCCGCATGCTGGCCA
 GCGCCACCGAGCTGCAGAAGGCCAACGAGCTGGTGTGCCCCAGCACCTGGTGGCG
 CTGCTGTACTACACCCAGAACATCAGCGCCACCACCGGCAGCAACAACCTGGGCTA
 CATCGAGCAGCACCGCGAGGAGTTCAAGGAGATATTCGAGAAGATCATCGACTTCA
 35 GCGAGAAGTACATCCTGAAGAACAAGGTGAACAGCAACCTGAAGAGCAGCTTCGA
 CGAGCAGTTCGCCGTGAGCGACAGCATCCTGCTGAGCAACAGCTTCGTGAGCCTGC
 TGAAGTACACCAGCTTCGGCGCCAGCGGCGGCTTCACCTTCCTGGACCTGGACGTG
 AAGCAGGGCCGCTGCGCTACCAGACCGTGACCGAGGTGCTGGACGCCACCCTGAT
 CTACCAGAGCATCACCGGCCTGTACGAGACCCGCACCGACCTGAGCCAGCTGGGCG
 40 GCGAC

SEQ ID NO: 41, TadA-TadA*: Nucleotide sequence:

TCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGG
 CAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACAA
 45 CAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACC
 GCACACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTACCG
 CCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAG
 CAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGAC
 CGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATGAACCACCGGG
 50 TGGAGATCACAGAGGGAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTC
 TTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGAGCTCCACCG

ACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGA
 GTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGT
 TTTCCACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGGGCACGCGAT
 GAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCG
 5 AGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCCCATGCCGAAATTATG
 GCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCACCCT
 GTACGTGACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGAT
 CGGCCGCGTGGTGTGTTGGCGTGAGGAACGCAAAAACCGGGCGCCGAGGCTCCCTGA
 TGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTCGAAATTACCGAGGGGAATC
 10 CTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTTCGGATGCCTAGACAGGTG
 TTCAATGCTCAGAAGAAGGCCAGAGCTCCACCGAC

SEQ ID: 42, TadA*(A56G_V82G): Nucleotide sequence:

TCTGAGGTGGAGTTTTTCCACGAGTACTGGATGAGACATGCCCTGACCCTGG
 15 CCAAGAGGGGCACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGCTGGTGCTGAA
 CAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACA
 GgCCATGCCGAAATTATGGCCCTGAGgCAGGGCGGCCTGGTCATGCAGAACTACAG
 ACTGATTGACGCCACCCTGTACgGACATTCGAGCCTTGCGTGATGTGCGCCGGCGC
 CATGATCCACTCTAGGATCGGCCGCGTGGTGTGTTGGCGTGAGGAACGCAAAAACCG
 20 GCGCCGAGGCTCCCTGATGGACGTGCTGCACTACCCCGGCATGAATCACCGCGTC
 GAAATTACCGAGGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCTATTTCTTT
 CGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGGCCAGAGCTCCACCGAC

SEQ ID NO: 43, Gp41-1-N: Nucleotide sequence:

TGTTTGGATCTGAAAACGCAAGTTCAAACGCCACAGGGTATGAAAGAAATAT
 25 CCAATATACAGGTCGGCGATCTCGTCTTGTCTAACTGGCTATAACGAGGTGCTGA
 ATGTATTTCCAAAAAGCAAGAAAAAAGTTACAAGATAACTCTGGAAGATGGA
 AGAAATTATCTGTTCTGAGGAGCATCTGTTTCCGACCCAACAGGGGAGATGAATA
 TCAGTGGCGGTCTCAAAGAGGGTATGTGTTTGTATGTCAAGGaataa
 30

SEQ ID NO: 44, Gp41-1-C: Nucleotide sequence:

ATGATGCTCAAGAAGATCCTCAAGATTGAAGAGTTGGACGAGCGCGAGCTT
 ATAGACATAGAAGTCAGTGGTAATCACCTTTTCTACGCAAATGACATTTTGACTCAC
 AACTCC
 35

SEQ ID NO: 45, Cfa-N: Nucleotide sequence:

TGTCTCAGTTATGAcACCGAAATCCTGACAGTCGAGTATGGAiTtCTGCCGATC
 GGCAAGATTGTGGAGgAGAGAATTGAATGTACGGTCTATAcgGTCGACAAGaATGGTt
 tCgTCTACACCCAACCAaTTGCTCAATGGCAtaATCGAGGGGAGCAGGAGGTGTTTGA
 40 GTATTGCTGGAGGACGGGTCAaTCATTAGAGCTACAAAGGACCATAAGTTTATGAC
 AacCGATGGTCAAATGCTGCCGATAGATGAAATATTCGAAAGGGgACTGGATCTTAA
 GCaAGTCGATggCCTTCCAaac

SEQ ID NO: 46, Cfa-C, Nucleotide sequence:

ATGgTcAAgATTatCAgcCGCAAATcCTTGGGGAcACAGAATGTATATGACATCG
 GCGTGGAAaaGGATCACAATTTTctgCTGAAGAATGGTcTTgTTGCTTccAAAt
 45

SEQ ID NO: 47, TadA,

SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP
 50 TAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
 GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

SEQ ID NO: 48

TGAGTATCATCGTGTGAAAGCTgAGGGGACGAGGCAGGCCTATAA

SEQ ID NO: 49

5 AAAAACATCAACTTCAGCCATC_cATTTCTTCAGGGTTTGTATGTG

SEQ ID NO: 50

TATATCATAATGAAAACGCCGC_cATTTCTCAACAGATCTGTCAAA

SEQ ID NO: 51

CTGTGTGAAATGGCTGCAAATCgATGGTTGAGCTCTGAGATTTGG

10 SEQ ID NO: 52

GTTCTGCTTTTGCTACTACTCA_cGTTTCCATGTTGTCCCCCTCTA

SEQ ID NO: 53

ATTTTATGGCCTTTTGCAACTCgACCAGAAAAAAGCAGCTTTGG

SEQ ID NO: 54

15 TGAGGAGATCGCCACGGGCTG_cCAGGATCCCTTGATCACCTCAG

SEQ ID NO: 55

GGCTGCTCTGTCAGAAATATTCgTACAGTCTCAAGAGTACTCATG

SEQ ID NO: 56

GTGTAGGCATAGCTCTTGAATCgAGGCTTAGGGGAAGAAGTTCTC

20 SEQ ID NO: 57

CCTGTTCTTCAGTAAGACGTTG_cCATTTGAGAAGGATGTCTTGTA

SEQ ID NO: 58

GCCATTTTAGGCTTTTTACTTA_cTTGTCTGTAGCTCTTTCTCTCT

SEQ ID NO: 59

25 CTAGTTTCTCACACATGACACA_cCTGTTCTTCAGTAAGACGTTGC

SEQ ID NO: 60

GTGAAGTTGATTACATTAACCTgTGGATAATTACGAGTTGATTGT

SEQ ID NO: 61

ttttgtatATCTGAGTTAAACTgCTCCAATTCCTTCAAAGGAATG

30 SEQ ID NO: 62

TCTGCAATATAAGCTGCCAACTgCTTGTCATGAATGTGAGGGAC

SEQ ID NO: 63

GGACTGGGGTTCCAGTCTCATC_cAGTCTAGGAAGAGGGCCGCTTC

SEQ ID NO: 64

AGTCGTTGTGTGGCTGACTGCTgGCAAACCACACTATTCCAGTCA

SEQ ID NO: 65

ATTTGtgtcttctgagaaactgttcaGCTTCTGTTAGCCACTGA

5 SEQ ID NO: 66

AAGAACCCAGCGGTCTTCTGTCcATCTACAGATGTTTGCCCATCG

SEQ ID NO: 67

AGACTTTTTCCGAAGTTCACCTCcACTTGAAGTTCATGTTATCCAA

SEQ ID NO: 68

10 TACCTGTTGGCACATGTGATCCcACTGAGTGTTAAGTTCTTTGAG

SEQ ID NO: 69

CAAAGGGCCTTCTGCAGTCTTCgGAGTTTCATGGCAGTCCTATAA

SEQ ID NO: 70

ATGGTTAATGTCTAACCTTTATcCACTGGAGATTTGTCTGCTTGA

15 SEQ ID NO: 71

CAAAATAATCTGACCTTAAGTTgTTCTTCCAAAGCAGCAGTTGCG

SEQ ID NO: 72

CATCTACAGATGTTTGCCCATCgATCTCCCAATACCTGGAGAAGA

SEQ ID NO: 73

20 AAGGTGTTCTTGTACTTCATCCcACTGATTCTGAATTCTTTCAac

SEQ ID NO: 74

GAAGGTGTTCTTGTACTTCATCcCACTGATTCTGAATTCTTTCAa

SEQ ID NO: 75

GCTATGCTTTGAATTTTAAATCgTTCAATTTGAGGTTGAAGATCT

25 SEQ ID NO: 76

TCTAGGAGGCGCCTCCCATCCTgTAGGTCACCTGAAGAGGTTCTCA

SEQ ID NO: 77

GTGTAATTACCATTCACCATCTgTTCCACCAGGGCCTGAGCTGAT

SEQ ID NO: 78

30 TGAGCATGCTTTACCAGGATCTgTTCCCTTGTTGGTCACCGTAGTT

SEQ ID NO: 79

CTACTGTATAGGGACCCTCCTTcCATGACTCAAGCTTGGCTCTGG

SEQ ID NO: 80

CAGCTTCTTCCTTAGCTTCCAGcCATTGTGTTGAATCCTTTAACA

SEQ ID NO: 81

AGCTGCCCAAGGTCTTTTATTTgAGCTTCAATTTCTCCTTGTTTC

SEQ ID NO: 82

TGCCAGTAACAACCTCACAATTTgTGCAAAGTTGAGTCTTCGAAAC

5

SEQ ID NO: 83

AGAGTAACAGTCTGAGTAGGAGctaaaatattttgggtttttgca

SEQ ID NO: 84

CTTCAGCAAAAAAAGTACTCACgCAGAATCTACTGGCCAGAAGTT

SEQ ID NO: 85

10

ACATCTACATTTGTCTGCCACTgGCGGAGGTCTTTGGCCAACTGC

SEQ ID NO: 86

CTGAGATAGTATAGGCCACTTTgTTGCTCTTGCAGAGAACTTTGT

SEQ ID NO: 87

TAGCTGTCCTTTACACACTTTAcCTGTTGAGAATAGTGCATTTGA

15

SEQ ID NO: 88

GCCTGGGCTTCCTGAGGCATTTgAGCTGCGTCCACCTTGTCTGCA

SEQ ID NO: 89

TGCACTGGCAGGTAGCCCATTCgGGGATGCTTCGCAAAATACCTT

SEQ ID NO: 90

20

TTATAGTTCCACATTCAATTACcTCTGGGCTCCTGGTAGAGTTTC

SEQ ID NO: 91

CGTCAGGCTGGCGTCAAACCTTAcCGGAGTGCAATATTCCACCATG

SEQ ID NO: 92

tcaTCCAAAAGTGTGTCAGCCTgAATGATCCACTTTGTGATGTGG

25

SEQ ID NO: 93

TGTAGCCACACCAGAAGTTCCTgCAGAGAAAGGTGCAGACGCTTC

SEQ ID NO: 94

TGTAAGGATTTTTTCAGTCTCCTgGGCAGACTGGATGCTCTGTTCA

SEQ ID NO: 95

30

AGAATGGGATCCAGTATACTTAcAGGCTCCAATAGTGGTCAGTCC

SEQ ID NO: 96

TTCAGAGGCGCAATTTCTCCTCgAAGTGCCTGTGTGCAATAGTCA

SEQ ID NO: 97

GTTTCTTCCAAAGCAGCCTCTCgCTCACTCACCTGCAAAGGACC

SEQ ID NO: 98

CTGAACTTCTCAGCTTTTTCTCgCTCTATGGCCTGCAGCATGAGA

SEQ ID NO: 99

AGATTTAACCACTCTTCTGCTCgGGAGGTGACAGCTATCCAGTTA

5 SEQ ID NO: 100

tgttttatctttatttcctctCgCTTTCTCTCATCTGTGATTCTT

SEQ ID NO: 101

CCAGCTGGGAGGAGAGCTTCTTcCAGCGTCCCTCAATTTCTTCAA

SEQ ID NO: 102

10 ACACAGCTTCTGAGCGAGTAATcCAGCTGTGAAGTTCAGTTATAT

SEQ ID NO: 103

AAGTAAACGGTTTACCGCCTTcACTCAGAGCTCAGATCTTCTAA

SEQ ID NO: 104

aaatagaaaaattagatgacttgccaaaggtcacaaaGGTAACTG

15 SEQ ID NO: 105

TTGACTTTCTCGAGGTGATCTTgGAGAGAGTCAATGAGGAGATCG

SEQ ID NO: 106

TCTAAAATCATCTTACTTTCTTgTAGACGCTGCTCAAAATTGGCT

SEQ ID NO: 107

20 GAATTGACCCTGACTTGTCTTgTTCTAGATCTTCTTGAAGCACc

SEQ ID NO: 108

TGGATGGCTTCAATGCTCACTTgTTGAGGCAAACTTGGAAGAGT

SEQ ID NO: 109

AACAGTCCTCTACTTCTTCCCAcCAAAGCATTTTGAAAAGTGTAT

25 SEQ ID NO: 110 AGGCCTCCTTTCTGGCATAGACcttcacaaaacaaacaaacaaa

SEQ ID NO: 111

TTTGGTTTCTGACTGCTGGACCcATGTCCTGATGGCACTCATGGT

SEQ ID NO: 112

ATCTTACTTTCTTGTAGACGCTgCTCAAAATTGGCTGGTTTCTGG

30 SEQ ID NO: 113

AGATTTTCACTTATCTTCATAcCTCTTCATGTAGTTCCTCCAA

SEQ ID NO: 114

CTGTTCAAGTTGTTCTGAGGCTTgTTTGATGCTATCTGCATTAACA

SEQ ID NO: 115

CAGCATTAATATACACGACTTA_cATCTGTACTTGTCTTCCAAATG

SEQ ID NO: 116

TCATGACTTGTCAAATCAGATT_gGATTTTCTGTTGGGAGGATAGC

5 SEQ ID NO: 117

ATCTGCTCCAATTGTTGTAGCT_gATTATAGAAAGCGATGATGTTG

SEQ ID NO: 118

CAAATTTGCTCTCAATTTCCCG_cCAGCGCTTGCTGAGCTGGATCT

SEQ ID NO: 119

10 CATTCAAAGCCAGGCCATCAGAcCAGCTGGTGGTGAAGTTGATTA

SEQ ID NO: 120

ttcatcTCTTCAACTGCTTTCT_gTAATTCATCTGGAGTTTTATAT

SEQ ID NO: 121

ATTGAAAGCTAGAAAGTACATA_cGGCCAGTTTTTTGAAGACTTGAT

15 SEQ ID NO: 122

TTCAAATACTGGCCAATACTTA_cAGCAAAGGGCCTTCTGCAGTCT

SEQ ID NO: 123

GTTGTCTGTGTTAGTGATGGCT_gAGTGGTGGTGACAGCCTGTGAA

SEQ ID NO: 124

20 TCATCAGCCTGCCTCTTGTACT_gATACCACTGATGAGAAATTTCT

SEQ ID NO: 125

TACTGTATAGGGACCCTCCTTC_cATGACTCAAGCTTGGCTCTGGC

SEQ ID NO: 126

ATGTTGAATGCATGTTCCAGTC_gTTGTGTGGCTGACTGCTGGCAA

25 SEQ ID NO: 127

TGCCATTTGAGAAGGATGTCTT_gTAAAAGAACCCAGCGGTCTTCT

SEQ ID NO: 128

GAGACTTTTTCCGAAGTTCAC_{Tc}CACTTGAAGTTCATGTTATCCA

SEQ ID NO: 129

30 TTAGCAACTGGCAGAATTCGAT_cCACCGGCTGTTCAAGTTGTTCTG

SEQ ID NO: 130

TTGCCACATCTACATTTGTCTG_cCACTGGCGGAGGTCTTTGGCCA

SEQ ID NO: 131

TCCCATTCAGCCTAGTGCAGAG_cCACTGGTAGTTGGTGGTTAGAG

SEQ ID NO: 132

GACTTACTGGAAAGAAAGTGCTgAGATGCTGGACCAAAGTCCCTG

SEQ ID NO: 133

TTTAATCGTTCAATTTGAGGTTgAAGATCTGATAGCCGGTTGACT

5 SEQ ID NO: 134

AAAGAGATTGTCTATACCTGTTgGCACATGTGATCCCACTGAGTG

SEQ ID NO: 135

AAGTTTTTGGACTAAATTATCCcAACACCGGGCAAAGTTATCCAG

SEQ ID NO: 136

10 AGCTCAGCATCCCGGGGACTCTgGGGAGAGGTGGGCATCATTTC

SEQ ID NO: 137

TTGTCCCCCTCTAAGACAGTCTgCACTGGCAGGTAGCCCATTCGG

SEQ ID NO: 138

TTCGCAAAATACCTTTTGGTTCgAAATTGTTTTTTAGTACCTTG

15 SEQ ID NO: 139

TGCAACTCGACCAGAAAAAAAGcAGCTTTGGCAGATGTCATAATT

SEQ ID NO: 140

TTGCAGATGTTACATTTGGCCTgATGCTTGGCAGTTTCTGCAGCA

SEQ ID NO: 141

20 AAATAAAAACATGCCATACGTAcGTATCATAAACATTCAGCAGCC

SEQ ID NO: 142

TACTTACAGCAAAGGGCCTTCTgCAGTCTTCGGAGTTTCATGGCA

SEQ ID NO: 143

TTGACCTCCTCAGCCTGCTTTCgTAGAAGCCGAGTGACATTCTGG

25 SEQ ID NO: 144

ATTCAATTACCTCTGGGCTCCTgGTAGAGTTTCTCTAGTCCTTCC

SEQ ID NO: 145

AATGCCTGACTTACTTGCCATTgTTTCATCAGCTCTTTTACTCCC

SEQ ID NO: 146

30 TGTACTTCATCCCACTGATTCTgAATTCTTTCaactagaataaaa

SEQ ID NO: 147

TGCTTCATTACCTTCACTGGCTgAGTGGCTGGTTTTTCCTTGTAC

SEQ ID NO: 148

TTTAATTGTTTGAGAATTCCCTgGCGCAGGGGCAACTCTTCCACC

SEQ ID NO: 149

atatgtgttaCCTACCCTTGTCgGTCCTTGTacattttgttaact

SEQ ID NO: 150

TCAAGCTGGGAGAGAGCTTCCTgTAGCTTCACCCTTTCCACAGGC

5

SEQ ID NO: 151

ATGTCAATCCGACCTGAGCTTTgTTGTAGACTATCTTTTATATTc

SEQ ID NO: 152

GTTGTAGACTATCTTTTATATTctgtaataaaaaattttaaaac

SEQ ID NO: 153

10

TCCCGCCAGCGCTTGCTGAGCTgGATCTGAGTTGGCTCCACTGCC

SEQ ID NO: 154 TTATGTTTTGTCTGTAAACAGCTgctgttttatctttatttcctct

SEQ ID NO: 155

TGTTTTGTCTGTAAACAGCTgctgttttatctttatttcctctCGC

SEQ ID NO: 156

15

TTTCTCTCATCTGTGATTCTTTgTTGTAAGTTGTCTCCTCTTTGC

SEQ ID NO: 157

aCCTTAAGCACGTCTTCTTTTTgCtggggtttcttttctctgat

SEQ ID NO: 158

TTAAGCACGTCTTCTTTTTGtggggtttcttttctctgattca

20

SEQ ID NO: 159

ATACTCTTCAGGTGCACCTTCTgTTTCTCAATCTCTTTTTGAGTA

SEQ ID NO: 160

AGCTGTGACTGTACTACTTCCTgTTCCACACTCTTTGTTTCCAAT

SEQ ID NO: 161

25

TTGGCTGGTTTCTGGAATAATCgAAACTTCATGGAGACATCTTGT

SEQ ID NO: 162

TGCATCTCTGATAGATCTTCTgGAGGCTTACAGTTTTCTCCAAA

SEQ ID NO: 163

ACAGTGAAAGAGATTGTCTATAcCTGTTGGCACATGTGATCCCAC

30

SEQ ID NO: 164

AAGGCATCATATAAAAATCTTAcTCTGCACTGTTTCAGCTGCTTT

SEQ ID NO: 165

CTTACTCTGCACTGTTTCAGCTgCTTTTTTAGAATTTCTGAATCC

SEQ ID NO: 166

TCTTGAATTACCTGAATTTTTCgGAGTTTATTCATTTGCTCCTCT

SEQ ID NO: 167

TGTTGCTCTTGCAGAGAACTTTgTAAAgcctaaaaaacaattttt

5

SEQ ID NO: 168

ATTGGTGGCAAAGTGTCAAAAACtTtatcaaaagggaaaaaagaa

SEQ ID NO: 169

TAGGCTTTTTACTTACTTGTCTgTAGCTCTTCTCTCTGGCCTGC

SEQ ID NO: 170

10

GCCTGCACATCAGAAAAGACTTgCTTAAAATGATTTGTAAAGGCC

SEQ ID NO: 171

ATGGAAGGAGAAGAGATTCTTAcCTTACAAATTTTTAACTGACTT

SEQ ID NO: 172

GGTGGTGGGTTGGATTTTCAACcAGTTTTTCAGCAGTAGTTGTCAT

15

SEQ ID NO: 173

TCGATCCACCGGCTGTTCAgTTCTGAGGCTTGTTTGATGCTA

SEQ ID NO: 174

TGAGCTGATCTGCTGGCATCTTgCAGTTTTCTGAACTTCTCAGCT

SEQ ID NO: 175

20

ATAAAAGCTTAAGATGCTCTCAcCTTTTCCTAATTTCAGAATCCA

SEQ ID NO: 176

ATTTCAGAATCCACAGTAATCTgCCTCTTCTTtggggaggtggt

SEQ ID NO: 177

TGATAATTGGTATCACTAACCTgTGCTGTACTCTTTTCAAGTTTT

25

SEQ ID NO: 178

TCCAGCCATGCTTCCGTCTTCTgGGTCACTGACTTATTCTTCAGT

SEQ ID NO: 179

GAAGGATGTCTTGTAAGAAGcCAGCGGTCTTCTGTCCATCTAC

SEQ ID NO: 180

30

TGTTCTTGTTCTAGATCTTCTTgAAGCACctgaaagataaaatgt

SEQ ID NO: 181

CCTACCTTATGTTGTTGTACTTgGCGTTTTAGGTCTTCAAGATCA

SEQ ID NO: 182

tctttctctgtttttgttagCcAGTCATTCAACTCTTTCAGTTT

SEQ ID NO: 183

attaaaaacaaataaggacTTAcTTGCTTTGTTTTTCCATGCTAG

SEQ ID NO: 184

caaataaggacTTACTTGCTTTgTTTTTCCATGCTAGCTACCCTG

5

SEQ ID NO: 185

TTTAGGAGATTCATCTGCTCTgtgacttcagttttcatcttct

SEQ ID NO: 186

ACATCATTAGAAATCTCTCCTTgTGCTTGCAATGTGTCCTCAGCA

SEQ ID NO: 187

10

TGGTAGTCCAGAAATTTACCAAcCTTCAGGATCGAGTAGTTTCTC

SEQ ID NO: 188

tatitttcattacatttttgaCcTACATGtggaataaaattttca

SEQ ID NO: 189

CCATTCATCAGGATTCTTACCTgCCAGTGGAGGATTATATTCCAA

15

SEQ ID NO: 190

tttcttaaaaataagteaCATAcCAGTTTTTGGCCCTGTCAGGCCT

SEQ ID NO: 191

aataagtcaCATACCAGTTTTTgCCCTGTCAGGCCTTCGAGGAGG

SEQ ID NO: 192

20

GTAAAGTAAACAAACCATTCTTAcCTTAGAAAATTGTGCATTTACC

SEQ ID NO: 193

TTTACTAAGCAAAATAATCTGAcCTTAAGTTGTTCTTCCAAAGCA

SEQ ID NO: 194

ACGGATCCTCCCTGTTCGTCCCcTATTATGAAGAATCAAAGCAGA

25

SEQ ID NO: 195

TTCTCAACAGATCTGTCAAATCgCCTGCAGGTAAAAGCATATGGA

SEQ ID NO: 196

CTGTCAAATCCATCATGTACCCcTGACAAAGAAGGAAGTTAACAA

SEQ ID NO: 197 TCTCAATATGCtgcttcccaaactgaaattaaaaaaatacactc

30

SEQ ID NO: 198

CTTAATTCATCATCTTTCAGCTgTAGCCACACCAGAAGTTCCTGC

SEQ ID NO: 199

GTCAAGACATTCATTTCTTTTCgCATCTTACGGGACAATTTCAAG

SEQ ID NO: 200

TGTGTCCTCAGCAGAAAGAAGCcACGATAATACTTCTTCTAAAGC

SEQ ID NO: 201

TAGAAAGCGATGATGTTGTTCTgATACTCCAGCCAGTTAAGTCTC

5 SEQ ID NO: 202

CTCTCTAAGGAAATCAAGATCTgGGCAGGACTACGAGGCTGGCTC

SEQ ID NO: 203

TCAAAAGTTTCCATGTGTTTCTgGTATTCCTTAATTGTACAGAGA

SEQ ID NO: 204

10 ACTGTTTCCATTACAGTTGTCTgTGTTAGTGATGGCTGAGTGGTG

SEQ ID NO: 205

TTTAGTACCTTGGCAAAGTCTCgAACATCTTCTCCTGATGTAGTC

SEQ ID NO: 206

ATTTGTGCAAAGTTGAGTCTTCgAAACTGAGCAAATTTGCTCTCA

15 SEQ ID NO: 207

TGGGGACGCCTCTGTTCCAAATcCTGCATTGTTGCCTGTAAGAAC

SEQ ID NO: 208

GTCTCCTATGAACTCGAGAAGCcGCAAAaccaaggaagagaaaga

SEQ ID NO: 209

20 GAGAGTTTGGTTTCTGACTGCTgGACCCATGTCCTGATGGCACTC

SEQ ID NO: 210

TGCGTATTTGCCACCAGAAATAcATACCACACAATGATTTAGCTG

SEQ ID NO: 211

TTTGGGTTATCCTCTGAATGTCgCATCAAATTTTCAAGTGACTGA

25 SEQ ID NO: 212

AGGACACGGATCCTCCCTGTTCgTCCCCTATTATGAAGAATCAAA

SEQ ID NO: 213

TGCTTGTTAAAAAACTTACTTCgATCCGTAATGATTGTTCTAGCC

SEQ ID NO: 214

30 TTTTGCTCCACATCTTTTCCTAcCTAATGTTGAGAGACTTTTTC

SEQ ID NO: 215

CCTGCCAGTGGAGGATTATATTcCAAATCAAACCAAGAGTCAGTT

SEQ ID NO: 216

GGATAATTACGAGTTGATTGTCgGACCCAGCTCAGGAGAATCTTT

	SEQ ID NO: 217	
	TTTAGACTGGGCTGAATTGTCTgAATATCACTGACTAAAagctaa	
	SEQ ID NO: 218	
	GTACTACTTACATTATTGTTCTgCAAAACCCGCAGTGCCTTGTTG	
5	SEQ ID NO: 219	
	TTCATTTGCTCCTCTAGCTTTTgACAATGCTCAACCAGCTGGGAG	
	SEQ ID NO: 220	
	TCAATCTGAGACAGGACTCTTTgGGCAGCCTCCTTCCCCTGAtta	
	SEQ ID NO: 221	TTcagctcctcttttcttctgcaaTTCCCGATCAATTTCCCTAT
10	SEQ ID NO: 222	AAAGCTaAGGGGACGAGGCAGGC
	SEQ ID NO: 223	GAAAGCTaAGGGGACGAGGCAGG
	SEQ ID NO: 224	AAATaGATGGCTGAAGTTGATGT
	SEQ ID NO: 225	GAAATaGATGGCTGAAGTTGATG
	SEQ ID NO: 226	AAGAAATaGATGGCTGAAGTTGA
15	SEQ ID NO: 227	CTGAAGAAATaGATGGCTGAAGTTGAT
	SEQ ID NO: 228	AAATaGCGGCGTTTTTCATTATGA
	SEQ ID NO: 229	GAGAAATaGCGGCGTTTTTCATTATGAT
	SEQ ID NO: 230	AAATCaATGGTTGAGCTCTGAGA
	SEQ ID NO: 231	GCAAATCaATGGTTGAGCTCTGA
20	SEQ ID NO: 232	TGCAAATCaATGGTTGAGCTCTGAGAT
	SEQ ID NO: 233	AACaTGAGTAGTAGCAAAAGCAG
	SEQ ID NO: 234	GAAACaTGAGTAGTAGCAAAAGC
	SEQ ID NO: 235	GGAAACaTGAGTAGTAGCAAAAG
	SEQ ID NO: 236	TGGAAACaTGAGTAGTAGCAAAA
25	SEQ ID NO: 237	AACTCaACCAGAAAAAAAGCAGC
	SEQ ID NO: 238	CAACTCaACCAGAAAAAAAGCAG
	SEQ ID NO: 239	AAGGGATCCTGaCAGCCCGTG GGCGAT
	SEQ ID NO: 240	AATATTCaTACAGTCTCAAG AGT
	SEQ ID NO: 241	AATCaAGGCTTAGGGGAAGAAGT
30	SEQ ID NO: 242	GAATCaAGGCTTAGGGGAAGAAG
	SEQ ID NO: 243	TGAATCaAGGCTTAGGGGAAGAA
	SEQ ID NO: 244	TTGAATCaAGGCTTAGGGGAAGA
	SEQ ID NO: 245	CTTGAATCaAGGCTTAGGGGAAGAAGT
	SEQ ID NO: 246	AATGaCAACGTCTTACTGAA GAA

	SEQ ID NO: 247	AAATGaCAACGTCTTACTGA AGA
	SEQ ID NO: 248	CAAATGaCAACGTCTTACTG AAG
	SEQ ID NO: 249	TCAAATGaCAACGTCTTACT GAA
	SEQ ID NO: 250	ACAAaTAAGTAAAAAGCCTA AAA
5	SEQ ID NO: 251	ACAGaTGTGTCATGTGTGAG AAA
	SEQ ID NO: 252	AACAGaTGTGTCATGTGTGA GAA
	SEQ ID NO: 253	GAACAGaTGTGTCATGTGTG AGA
	SEQ ID NO: 254	AGAACAGaTGTGTCATGTGT GAG
	SEQ ID NO: 255	ACCTaTGGATAATTACGAGT TGA
10	SEQ ID NO: 256	AACCTaTGGATAATTACGAG TTG
	SEQ ID NO: 257	TTAACCTaTGGATAATTACG AGT
	SEQ ID NO: 258	TTAACCTaTGGATAATTACGA GTTGAT
	SEQ ID NO: 259	ACTaCTCCAATTCCTTCAAA GGA
	SEQ ID NO: 260	AACTaCTCCAATTCCTTCAA AGG
15	SEQ ID NO: 261	AAACTaCTCCAATTCCTTCAA AGGAAT
	SEQ ID NO: 262	ACTaCTTGTCAATGAATGTG AGG
	SEQ ID NO: 263	AACTaCTTGTCAATGAATGT GAG
	SEQ ID NO: 264	CAACTaCTTGTCAATGAATG TGA
	SEQ ID NO: 265	CCAACTaCTTGTCAATGAAT GTG
20	SEQ ID NO: 266	GCCAACTaCTTGTCAATGAA TGT
	SEQ ID NO: 267	ACTaGATGAGACTGGAACCC CAG
	SEQ ID NO: 268	TAGACTaGATGAGACTGGAAC CCCAGT
	SEQ ID NO: 269	ACTGCTaGCAAACCACTAT TCCAGT
	SEQ ID NO: 270	agaaactattcaGCTTCTGT TAG
25	SEQ ID NO: 271	AGATaGACAGAAGACCGCTG GGT
	SEQ ID NO: 272	TAGATaGACAGAAGACCGCT GGG
	SEQ ID NO: 273	GTAGATaGACAGAAGACCGC TGG
	SEQ ID NO: 274	TGTAGATaGACAGAAGACCG CTG
	SEQ ID NO: 275	CTGTAGATaGACAGAAGACCG CTGGGT
30	SEQ ID NO: 276	AGTaGAGTGAACCTTCGGAAA AAG
	SEQ ID NO: 277	AAGTaGAGTGAACCTTCGGAA AAA
	SEQ ID NO: 278	CAAGTaGAGTGAACCTTCGGA AAA
	SEQ ID NO: 279	TCAAGTaGAGTGAACCTTCGG AAA
	SEQ ID NO: 280	TCAAGTaGAGTGAACCTTCGGA AAAAGT

	SEQ ID NO: 281	AGTaGGATCACATGTGCCAA CAG
	SEQ ID NO: 282	CTCAGTaGGATCACATGTGC CAA
	SEQ ID NO: 283	TCAGTaGGATCACATGTGCCA ACAGGT
	SEQ ID NO: 284	AGTCTTCaGAGTTTCATGGC AGT
5	SEQ ID NO: 285	CTGCAGTCTTCaGAGTTTCAT GGCAGT
	SEQ ID NO: 286	AGTGaATAAAGGTTAGACAT TAA
	SEQ ID NO: 287	AGTTaTTCTTCCAAAGCAGC AGT
	SEQ ID NO: 288	AAGTTaTTCTTCCAAAGCAG CAG
	SEQ ID NO: 289	TTAAGTTaTTCTTCCAAAGC AGC
10	SEQ ID NO: 290	CTTAAGTTaTTCTTCCAAAGC AGCAGT
	SEQ ID NO: 291	ATCaATCTCCCAATACCTGG AGA
	SEQ ID NO: 292	CATCaATCTCCCAATACCTG GAG
	SEQ ID NO: 293	CCATCaATCTCCCAATACCT GGA
	SEQ ID NO: 294	CCCATCaATCTCCCAATACC TGG
15	SEQ ID NO: 295	GCCCATCaATCTCCCAATAC CTG
	SEQ ID NO: 296	ATCAGTaGGATGAAGTACAA GAA
	SEQ ID NO: 297	AATCAGTaGGATGAAGTACA AGA
	SEQ ID NO: 298	ATCAGTGaGATGAAGTACAA GAA
	SEQ ID NO: 299	ATCaTTCAATTTGAGGTTGA AGA
20	SEQ ID NO: 300	AATCaTTCAATTTGAGGTTG AAG
	SEQ ID NO: 301	TAATCaTTCAATTTGAGGTT GAA
	SEQ ID NO: 302	TTAATCaTTCAATTTGAGGT TGA
	SEQ ID NO: 303	TTTAATCaTTCAATTTGAGG TTG
	SEQ ID NO: 304	ATCCTaTAGGTCACCTGAAGA GGT
25	SEQ ID NO: 305	CATCCTaTAGGTCACCTGAAG AGG
	SEQ ID NO: 306	CCATCCTaTAGGTCACCTGAA GAG
	SEQ ID NO: 307	TCCCATCCTaTAGGTCACCTGA AGAGGT
	SEQ ID NO: 308	ATCTaTTCCACCAGGGCCTG AGC
	SEQ ID NO: 309	ACCATCTaTTCCACCAGGGC CTG
30	SEQ ID NO: 310	ATCTaTTCCACCAGGGCCTGA GCTGAT
	SEQ ID NO: 311	ATCTaTTCCCTTGTGGTCAC CGT
	SEQ ID NO: 312	GATCTaTTCCCTTGTGGTCA CCG
	SEQ ID NO: 313	GATCTaTTCCCTTGTGGTCAC CGTAGT
	SEQ ID NO: 314	ATGaAAGGAGGGTCCCTATA CAG

	SEQ ID NO: 315	GTCATGaAAGGAGGGTCCCTA TACAGT
	SEQ ID NO: 316	ATGaCTGGAAGCTAAGGAAG AAG
	SEQ ID NO: 317	AATGaCTGGAAGCTAAGGAA GAA
	SEQ ID NO: 318	CAATGaCTGGAAGCTAAGGA AGA
5	SEQ ID NO: 319	ACAATGaCTGGAAGCTAAGG AAG
	SEQ ID NO: 320	CACAATGaCTGGAAGCTAAG GAA
	SEQ ID NO: 321	ATTTaAGCTTCAATTTCTCC TTG
	SEQ ID NO: 322	ATTTaTGCAAAGTTGAGTCT TCG
	SEQ ID NO: 323	attttaaCTCCTACTCAGAC TGT
10	SEQ ID NO: 324	tatttttaaCTCCTACTCAGA CTG
	SEQ ID NO: 325	CACaCAGAATCTACTGGCCA GAA
	SEQ ID NO: 326	CTCACaCAGAATCTACTGGC CAG
	SEQ ID NO: 327	CACTaGCGGAGGTCTTTGGC CAA
	SEQ ID NO: 328	CACTTTaTTGCTCTTGCAGA GAA
15	SEQ ID NO: 329	CCACTTTaTTGCTCTTGCAG AGA
	SEQ ID NO: 330	CAGaTAAAGTGTGTAAAGGA CAG
	SEQ ID NO: 331	CAACAGaTAAAGTGTGTAAA GGA
	SEQ ID NO: 332	TCAACAGaTAAAGTGTGTAA AGG
	SEQ ID NO: 333	CATTTaAGCTGCGTCCACCT TGT
20	SEQ ID NO: 334	GCATTTaAGCTGCGTCCACC TTG
	SEQ ID NO: 335	CCATTCaGGGATGCTTCGCA AAA
	SEQ ID NO: 336	CCCATTCaGGGATGCTTCGC AAA
	SEQ ID NO: 337	CCCAGaAaGTAATTGAATGTG GAA
	SEQ ID NO: 338	GCCCAGaAaGTAATTGAATGT GGA
25	SEQ ID NO: 339	CCGaTAAGTTTGACGCCAGC CTG
	SEQ ID NO: 340	ACTCCGaTAAGTTTGACGCC AGC
	SEQ ID NO: 341	CACTCCGaTAAGTTTGACGC CAG
	SEQ ID NO: 342	CCTaAATGATCCACTTTGTG ATG
	SEQ ID NO: 343	CCTaCAGAGAAAGGTGCAGA CGC
30	SEQ ID NO: 344	TCCTaCAGAGAAAGGTGCAG ACG
	SEQ ID NO: 345	GTTTCCTaCAGAGAAAGGTGC AGA
	SEQ ID NO: 346	AGTTCCTaCAGAGAAAGGTG CAG
	SEQ ID NO: 347	CCTaGGCAGACTGGATGCTC TGT
	SEQ ID NO: 348	TCCTaGGCAGACTGGATGCT CTG

	SEQ ID NO: 349	CTATTGGAGCCTaTAAGTATA CTGGAT
	SEQ ID NO: 350	CTCaAAGTGCCTGTGTGCAA TAG
	SEQ ID NO: 351	CTCCTCaAAGTGCCTGTGTG CAA
	SEQ ID NO: 352	CTCCTCaAAGTGCCTGTGTGC AATAGT
5	SEQ ID NO: 353	TTTCTCCTCaAAGTGCCTGTG TGCAAT
	SEQ ID NO: 354	CTCaCTCACTCACCCCTGCAA AGG
	SEQ ID NO: 355	TCTCaCTCACTCACCCCTGCA AAG
	SEQ ID NO: 356	CTCTCaCTCACTCACCCCTGC AAA
	SEQ ID NO: 357	CCTCTCaCTCACTCACCCCTG CAA
10	SEQ ID NO: 358	CTCaCTCTATGGCCTGCAGC ATG
	SEQ ID NO: 359	TTTCTCaCTCTATGGCCTGC AGC
	SEQ ID NO: 360	TTTTCTCaCTCTATGGCCTG CAG
	SEQ ID NO: 361	CTCaGGAGGTGACAGCTATC CAG
	SEQ ID NO: 362	CTGCTCaGGAGGTGACAGCTA TCCAGT
15	SEQ ID NO: 363	ctctCaCTTTCTCTCATCTG TGA
	SEQ ID NO: 364	cctctCaCTTTCTCTCATCT GTG
	SEQ ID NO: 365	tcctctCaCTTTCTCTCATC TGT
	SEQ ID NO: 366	ttcctctCaCTTTCTCTCATC TGTGAT
	SEQ ID NO: 367	CTGaAAGAAGCTCTCCTCCC AGC
20	SEQ ID NO: 368	GCTGaAAGAAGCTCTCCTCC CAG
	SEQ ID NO: 369	CTGaATTACTCGCTCAGAAG CTG
	SEQ ID NO: 370	AGCTGaATTACTCGCTCAGA AGC
	SEQ ID NO: 371	CAGCTGaATTACTCGCTCAG AAG
	SEQ ID NO: 372	ACAGCTGaATTACTCGCTCA GAA
25	SEQ ID NO: 373	CTGAGTaGAAGGCGGTAAAC CGT
	SEQ ID NO: 374	TCTGAGTaGAAGGCGGTAAA CCG
	SEQ ID NO: 375	cttaccaaaggtcacaaaGG TAA
	SEQ ID NO: 376	gacttaccaaaggtcacaaa GGT
	SEQ ID NO: 377	tgacttaccaaaggtcaca aGG
30	SEQ ID NO: 378	atgacttaccaaaggtcaca aaG
	SEQ ID NO: 379	agatgacttaccaaaggtcac aaaGGT
	SEQ ID NO: 380	CTTaGAGAGAGTCAATGAGG AGA
	SEQ ID NO: 381	TCTTaGAGAGAGTCAATGAG GAG
	SEQ ID NO: 382	ATCTTaGAGAGAGTCAATGA GGA

	SEQ ID NO: 383	GATCTTaGAGAGAGTCAATG AGG
	SEQ ID NO: 384	TGATCTTaGAGAGAGTCAAT GAG
	SEQ ID NO: 385	GATCTTaGAGAGAGTCAATGA GGAGAT
	SEQ ID NO: 386	CTTaTAGACGCTGCTCAAAA TTG
5	SEQ ID NO: 387	TTTCTTaTAGACGCTGCTCA AAA
	SEQ ID NO: 388	CTTTCTTaTAGACGCTGCTC AAA
	SEQ ID NO: 389	TACTTTCTTaTAGACGCTGCT CAAAAT
	SEQ ID NO: 390	CTTaTTCTAGATCTTCTTGA AGC
	SEQ ID NO: 391	CTTaTTGAGGCAAAAC TTGG AAG
10	SEQ ID NO: 392	TCACTTaTTGAGGCAAAAC TGG
	SEQ ID NO: 393	CTCACTTaTTGAGGCAAAAC TTG
	SEQ ID NO: 394	ACTTaTTGAGGCAAAAC TTGG AAGAGT
	SEQ ID NO: 395	CTTTGaTGGGAAGAAGTAGA GGA
	SEQ ID NO: 396	GCTTTGaTGGGAAGAAGTAG AGG
15	SEQ ID NO: 397	TGCTTTGaTGGGAAGAAGTA GAG
	SEQ ID NO: 398	gaaaGTCTATGCCAGAAAGG AGG
	SEQ ID NO: 399	ggaaaGTCTATGCCAGAAAG GAG
	SEQ ID NO: 400	tggaaaGTCTATGCCAGAAA GGA
	SEQ ID NO: 401	gtggaaaGTCTATGCCAGAA AGG
20	SEQ ID NO: 402	tgtggaaaGTCTATGCCAGA AAG
	SEQ ID NO: 403	GACATaGGTCCAGCAGTCAG AAA
	SEQ ID NO: 404	GGACATaGGTCCAGCAGTCA GAA
	SEQ ID NO: 405	AGGACATaGGTCCAGCAGTC AGA
	SEQ ID NO: 406	GACGCTaCTCAAAATTGGCT GGT
25	SEQ ID NO: 407	AGACGCTaCTCAAAATTGGC TGG
	SEQ ID NO: 408	TGTAGACGCTaCTCAAAATTG GCTGGT
	SEQ ID NO: 409	GAGaTATGAAGATAAGTGAA AAA
	SEQ ID NO: 410	AGAGaTATGAAGATAAGTGA AAA
	SEQ ID NO: 411	AAGAGaTATGAAGATAAGTG AAA
30	SEQ ID NO: 412	GAAGAGaTATGAAGATAAGT GAA
	SEQ ID NO: 413	TGAAGAGaTATGAAGATAAG TGA
	SEQ ID NO: 414	GAAGAGaTATGAAGATAAGTG AAAAAT
	SEQ ID NO: 415	GAGGCTTaTTTGATGCTATC TGC
	SEQ ID NO: 416	GATaTAAGTCGTGTATATTA ATG

	SEQ ID NO: 417	CAGATaTAAGTCGTGTATAT TAA
	SEQ ID NO: 418	GTACAGATaTAAGTCGTGTAT ATTAAT
	SEQ ID NO: 419	GATTaGATTTTCTGTTGGGA GGA
	SEQ ID NO: 420	AGATTaGATTTTCTGTTGGG AGG
5	SEQ ID NO: 421	CAGATTaGATTTTCTGTTGG GAG
	SEQ ID NO: 422	TCAGATTaGATTTTCTGTTG GGA
	SEQ ID NO: 423	TCAGATTaGATTTTCTGTTGG GAGGAT
	SEQ ID NO: 424	GCTaATTATAGAAAGCGATG ATG
	SEQ ID NO: 425	TAGCTaATTATAGAAAGCGA TGA
10	SEQ ID NO: 426	GTAGCTaATTATAGAAAGCG ATG
	SEQ ID NO: 427	TTGTAGCTaATTATAGAAAGC GATGAT
	SEQ ID NO: 428	GCTGaCGGGAAATTGAGAGC AAA
	SEQ ID NO: 429	CGCTGaCGGGAAATTGAGAG CAA
	SEQ ID NO: 430	AGCGCTGaCGGGAAATTGAG AGC
15	SEQ ID NO: 431	AGCGCTGaCGGGAAATTGAGA GCAAAT
	SEQ ID NO: 432	GCTGaTCTGATGGCCTGGCT TTG
	SEQ ID NO: 433	AGCTGaTCTGATGGCCTGGCT TTGAAT
	SEQ ID NO: 434	GCTTTCTaTAATTCATCTGG AGT
	SEQ ID NO: 435	AACTGCTTTCTaTAATTCATC TGGAGT
20	SEQ ID NO: 436	GGCCaTATGTACTTTCTAGCT TTCAAT
	SEQ ID NO: 437	GGCCCTTTGCTaTAAGTATTG GCCAGT
	SEQ ID NO: 438	GGCTaAGTGGTGGTGACAGC CTG
	SEQ ID NO: 439	GATGGCTaAGTGGTGGTGAC AGC
	SEQ ID NO: 440	GTACTaATACCACTGATGAG AAA
25	SEQ ID NO: 441	TGTACTaATACCACTGATGA GAA
	SEQ ID NO: 442	TTGTACTaATACCACTGATG AGA
	SEQ ID NO: 443	CTTGTACTaATACCACTGATG AGAAAT
	SEQ ID NO: 444	GTCATaGAAGGAGGGTCCCTA TACAGT
	SEQ ID NO: 445	GTCaTTGTGTGGCTGACTGC TGG
30	SEQ ID NO: 446	AGTCaTTGTGTGGCTGACTG CTG
	SEQ ID NO: 447	CCAGTCaTTGTGTGGCTGAC TGC
	SEQ ID NO: 448	TCCAGTCaTTGTGTGGCTGA CTG
	SEQ ID NO: 449	GTCTTaTAAAAGAACCCAGC GGT
	SEQ ID NO: 450	TGTCTTaTAAAAGAACCCAG CGG

	SEQ ID NO: 451	ATGTCTTaTAAAAGAACCCA GCG
	SEQ ID NO: 452	GGATGTCTTaTAAAAGAACCC AGCGGT
	SEQ ID NO: 453	GTGaAGTGAACCTTCGGAAAA AGT
	SEQ ID NO: 454	AGTGaAGTGAACCTTCGGAAA AAG
5	SEQ ID NO: 455	AAGTGaAGTGAACCTTCGGAA AAA
	SEQ ID NO: 456	CAAGTGaAGTGAACCTTCGGAA AAA
	SEQ ID NO: 457	TCAAGTGaAGTGAACCTTCGG AAA
	SEQ ID NO: 458	TCAAGTGaAGTGAACCTTCGG AAAAGT
	SEQ ID NO: 459	GTGaATCGAATTCTGCCAGT TGC
10	SEQ ID NO: 460	GGTGaATCGAATTCTGCCAG TTG
	SEQ ID NO: 461	GCCGGTGaATCGAATTCTGC CAG
	SEQ ID NO: 462	GTGaCAGACAAATGTAGATG TGG
	SEQ ID NO: 463	AGTGaCAGACAAATGTAGAT GTG
	SEQ ID NO: 464	CAGTGaCAGACAAATGTAGA TGT
15	SEQ ID NO: 465	CCAGTGaCAGACAAATGTAG ATG
	SEQ ID NO: 466	CTCCGCCAGTGaCAGACAAAT GTAGAT
	SEQ ID NO: 467	GTGaCTCTGCACTAGGCTGA ATG
	SEQ ID NO: 468	CAGTGaCTCTGCACTAGGCT GAA
	SEQ ID NO: 469	CCAGTGaCTCTGCACTAGGC TGA
20	SEQ ID NO: 470	ACCAGTGaCTCTGCACTAGG CTG
	SEQ ID NO: 471	TACCAGTGaCTCTGCACTAGG CTGAAT
	SEQ ID NO: 472	GTGCTaAGATGCTGGACCAA AGT
	SEQ ID NO: 473	AGTGCTaAGATGCTGGACCA AAG
	SEQ ID NO: 474	AAGTGCTaAGATGCTGGACC AAA
25	SEQ ID NO: 475	GAAAGTGCTaAGATGCTGGAC CAAAGT
	SEQ ID NO: 476	GTTaAAGATCTGATAGCCGG TTG
	SEQ ID NO: 477	GAGGTTaAAGATCTGATAGC CGG
	SEQ ID NO: 478	TGAGGTTaAAGATCTGATAG CCG
	SEQ ID NO: 479	TTTGAGGTTaAAGATCTGATA GCCGGT
30	SEQ ID NO: 480	GTTaGCACATGTGATCCCAC TGA
	SEQ ID NO: 481	TGTTaGCACATGTGATCCCA CTG
	SEQ ID NO: 482	CTGTTaGCACATGTGATCCCA CTGAGT
	SEQ ID NO: 483	GTTaGGATAATTTAGTCCAA AAA
	SEQ ID NO: 484	TGTTaGGATAATTTAGTCCA AAA

	SEQ ID NO: 485	GTGTTaGGATAATTTAGTCC AAA
	SEQ ID NO: 486	GGTGTaGGATAATTTAGTC CAA
	SEQ ID NO: 487	TAAGaTCAGATTATTTTGCT TAG
	SEQ ID NO: 488	TAATAaGGGACGAACAGGGA GGA
5	SEQ ID NO: 489	ATAATAaGGGACGAACAGGG AGG
	SEQ ID NO: 490	CATAATAaGGGACGAACAGG GAG
	SEQ ID NO: 491	TCATAATAaGGGACGAACAGG GAGGAT
	SEQ ID NO: 492	TCAAATCaCCTGCAGGTAAA AGC
	SEQ ID NO: 493	TCAaGGGTACATGATGGATT TGA
10	SEQ ID NO: 494	GTCAaGGGTACATGATGGAT TTG
	SEQ ID NO: 495	CTTCTTTGTCAaGGGTACATG ATGGAT
	SEQ ID NO: 496	tcaatttgggaagcaGCATA TTG
	SEQ ID NO: 497	TCAGCTaTAGCCACACCAGA AGT
	SEQ ID NO: 498	TTCAGCTaTAGCCACACCAG AAG
15	SEQ ID NO: 499	TCTTTCAGCTaTAGCCACACC AGAAGT
	SEQ ID NO: 500	TCCTTTCaCATCTTACGGGA CAA
	SEQ ID NO: 501	ATTTCTTTCaCATCTTACGG GACAAT
	SEQ ID NO: 502	TCGTaGCTTCTTTCTGCTGA GGA
	SEQ ID NO: 503	ATCGTaGCTTCTTTCTGCTG AGG
20	SEQ ID NO: 504	TATCGTaGCTTCTTTCTGCT GAG
	SEQ ID NO: 505	TTATCGTaGCTTCTTTCTGC TGA
	SEQ ID NO: 506	TCTaATACTCCAGCCAGTTA AGT
	SEQ ID NO: 507	TTCTaATACTCCAGCCAGTT AAG
	SEQ ID NO: 508	GTTCTaATACTCCAGCCAGT TAA
25	SEQ ID NO: 509	TTGTCTaATACTCCAGCCAG TTAAGT
	SEQ ID NO: 510	TGATGTTGTTCTaATACTCCA GCCAGT
	SEQ ID NO: 511	TCTaGGCAGGACTACGAGGC TGG
	SEQ ID NO: 512	ATCTaGGCAGGACTACGAGG CTG
	SEQ ID NO: 513	AGATCTaGGCAGGACTACGA GGC
30	SEQ ID NO: 514	AAGATCTaGGCAGGACTACG AGG
	SEQ ID NO: 515	TCTaGTATTCCTTAATTGTA CAG
	SEQ ID NO: 516	TGTTTCTaGTATTCCTTAAT TGT
	SEQ ID NO: 517	TCTaTGTTAGTGATGGCTGA GTG
	SEQ ID NO: 518	GTCTaTGTTAGTGATGGCTG AGT

	SEQ ID NO: 519	TGTCTaTGTTAGTGATGGCT GAG
	SEQ ID NO: 520	TTGTCTaTGTTAGTGATGGC TGA
	SEQ ID NO: 521	GTTGTCTaTGTTAGTGATGG CTG
	SEQ ID NO: 522	TGTCTaTGTTAGTGATGGCTG AGTGGT
5	SEQ ID NO: 523	AGTTGTCTaTGTTAGTGATGG CTGAGT
	SEQ ID NO: 524	TCTCaAACATCTTCTCCTGA TGT
	SEQ ID NO: 525	GTCTCaAACATCTTCTCCTG ATG
	SEQ ID NO: 526	AAGTCTCaAACATCTTCTCC TGA
	SEQ ID NO: 527	GTCTCaAACATCTTCTCCTGA TGTAGT
10	SEQ ID NO: 528	TCTTCaAAACTGAGCAAATT TGC
	SEQ ID NO: 529	GTCTTCaAAACTGAGCAAAT TTG
	SEQ ID NO: 530	TGCAGaATTTGGAACAGAGG CGT
	SEQ ID NO: 531	ATGCAGaATTTGGAACAGAG GCG
	SEQ ID NO: 532	AATGCAGaATTTGGAACAGA GGC
15	SEQ ID NO: 533	TGCaGCTTCTCGAGTTCATA GGA
	SEQ ID NO: 534	TTGCaGCTTCTCGAGTTCAT AGG
	SEQ ID NO: 535	TGCTaGACCCATGTCCTGAT GGC
	SEQ ID NO: 536	CTGCTaGACCCATGTCCTGA TGG
	SEQ ID NO: 537	ACTGCTaGACCCATGTCCTG ATG
20	SEQ ID NO: 538	TTCTGACTGCTaGACCCATGT CCTGAT
	SEQ ID NO: 539	TGGTATaTATTTCTGGTGGC AAA
	SEQ ID NO: 540	GTGGTATaTATTTCTGGTGG CAA
	SEQ ID NO: 541	GTGTGGTATaTATTTCTGGTG GCAAAT
	SEQ ID NO: 542	TGTCaCATCAAATTTTCAAG TGA
25	SEQ ID NO: 543	ATGTCaCATCAAATTTTCAA GTG
	SEQ ID NO: 544	TGTTCaTCCCCTATTATGAA GAA
	SEQ ID NO: 545	CTGTTCaTCCCCTATTATGA AGA
	SEQ ID NO: 546	CCTGTTCaTCCCCTATTATG AAG
	SEQ ID NO: 547	CCCTGTTCaTCCCCTATTATG AAGAAT
30	SEQ ID NO: 548	TTACTTCaATCCGTAATGAT TGT
	SEQ ID NO: 549	TTAGaTAGGAAAAGATGTGG AGC
	SEQ ID NO: 550	ATTAGaTAGGAAAAGATGTG GAG
	SEQ ID NO: 551	CATTAGaTAGGAAAAGATGT GGA
	SEQ ID NO: 552	ACATTAGaTAGGAAAAGATG TGG

	SEQ ID NO: 553	TTGaAATATAATCCTCCACT GGC
	SEQ ID NO: 554	TTTGaAATATAATCCTCCAC TGG
	SEQ ID NO: 555	ATTTGaAATATAATCCTCCA CTG
	SEQ ID NO: 556	TTGTCaGACCCAGCTCAGGA GAA
5	SEQ ID NO: 557	ATTGTCaGACCCAGCTCAGG AGA
	SEQ ID NO: 558	GATTGTCaGACCCAGCTCAG GAG
	SEQ ID NO: 559	TGATTGTCaGACCCAGCTCAG GAGAAT
	SEQ ID NO: 560	TTGTCTaAATATCACTGACT AAA
	SEQ ID NO: 561	ATTGTCTaAATATCACTGAC TAA
10	SEQ ID NO: 562	TTGTTCTaCAAAACCCGCAG TGC
	SEQ ID NO: 563	TTTaACAATGCTCAACCAGC TGG
	SEQ ID NO: 564	TTTTaACAATGCTCAACCAG CTG
	SEQ ID NO: 565	GCTTTTaACAATGCTCAACC AGC
	SEQ ID NO: 566	AGCTTTTaACAATGCTCAAC CAG
15	SEQ ID NO: 567	TTTaGGCAGCCTCCTTCCCC TGA
	SEQ ID NO: 568	CTTTaGGCAGCCTCCTTCCC CTG
	SEQ ID NO: 569	tttcttctctacaaTTCCCG ATCAAT
	SEQ ID NO: 570	GTTATCTCCTGTTCTGCAGC
	SEQ ID NO: 571	GTTTATGTCACCAGAGTAAC
20	SEQ ID NO: 572	GAGGTAATAGAGCCAAGCCCT
	SEQ ID NO: 573	GCAAGAATTCCACTTTTCACTTCCT
	SEQ ID NO: 574	CTGTCATCTCCAACTAGAAATGC
	SEQ ID NO: 575	GCAGCCTCTTGCTCACTTACTC
	SEQ ID NO: 576	GATGACAGGCAGGGGCACCG
25	SEQ ID NO: 577	TTCCAGTGGTTCAATGGTCA
	SEQ ID NO: 578	CTTTCAACCCGAACGGAGAC
	SEQ ID NO: 579	GAGCGAGCAGCGTCTTCGAG
	SEQ ID NO: 580	GCAGACGGCAGTCACTAGGG
	SEQ ID NO: 581	GGGAAGCTGGGTGAATGGAG
30	SEQ ID NO: 582	AGCTGTTTGGGAGGTCAGAA
	SEQ ID NO: 583	AGGGAGCAGGAAAGTGAGGT
	SEQ ID NO: 584	GTCGCAGGACAGCTTTTCCT
	SEQ ID NO: 585	TGTAGCTACGCCTGTGATGG
	SEQ ID NO: 586	TGCCCTGAGATCTTTTCCTC

SEQ ID NO: 587 GATCCAGGTGCTGCAGAAGG
SEQ ID NO: 588 CTCTTGCCTCCACTGGTTGT
SEQ ID NO: 589 TCGGTAGGATGCCCTACATC
SEQ ID NO: 590 ATCCTACAGCATGGTGGCTG
5 SEQ ID NO: 591 AGTGGTCTCCGAAACCTCCGCGCCCCGCAAC
SEQ ID NO: 592 TCCTTGAAGAAGATGGTGCG
SEQ ID NO: 593
ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAACTCATTACTGCTGCCCA
GA
10 SEQ ID NO: 594
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGACCTGTTCCGGCTTCTTCCT
TA
SEQ ID NO: 595
ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAATTTCCACTGTCTTCTCTT
15 GAGT
SEQ ID NO: 596
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGCTTGCTTCTGACCTGTCCT
AT
SEQ ID NO: 597
20 ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGACTAGGGGCAAAGCAA
GAT
SEQ ID NO: 598
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCTTCCAAACTTTCTGCCCAT
TC
25 SEQ ID NO: 599
ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACACAGCGTGCTCTTTCCT
TAC
SEQ ID NO: 600
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGTTCAGAAGAACATCCCGT
30 TGAC
SEQ ID NO: 601
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
SEQ ID NO: 602
CAAGCAGAAGACGGCATACGAGATCTTGTAGTGACTGGAGTTCAGACGT

SEQ ID NO: 603

CAAGCAGAAGACGGCATAACGAGATCAGATCGTGACTGGAGTTCAGACGT

SEQ ID NO: 604

CAAGCAGAAGACGGCATAACGAGATCCGTCCGTGACTGGAGTTCAGACGT

5 SEQ ID NO: 605

CAAGCAGAAGACGGCATAACGAGATATGTCAGTGACTGGAGTTCAGACGT

SEQ ID NO: 606

CAAGCAGAAGACGGCATAACGAGAT GTCCGC GTGACTGGAGTTCAGACGT

SEQ ID NO: 607

10 CAAGCAGAAGACGGCATAACGAGAT TTAGGC GTGACTGGAGTTCAGACGT

SEQ ID NO: 608

CAAGCAGAAGACGGCATAACGAGAT CGATGT GTGACTGGAGTTCAGACGT

SEQ ID NO: 609

CAAGCAGAAGACGGCATAACGAGAT TGACCA GTGACTGGAGTTCAGACGT

15 SEQ ID NO: 610

CAAGCAGAAGACGGCATAACGAGAT AGTCAA GTGACTGGAGTTCAGACGT

SEQ ID NO: 611 CAAGCAGAAGACGGCATAACGAGAT AGTTCC

GTGACTGGAGTTCAGACGT

SEQ ID NO: 612

20 CAAGCAGAAGACGGCATAACGAGAT GATCAG GTGACTGGAGTTCAGACGT

SEQ ID NO: 613

CAAGCAGAAGACGGCATAACGAGAT ACAGTG GTGACTGGAGTTCAGACGT

SEQ ID NO: 614

CAAGCAGAAGACGGCATAACGAGAT TATACT GTGACTGGAGTTCAGACGT

25 SEQ ID NO: 615

CAAGCAGAAGACGGCATAACGAGAT CAACAA GTGACTGGAGTTCAGACGT

SEQ ID NO: 616

CAAGCAGAAGACGGCATAACGAGAT GTTGTT GTGACTGGAGTTCAGACGT

SEQ ID NO: 617

30 CAAGCAGAAGACGGCATAACGAGAT TCGGTT GTGACTGGAGTTCAGACGT

SEQ ID NO: 618

CAAGCAGAAGACGGCATAACGAGAT AGTATT GTGACTGGAGTTCAGACGT

SEQ ID NO: 619

CAAGCAGAAGACGGCATAACGAGAT TCTTGT GTGACTGGAGTTCAGACGT

SEQ ID NO: 620 GAACAGCTGCAGAACAGGAGATAACAG
 SEQ ID NO: 621 GTTaTCTCCTGTTCTGCAGCTGT
 SEQ ID NO: 622 ATGACAGGCAGGGGCACCGCGG
 SEQ ID NO: 623 GAGCGAGCAGCGTCTTCGAGAGT
 5 SEQ ID NO: 624 GCAGACGGCAGTCACTAGGGGGC
 SEQ ID NO: 625 GTCGCAGGACAGCTTTTCCTAGA
 SEQ ID NO: 626 GGGAAGCTGGGTGAATGGAGCGA
 SEQ ID NO: 627 GATCCAGGTGCTGCAGAAGGGAT
 SEQ ID NO: 628 GTTATCTCCTGCTCTGCAGCAGA
 10 SEQ ID NO: 629 GATATCTCCTGTTCTGCAGGAGA
 SEQ ID NO: 630 GGATTTCCTCAAGTCTCCACCC
 SEQ ID NO: 631 TCCCACCGTACACGCCTAC
 SEQ ID NO: 632
 GGTTTCAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAG
 15 GA
 SEQ ID NO: 633
 GGTTTTAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAG
 GA
 SEQ ID NO: 634 GTGCTCACATTCCTTAAATTAAGG
 20 SEQ ID NO: 635 GGCTCACATTCCTTAAATTAAGGA

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of skill in the art to which the disclosed invention belongs. Publications cited herein and the materials for which they are cited are specifically
 25 incorporated by reference.

Those skilled in the art will appreciate that numerous changes and modifications can be made to the preferred embodiments of the invention and that such changes and modifications can be made without departing from the spirit of the invention. It is, therefore, intended that the appended claims cover all such equivalent variations as fall within the true spirit and scope of the
 30 invention.

CLAIMS

What is claimed is:

1. A fusion protein comprising a Cas9 nickase and a nucleotide deaminase, wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.
2. The fusion protein of claim 1, wherein the first amino acid substitution is selected from the group consisting of A262T, R324L, S409I, E480K, E543D, M694I, and E1219V when compared to SEQ ID NO: 11.
3. The fusion protein of claim 1, wherein the second amino acid substitution is selected from the group consisting of L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, R1335Q, R1335E, and T1337R when compared to SEQ ID NO: 11.
4. The fusion protein of any one of claims 1 to 3, wherein the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.
5. The fusion protein of any one of claims 1 to 4, wherein the Cas9 nickase when in conjunction with a bound guide RNA (gRNA) specifically binds to a target nucleic acid sequence.
6. The fusion protein of any one of claim 1 to 5, wherein the Cas9 nickase recognizes a NG protospacer adjacent motif (PAM) sequence.
7. The fusion protein of any one of claims 1 to 6, wherein the nucleotide deaminase is a cytidine deaminase or an adenine deaminase.

8. The fusion protein of claim 7, wherein the adenine deaminase is a dimeric adenine deaminase or a monomeric adenine deaminase domain.

9. The fusion protein of claim 8, wherein the monomeric adenine deaminase comprises amino acid substitutions A56G and V82G when compared to SEQ ID NO: 47.

10. The fusion protein of any one of claims 7 to 9, wherein the adenine deaminase comprises the amino acid sequence set forth in SEQ ID NO: 22 or 23.

11. A system for base editing comprising:
a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N; and
a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;
wherein the first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,
wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and
wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.

12. The system of claim 11, wherein the first amino acid substitution is selected from the group consisting of A262T, R324L, S409I, E480K, E543D, M694I, and E1219V when compared to SEQ ID NO: 11.

13. The system of claim 11, wherein the second amino acid substitution is selected from the group consisting of L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, R1335Q, R1335E, and T1337R when compared to SEQ ID NO: 11.

14. The system of any one of claims 11 to 13, wherein the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

15. The system of any one of claims 11 to 14, wherein the Cas9 nickase recognizes a NG protospacer adjacent motif (PAM) sequence.

16. The system of any one of claims 11 to 15, wherein the N-terminal portion of the Cas9 nickase comprises the amino acid positions 2-573 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20 and the C-terminal portion of the Cas9 nickase comprises the amino acid positions 574-1368 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

17. The system of any one of claims 11 to 16, wherein the nucleotide deaminase is a cytidine deaminase or an adenine deaminase.

18. The system of claim 17, wherein the adenine deaminase is a dimeric adenine deaminase or a monomeric adenine deaminase domain.

19. The system of claim 18, wherein the monomeric adenine deaminase comprises amino acid substitutions A56G and V82G when compared to SEQ ID NO: 47.

20. The system of any one of claims 17 to 19, wherein the adenine deaminase comprises the amino acid sequence as set forth in SEQ ID NO: 22 or 23.

21. The system of any one of claims 11 to 20, wherein the intein-N is a Cfa intein-N or a gp41-1 intein-N.

22. The system of any one of claims 11 to 20, wherein the intein-C is a Cfa intein-C or a gp41-1 intein-C.

23. The system of any one of claims 11 to 22, wherein the first nucleotide sequence or the second nucleotide sequence further comprises a nucleotide encoding a guide RNA (gRNA).

24. A composition comprising:

- a first recombinant adeno-associated virus (AAV) particle comprising a first nucleotide sequence encoding a N-terminal portion of a Cas9 nickase fused at its C-terminus to an intein-N; and
- a second recombinant AAV particle comprising a second nucleotide sequence encoding an intein-C fused to the N-terminus of a C-terminal portion of the Cas9 nickase;

wherein first nucleotide sequence further comprises a nucleotide sequence encoding a nucleotide deaminase fused to the N-terminus of the N-terminal portion of the Cas9 nickase,

wherein the Cas9 nickase comprises a first amino acid substitution at a position selected from the group consisting of 262, 324, 409, 480, 543, 694, and 1219 when compared to SEQ ID NO: 11, and

wherein the Cas9 nickase comprises a second amino acid substitution at a position selected from the group consisting of 1111, 1135, 1218, 1219, 1322, 1335, and 1337 when compared to SEQ ID NO: 11.

25. The composition of claim 24, wherein the first amino acid substitution is selected from the group consisting of A262T, R324L, S409I, E480K, E543D, M694I, and E1219V when compared to SEQ ID NO: 11.

26. The composition of claim 24, wherein the second amino acid substitution is selected from the group consisting of L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, R1335Q, R1335E, and T1337R when compared to SEQ ID NO: 11.

27. The composition of any one of claims 24 to 26, wherein the Cas9 nickase comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

28. The composition of any one of claims 24 to 27, wherein the Cas9 nickase recognizes a NG protospacer adjacent motif (PAM) sequence.

29. The composition of any one of claims 24 to 28, wherein the N-terminal portion of the Cas9 nickase comprises the amino acid positions 2-573 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20 and the C-terminal portion of the

Cas9 nickase comprises the amino acid positions 574-1368 of the amino acid sequence selected from the group consisting of SEQ ID NOs: 12-20.

30. The composition of any one of claims 24 to 29, wherein the nucleotide deaminase is a cytidine deaminase or an adenine deaminase.

31. The composition of claim 30, wherein the adenine deaminase is a dimeric adenine deaminase or a monomeric adenine deaminase domain.

32. The composition of claim 31, wherein the monomeric adenine deaminase comprises amino acid substitutions A56G and V82G when compared to SEQ ID NO: 47.

33. The composition of any one of claims 30 to 32, wherein the adenine deaminase comprises the amino acid sequence as set forth in SEQ ID NO: 22 or 23.

34. The composition of any one of claims 24 to 33, wherein the intein-N is a Cfa intein-N or a gp41-1 intein-N.

35. The composition of any one of claims 24 to 33, wherein the intein-C is a Cfa intein-C or a gp41-1 intein-C.

36. The composition of any one of claims 24 to 35, wherein the first nucleotide sequence or the second nucleotide sequence further comprises a nucleotide encoding a guide RNA (gRNA).

37. A method of treating a genetic disease in a subject, comprising administering a therapeutically effective amount of the composition of any one of claims 24 to 36.

38. The method of claim 37, wherein the genetic disease is a muscular dystrophy or spinal muscular atrophy.

39. The method of claim 38, wherein the muscular dystrophy is Duchenne muscular dystrophy.

- 40. The method of claim 38, wherein the muscular dystrophy is dysferlinopathy.
- 41. The method of claim 37, wherein the genetic disease is hypercholesterolemia.
- 42. The method of claim 37, wherein the genetic disease is due to a point mutation in a causative gene from a G:C pair to an A:T pair.

mdx^{4cv}: GAA **CAG** CAG CTG CTG CAG AAC AGG AGA **TAA** CAG
E Q L Q N R R **X** Q
gRNA: gGTT**a**TCTCCTGTTCTGCAGC **TGT**

FIG. 1A

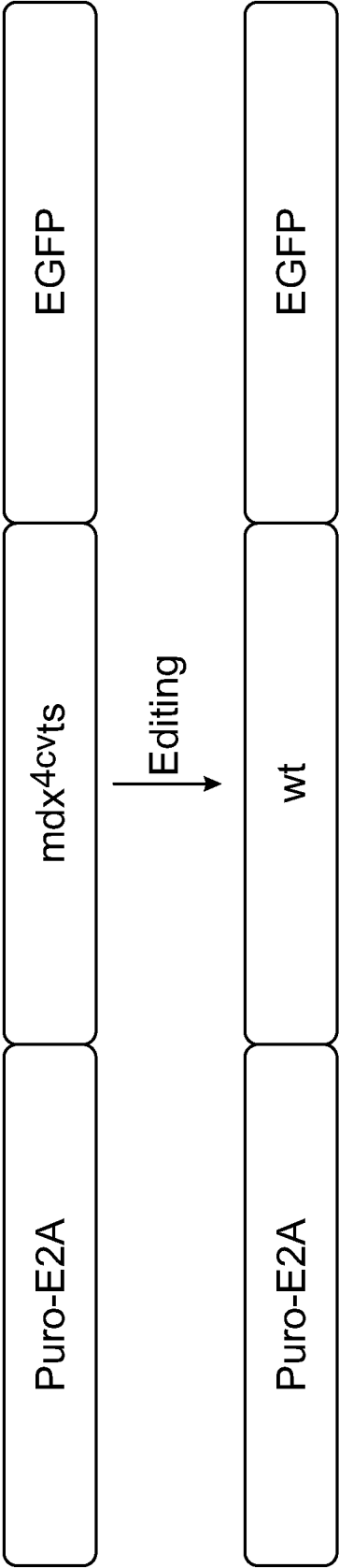


FIG. 1B

2/63

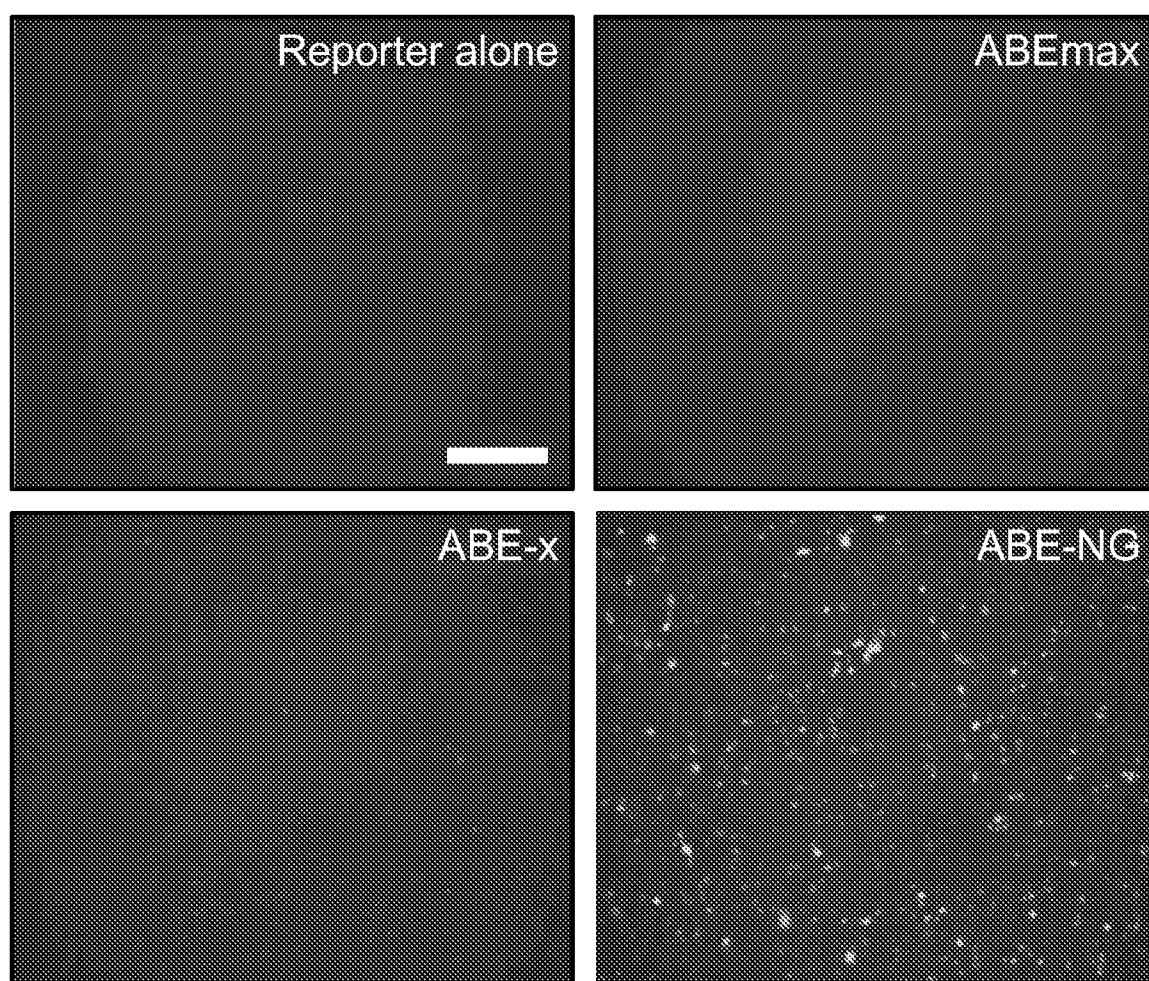


FIG. 1C

3/63

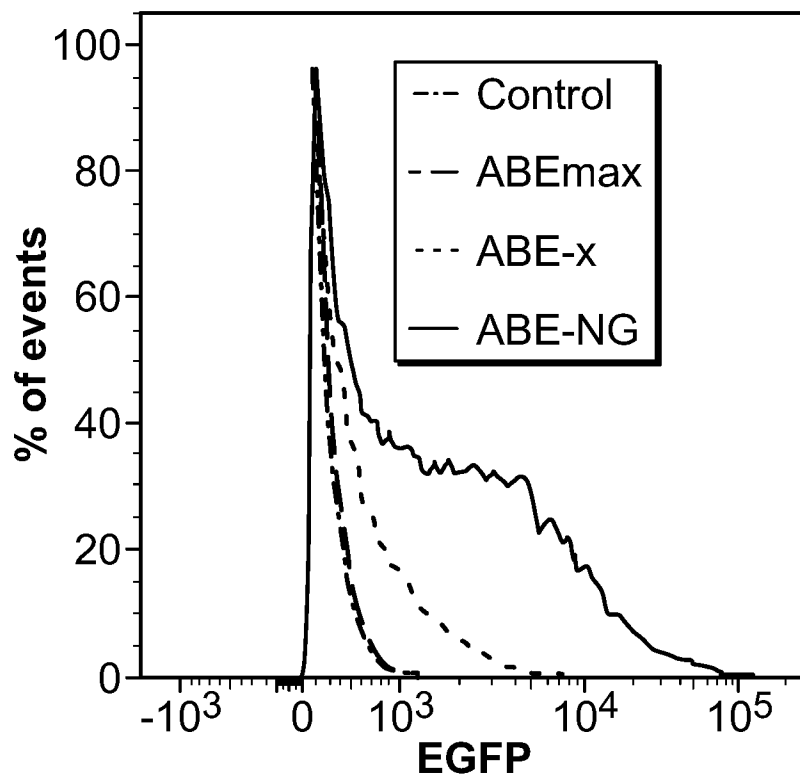


FIG. 1D

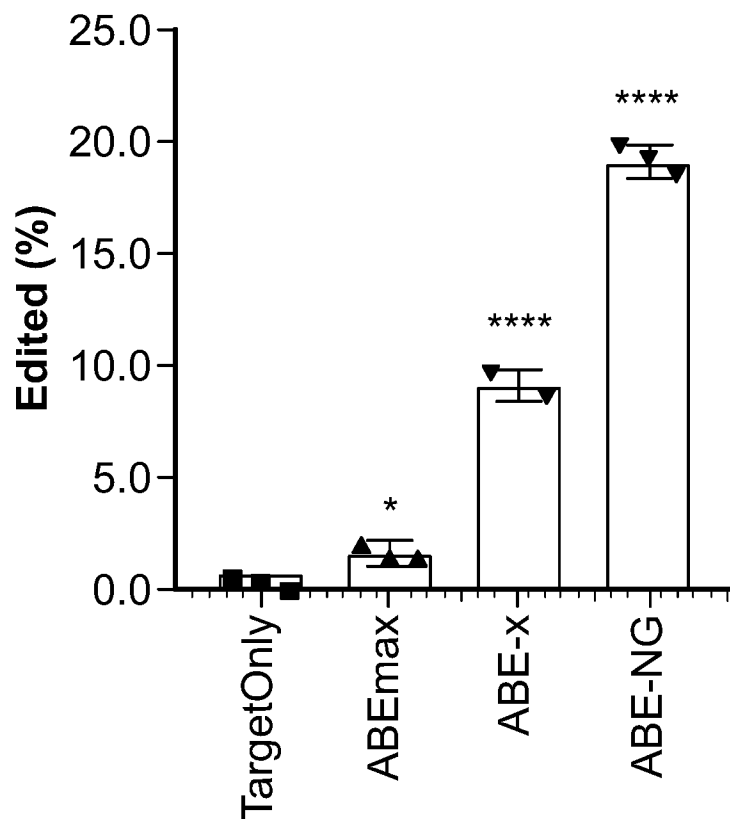
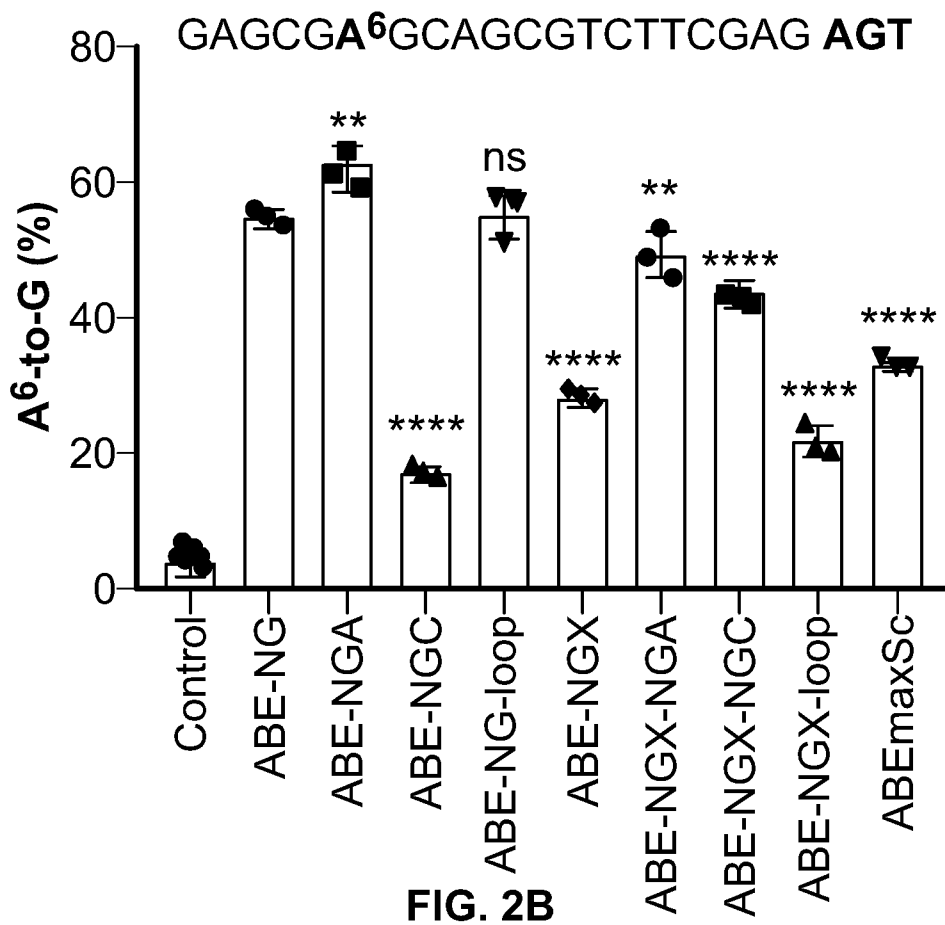
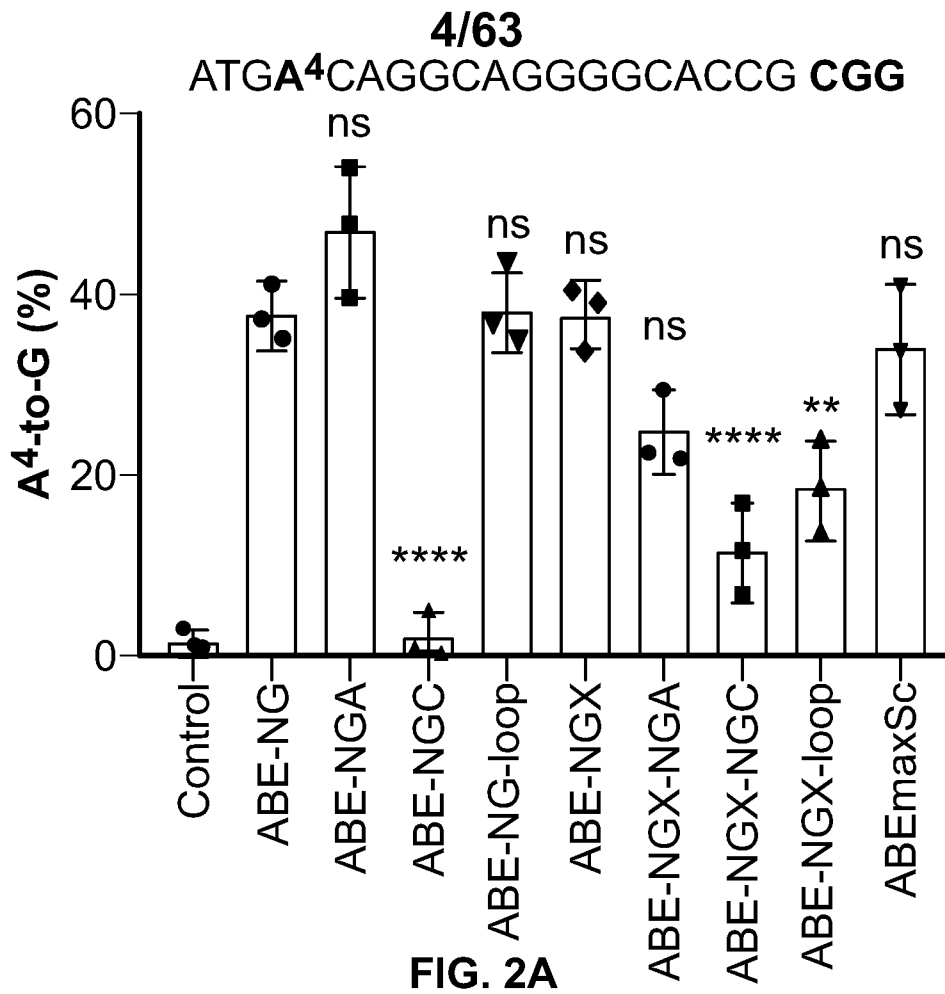
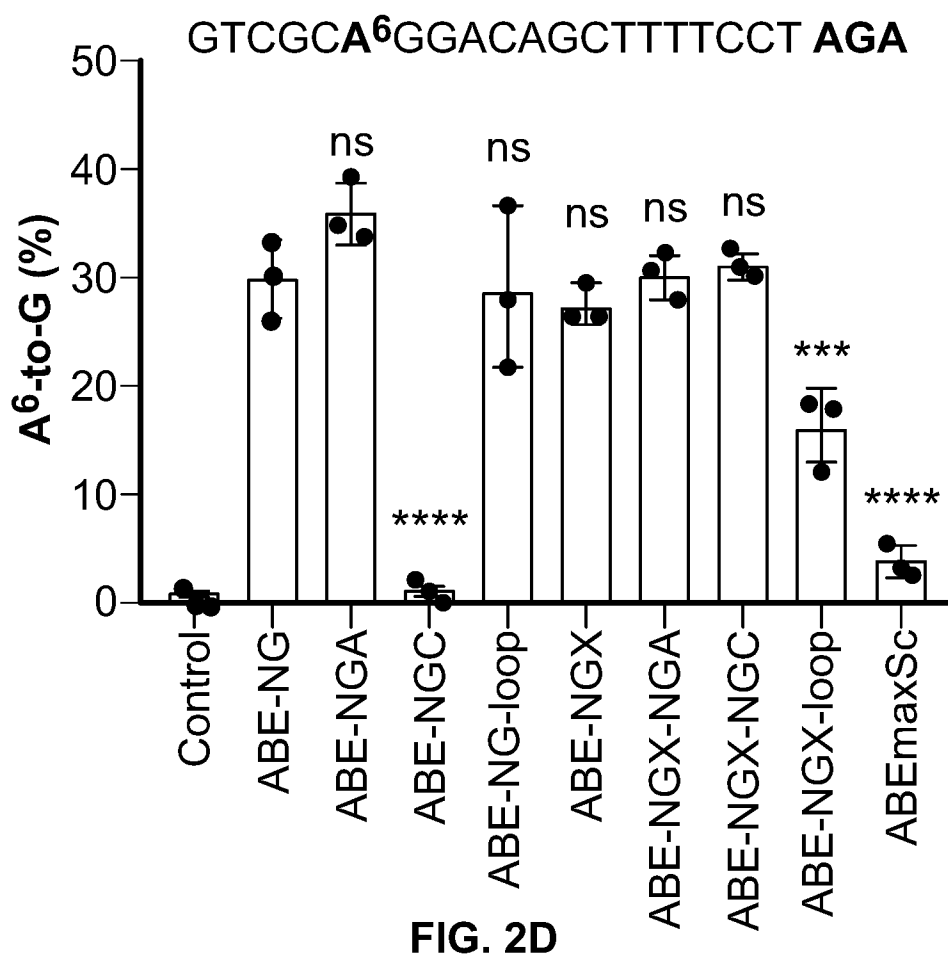
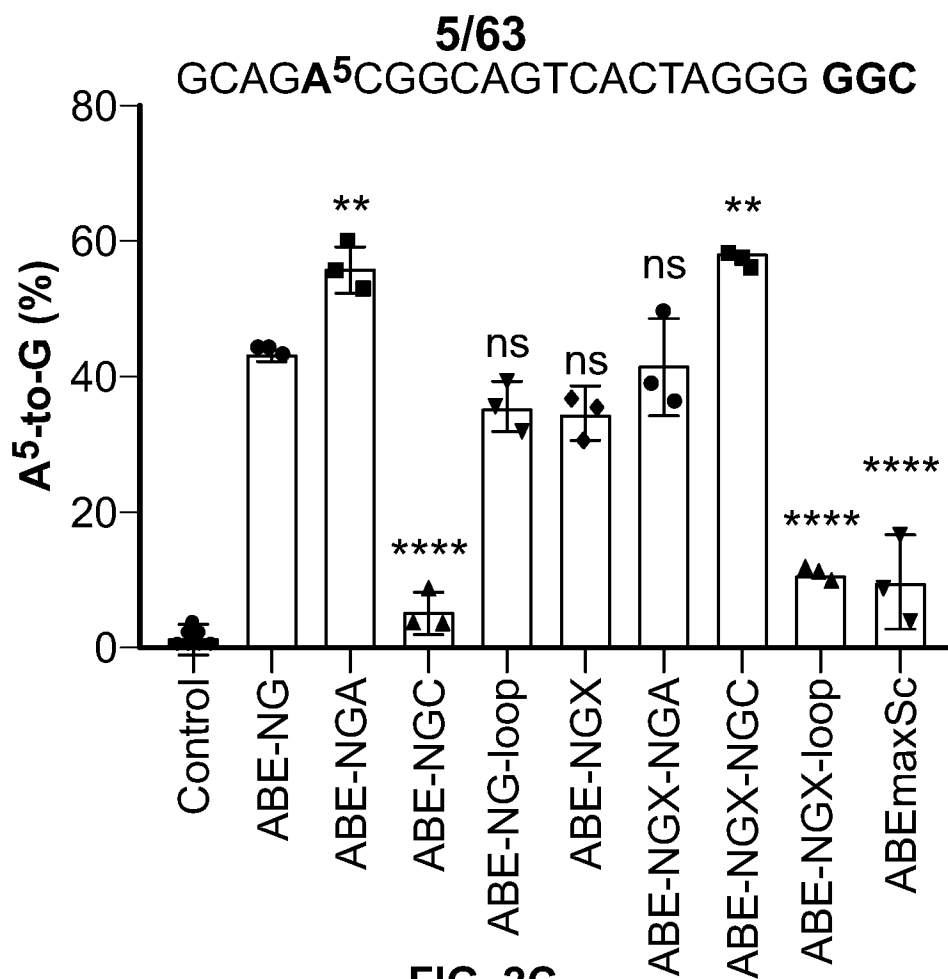
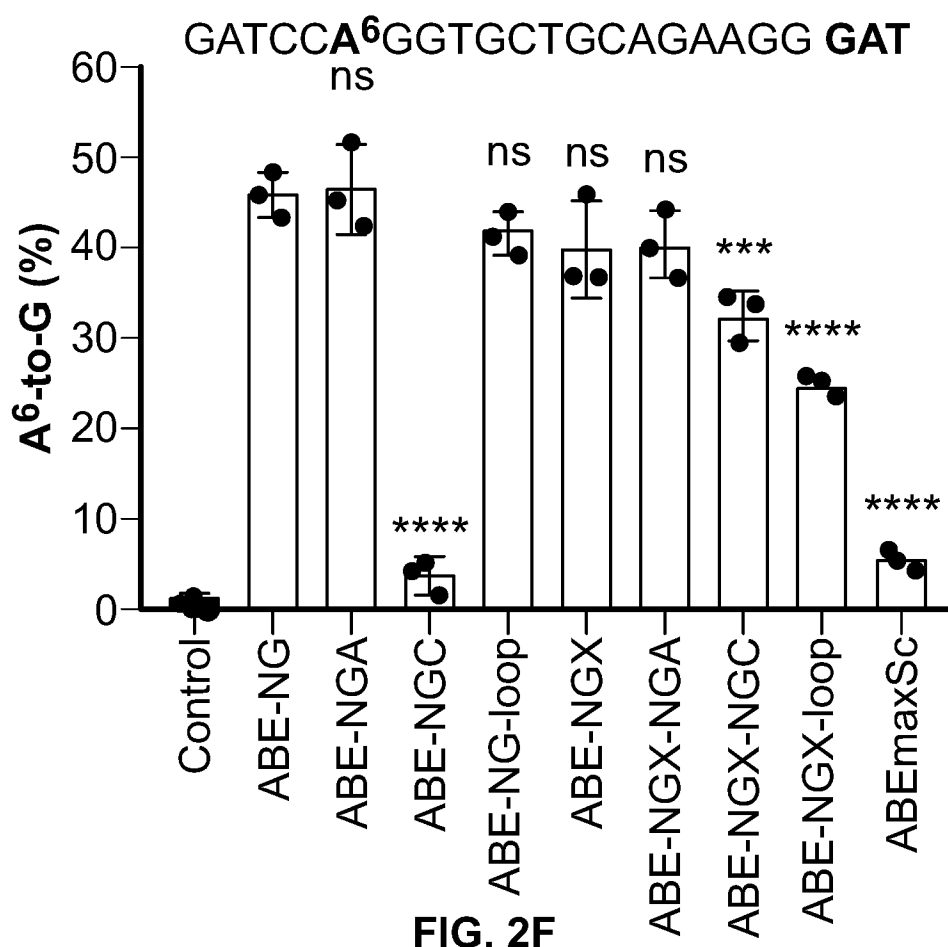
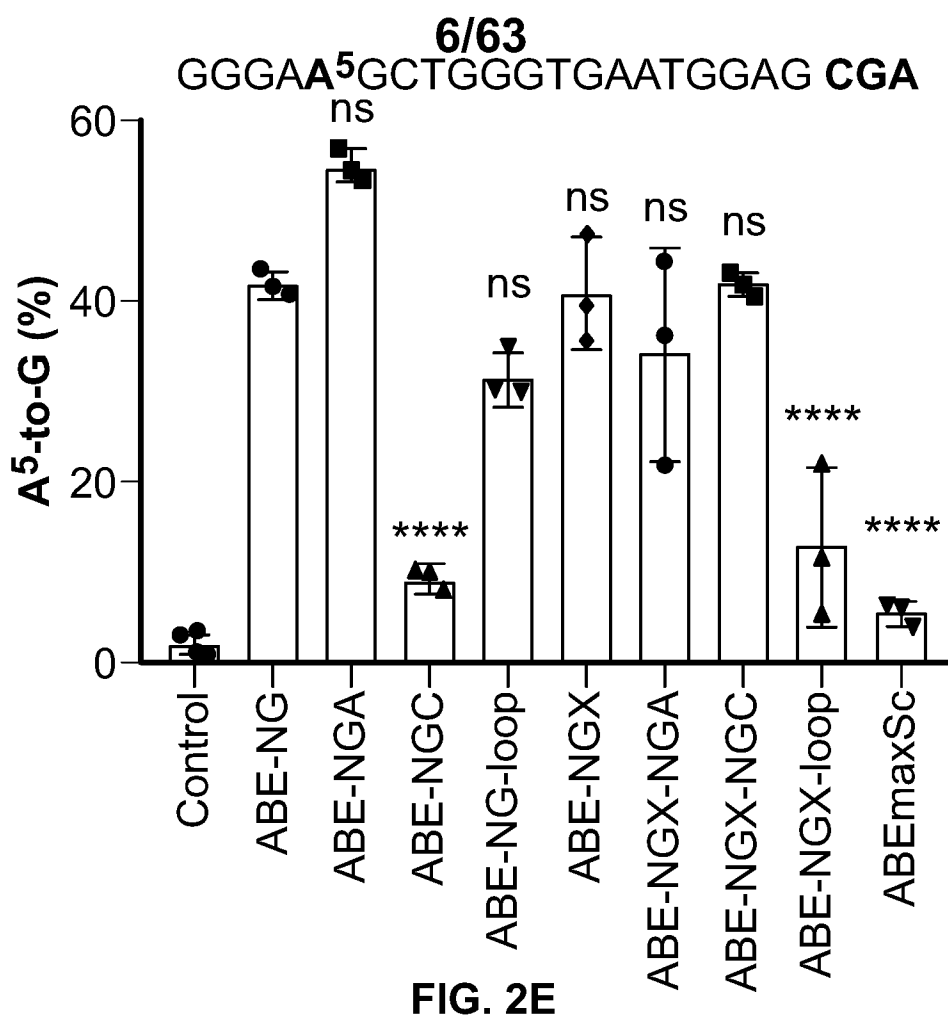


FIG. 1E







7/63

ABE7.10	WT TadA	TadA*
miniABE		TadA*
miniABE(V82G)		TadA*(V82G)
miniABE(A56G)		TadA*(A56G)
miniABE(GG)		TadA*(GG)

FIG. 3A

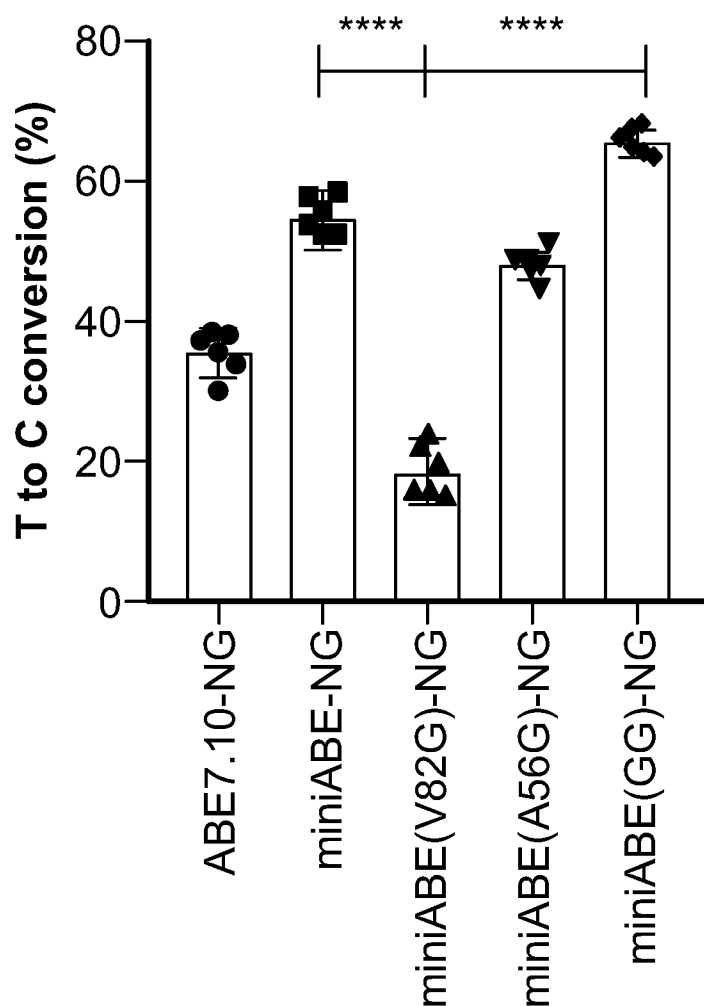
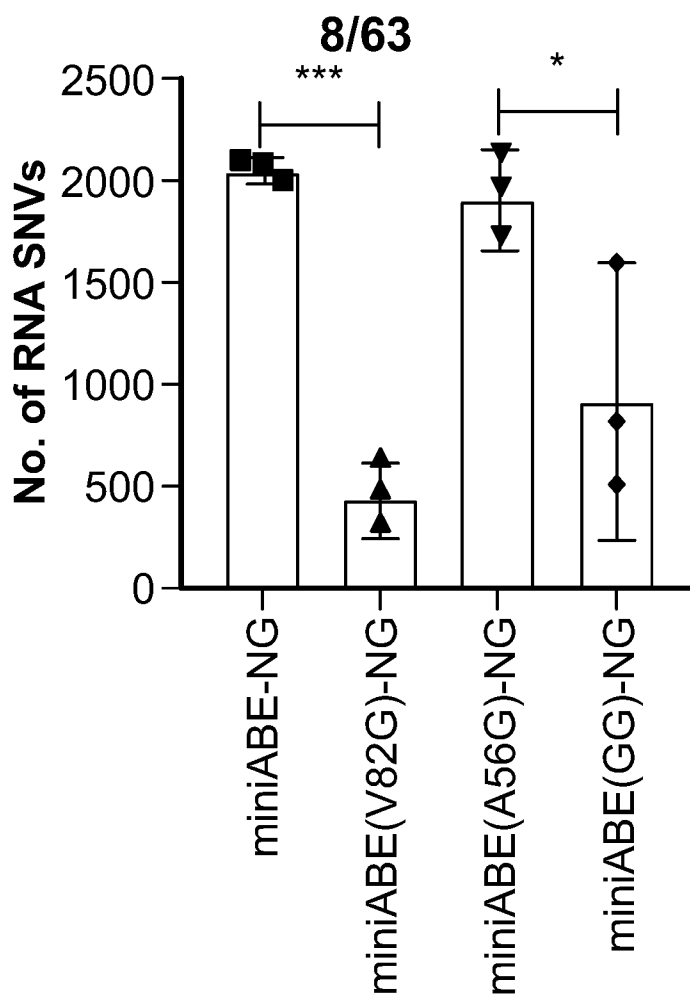
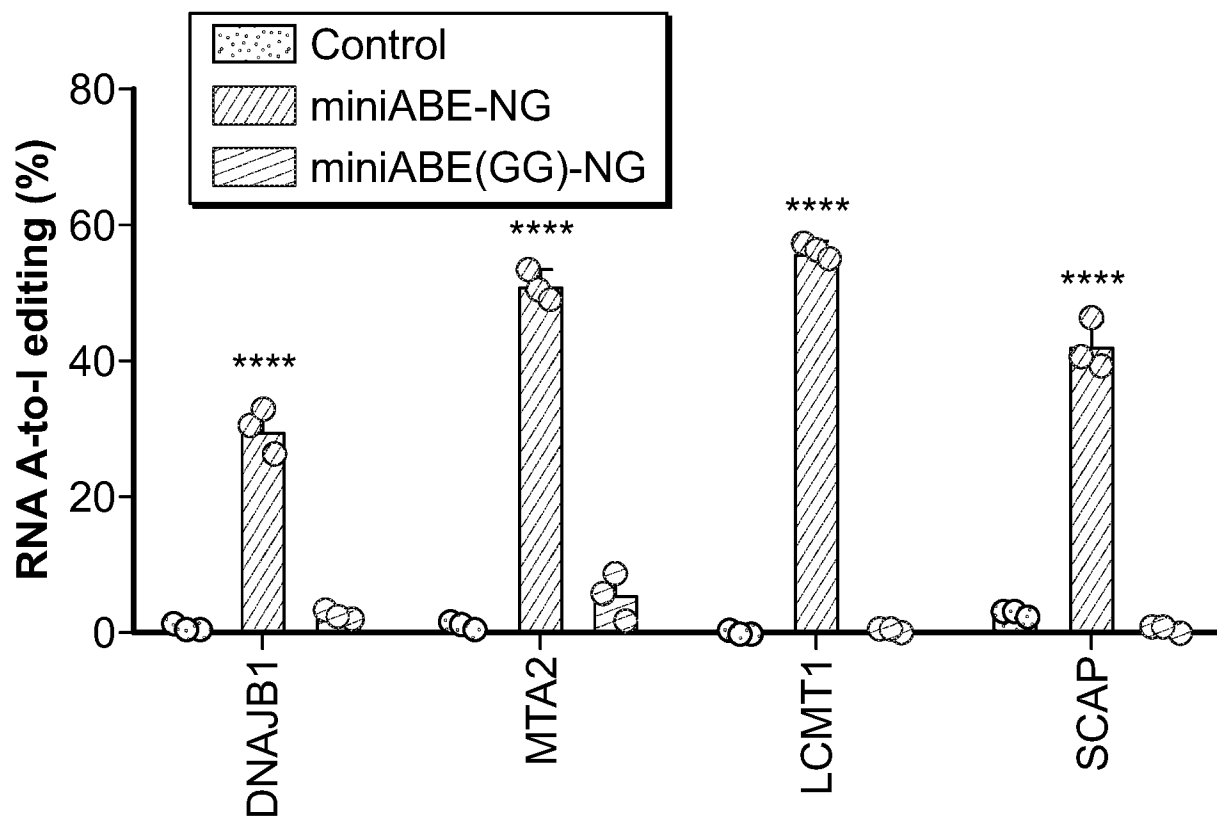


FIG. 3B

**FIG. 3C****FIG. 3D**

9/63

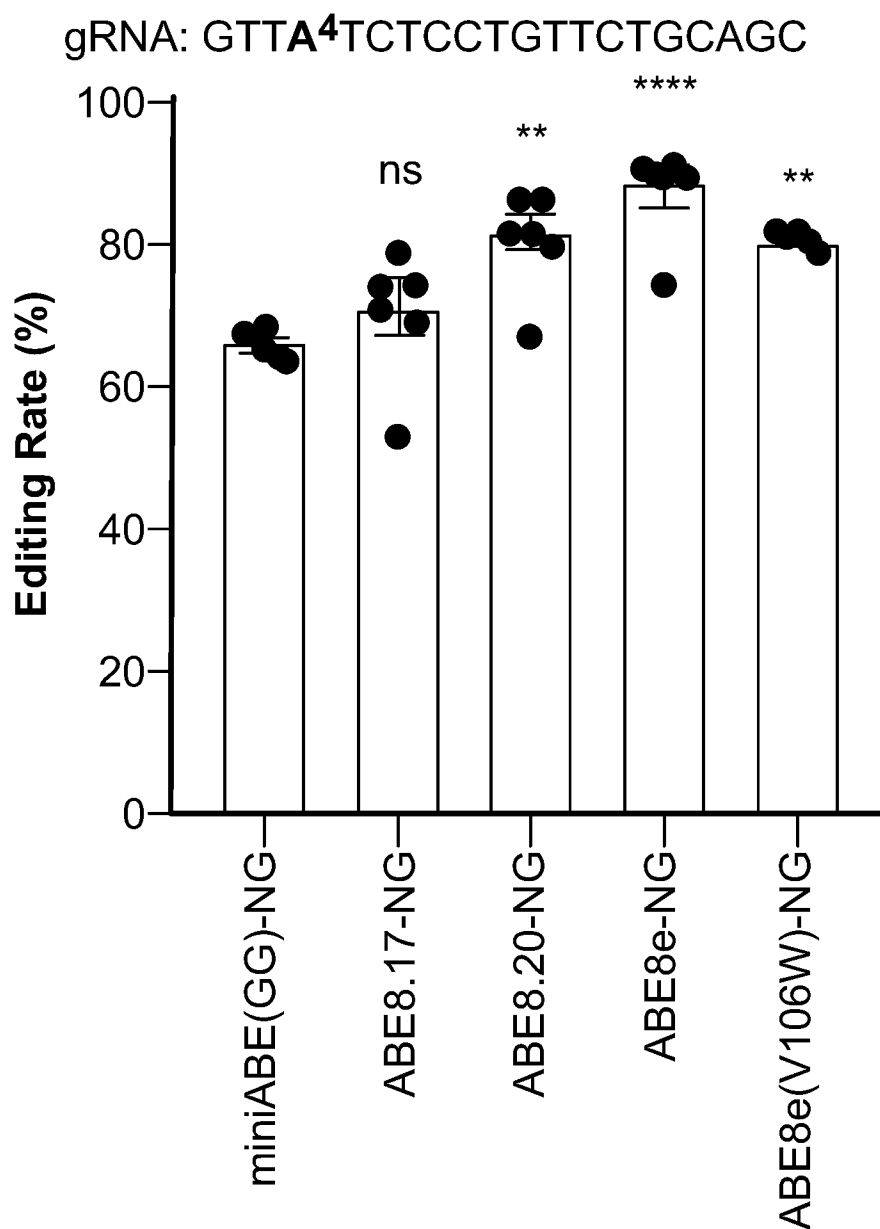
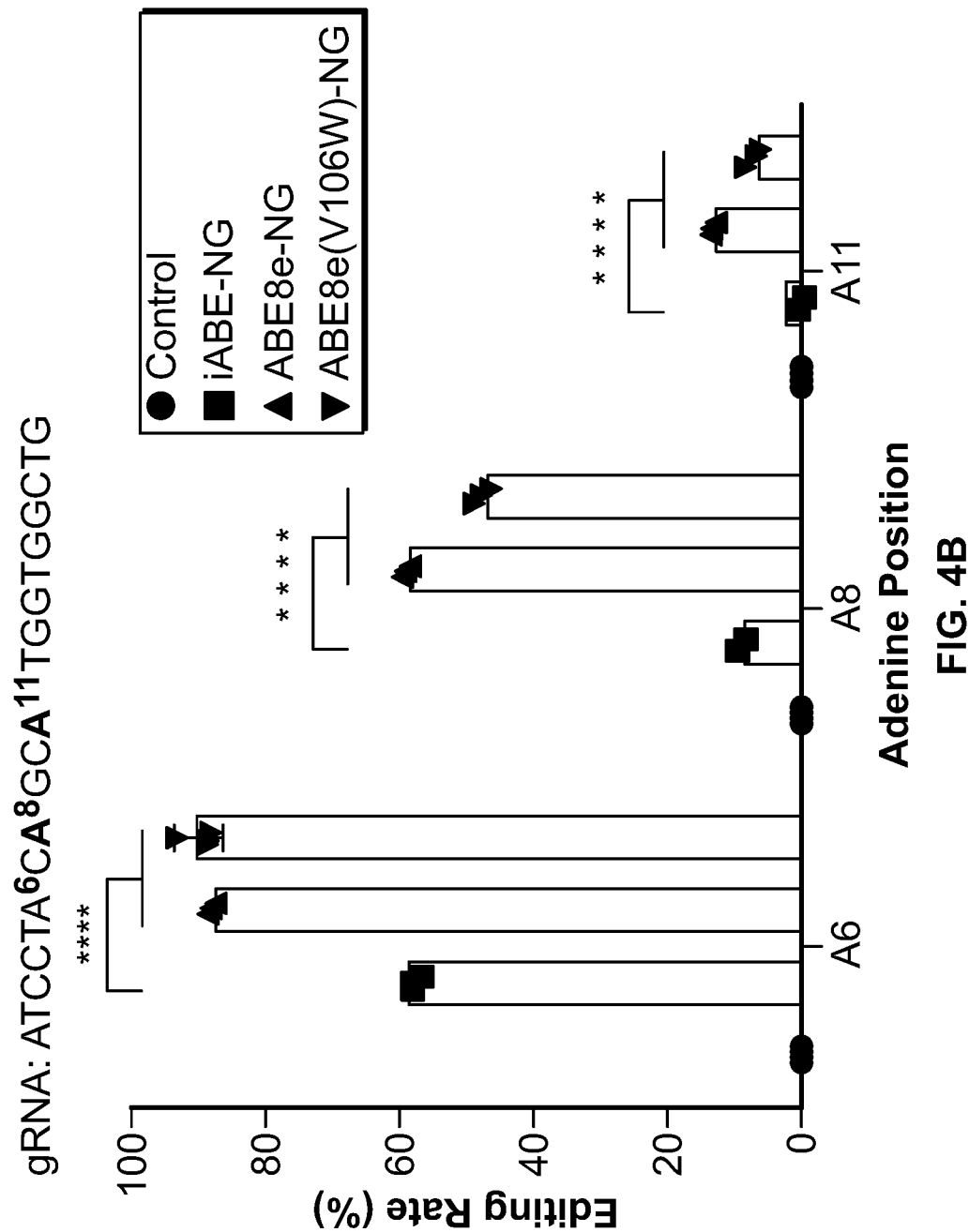


FIG. 4A



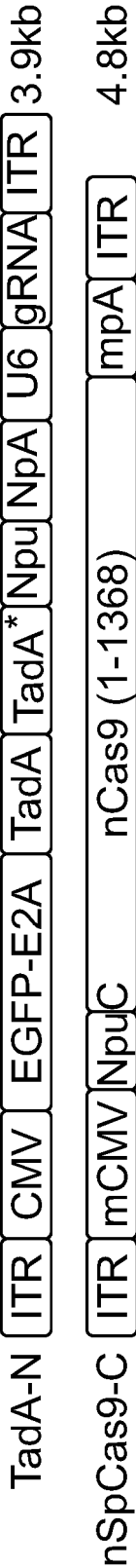


FIG. 5A

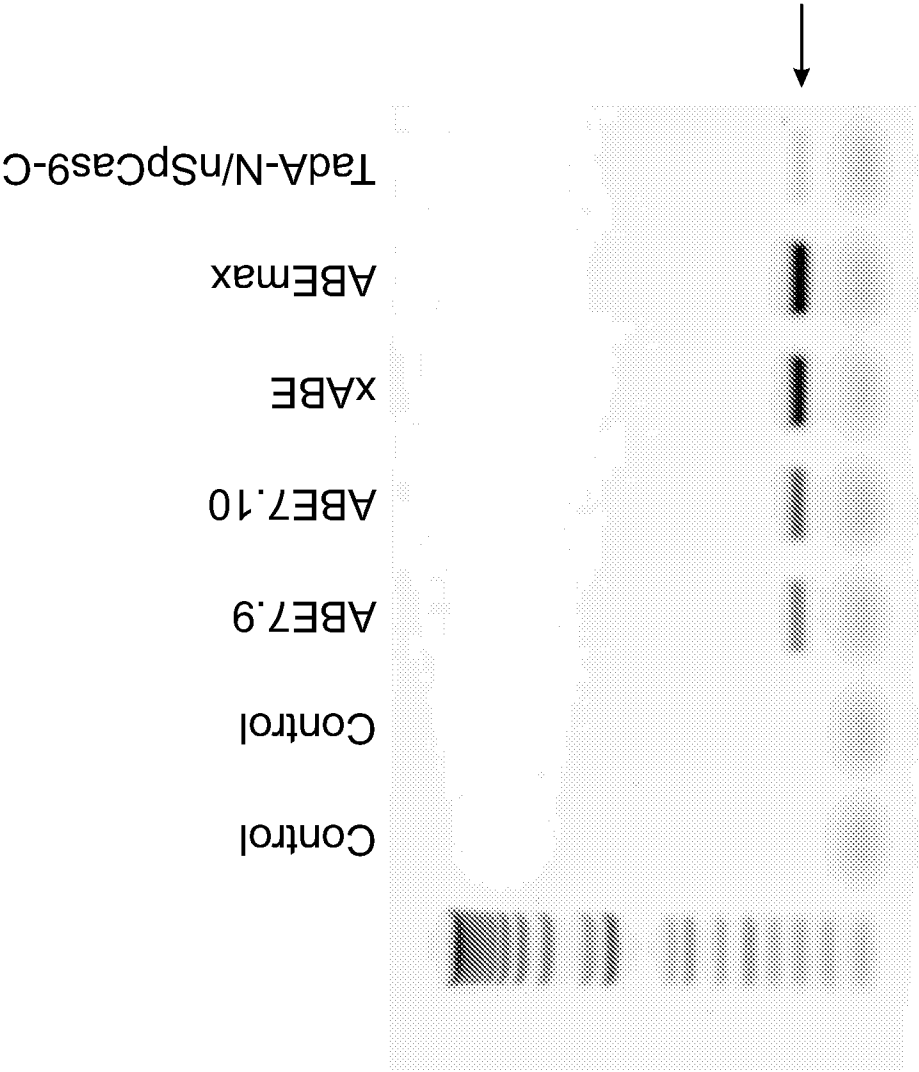


FIG. 5B

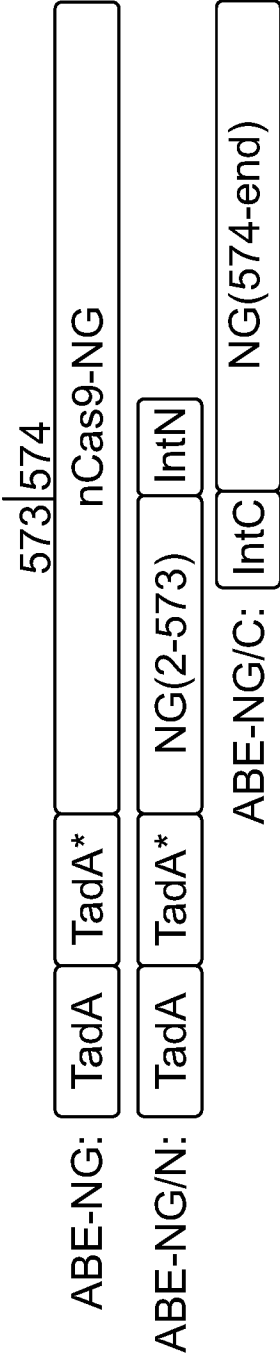


FIG. 6A

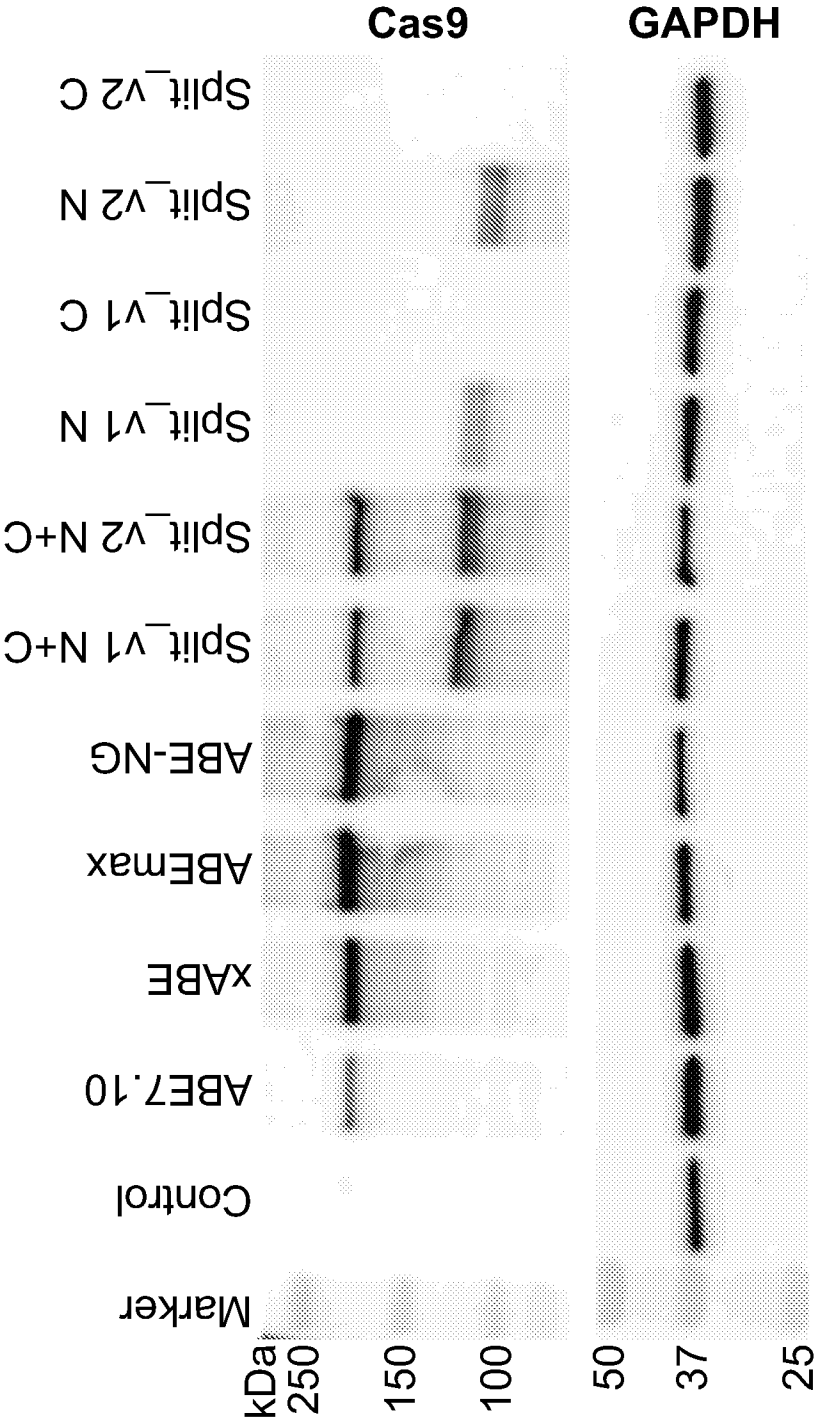


FIG. 6B

13/63

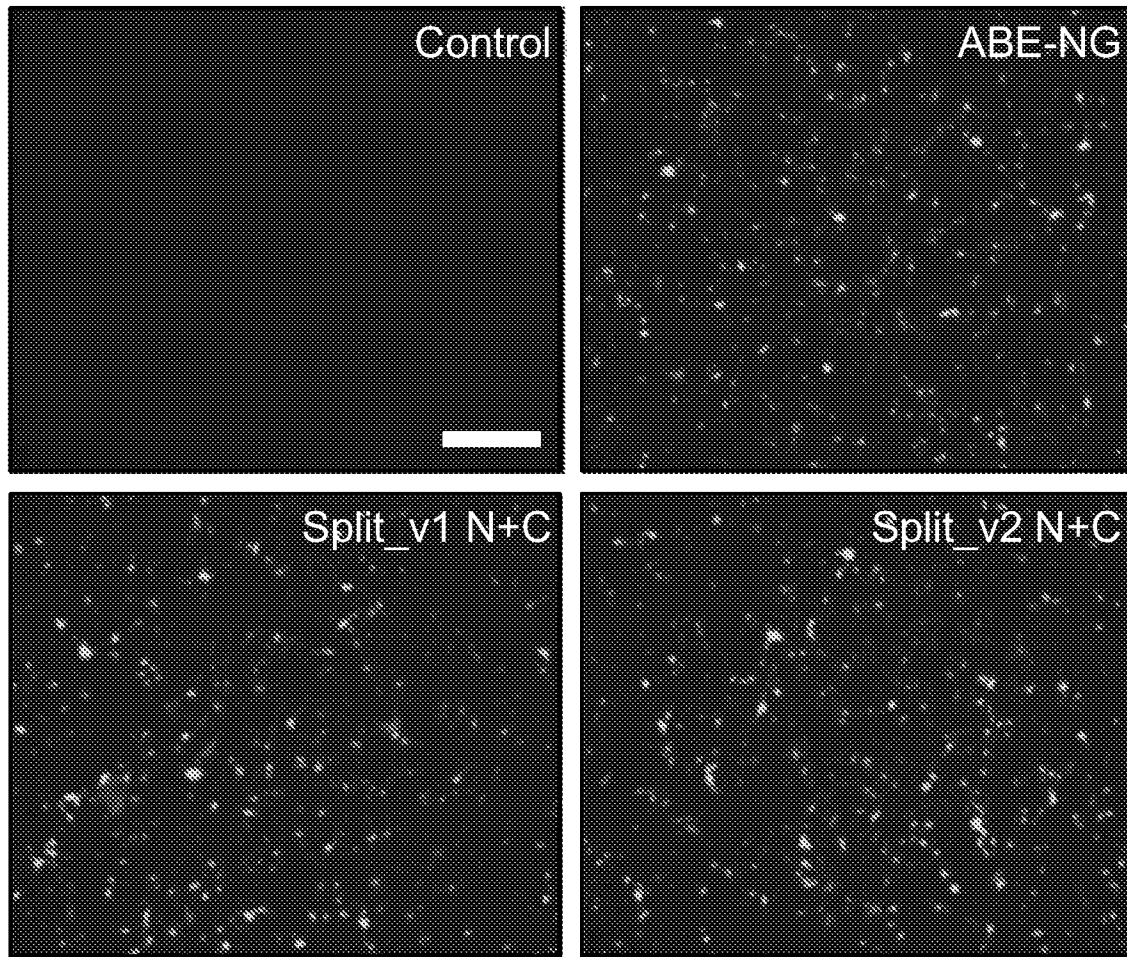


FIG. 6C

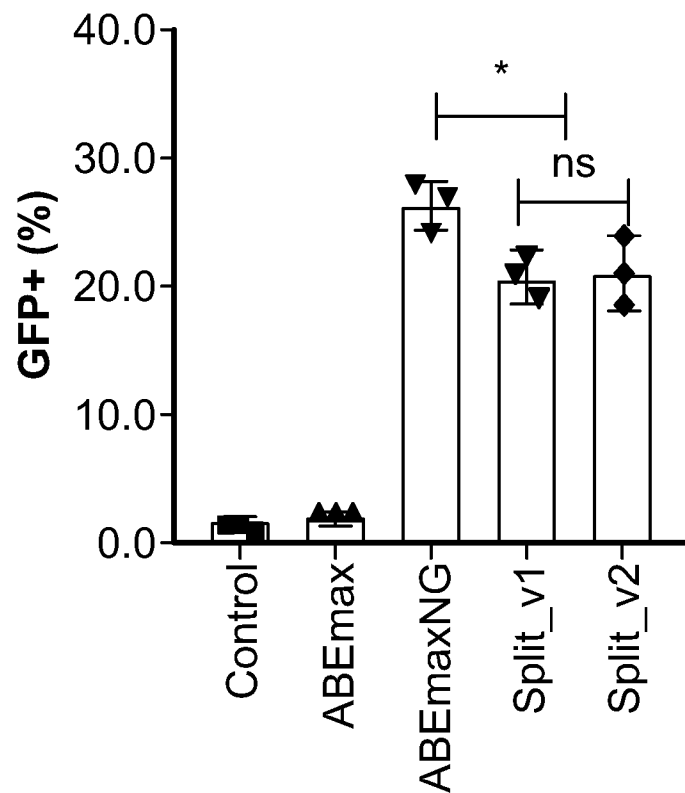


FIG. 6D

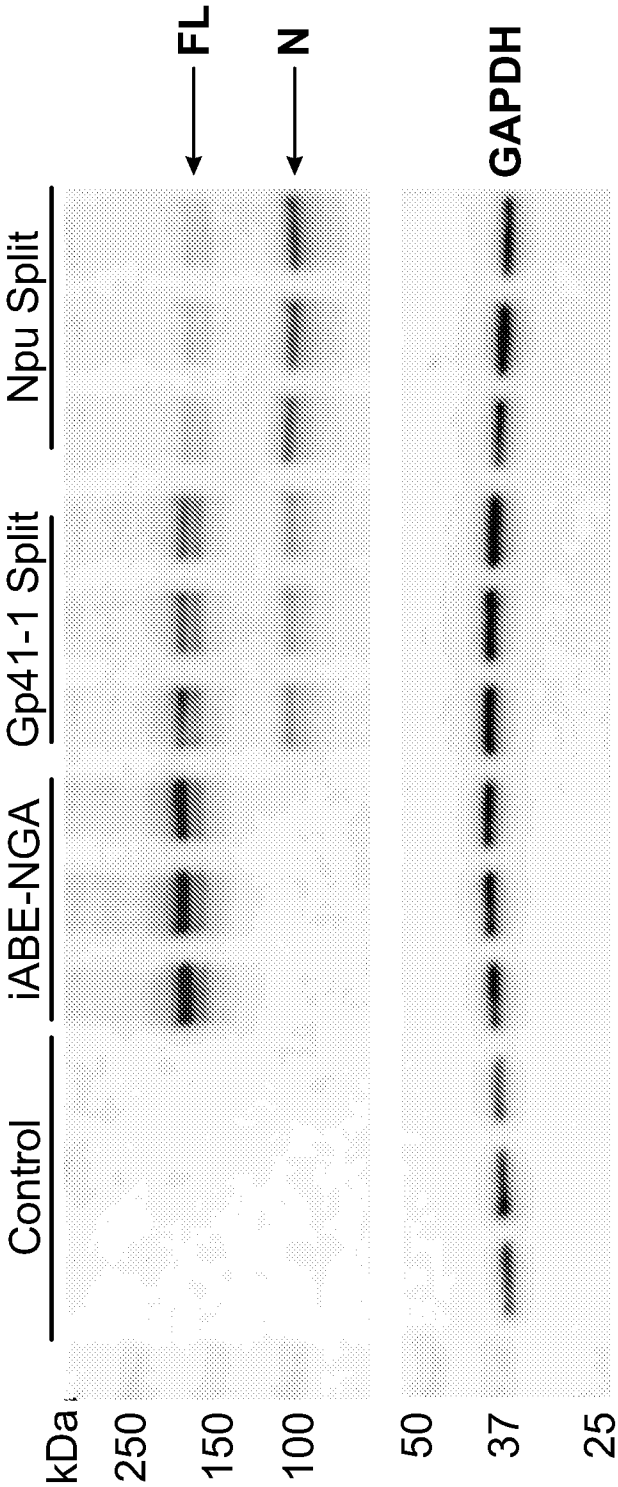


FIG. 6E

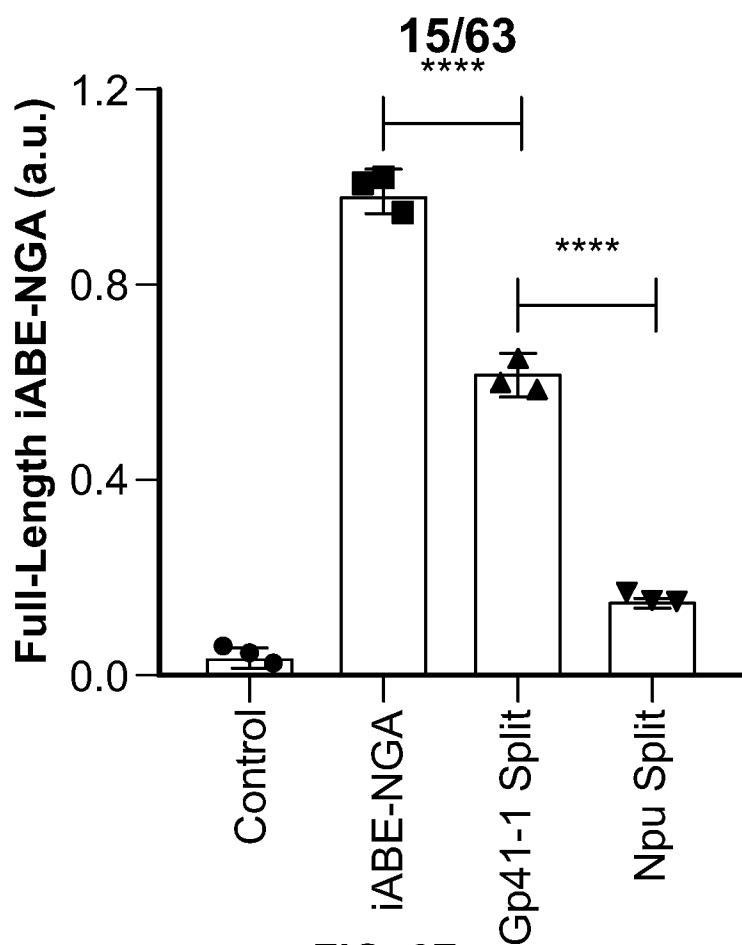


FIG. 6F

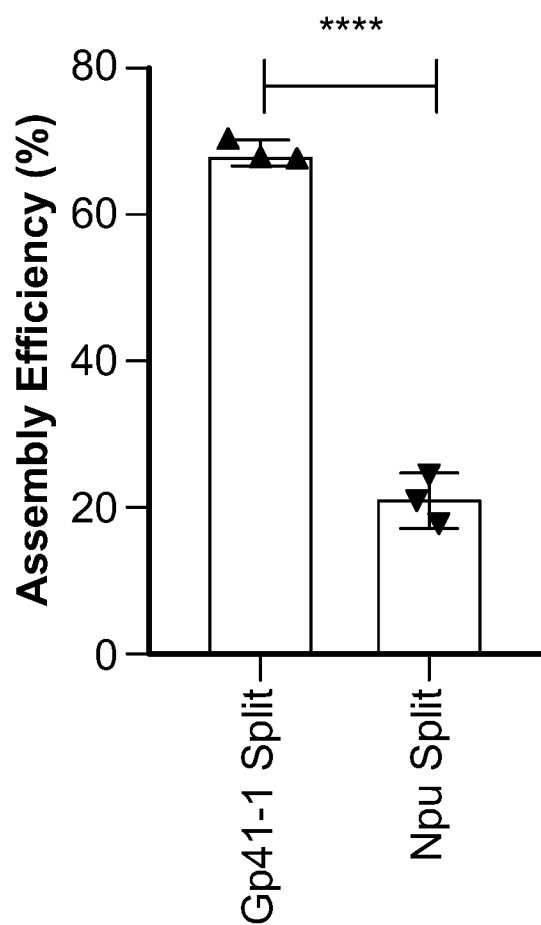


FIG. 6G

16/63

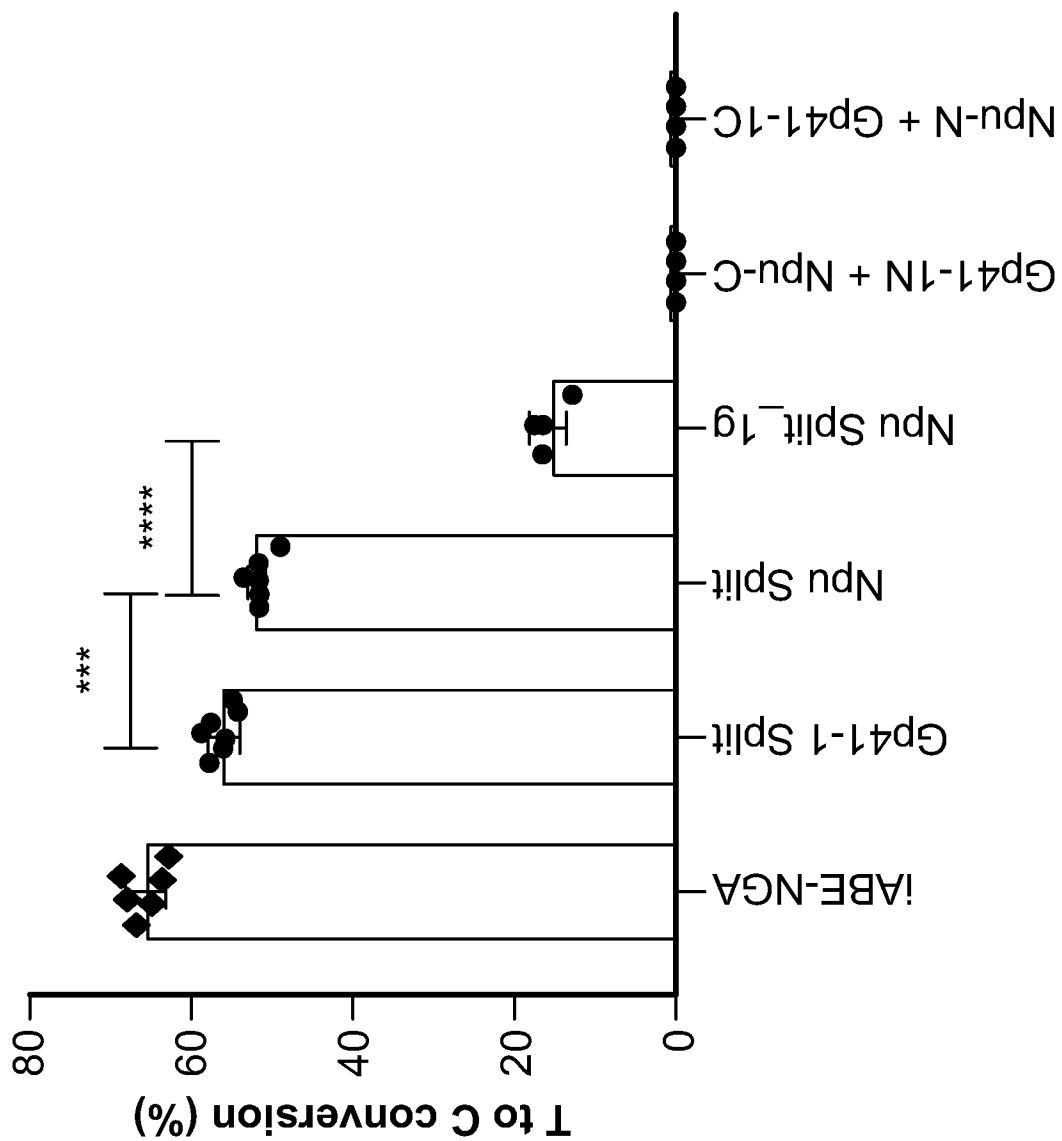


FIG. 6H

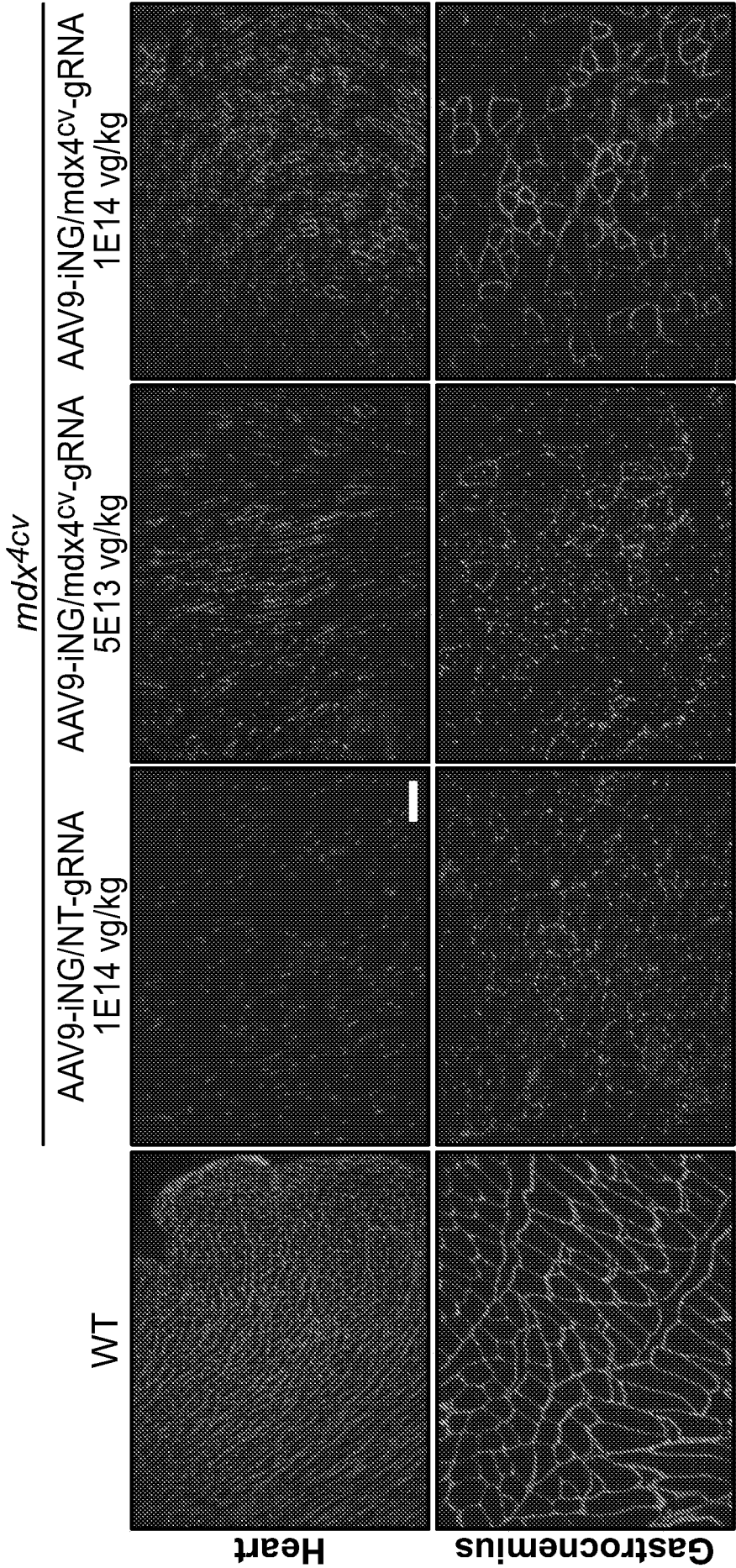


FIG. 7

18/63

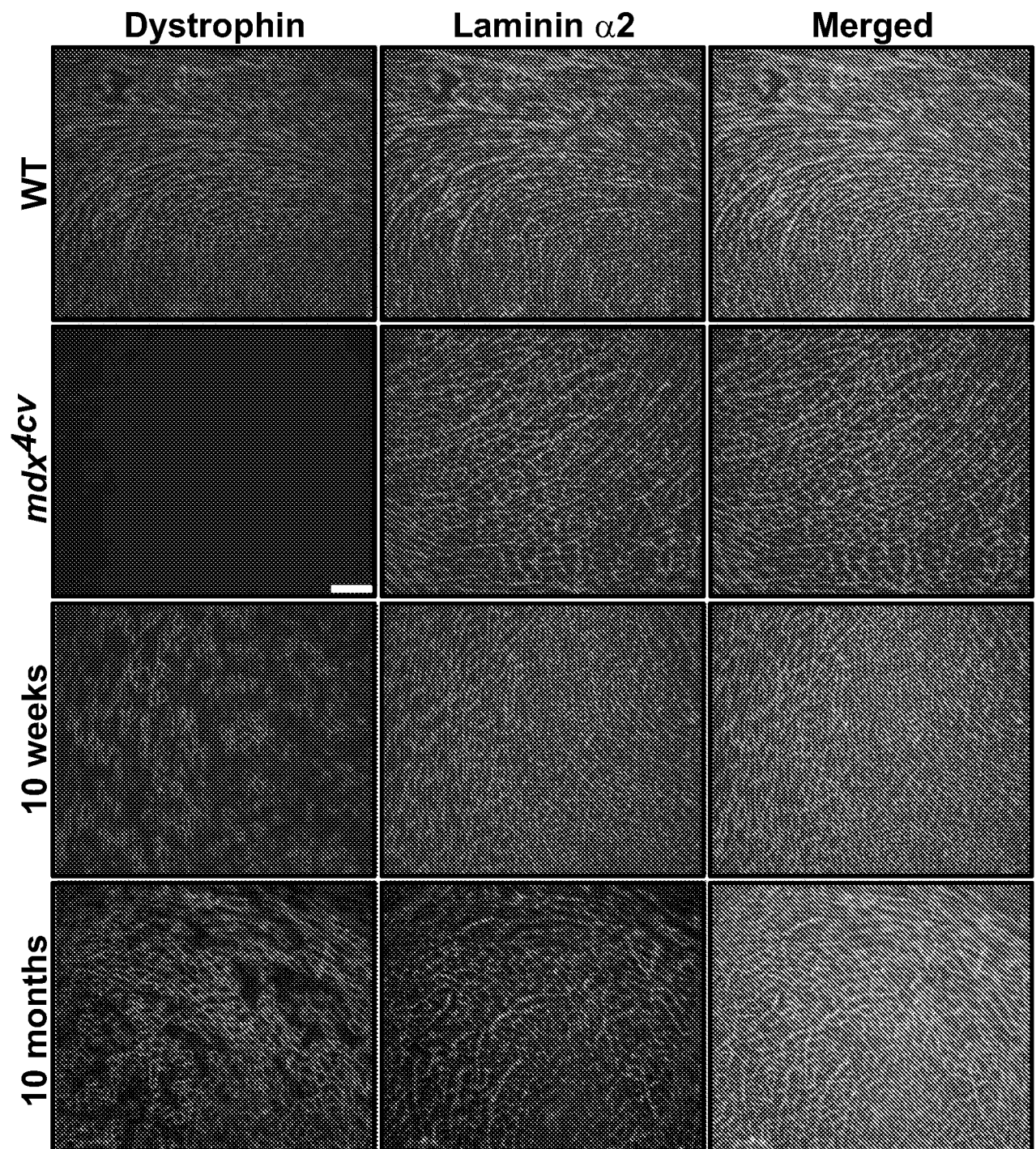
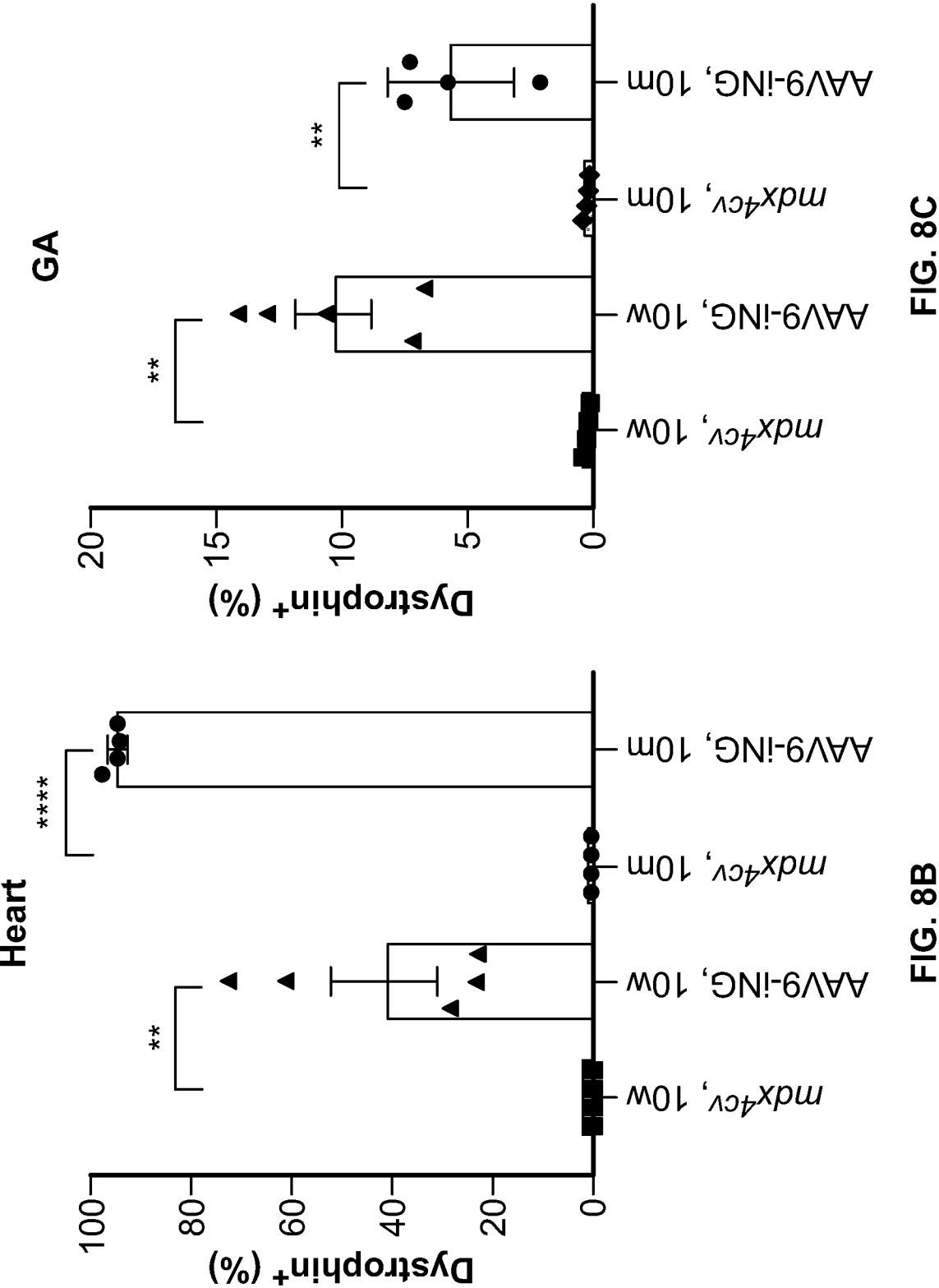
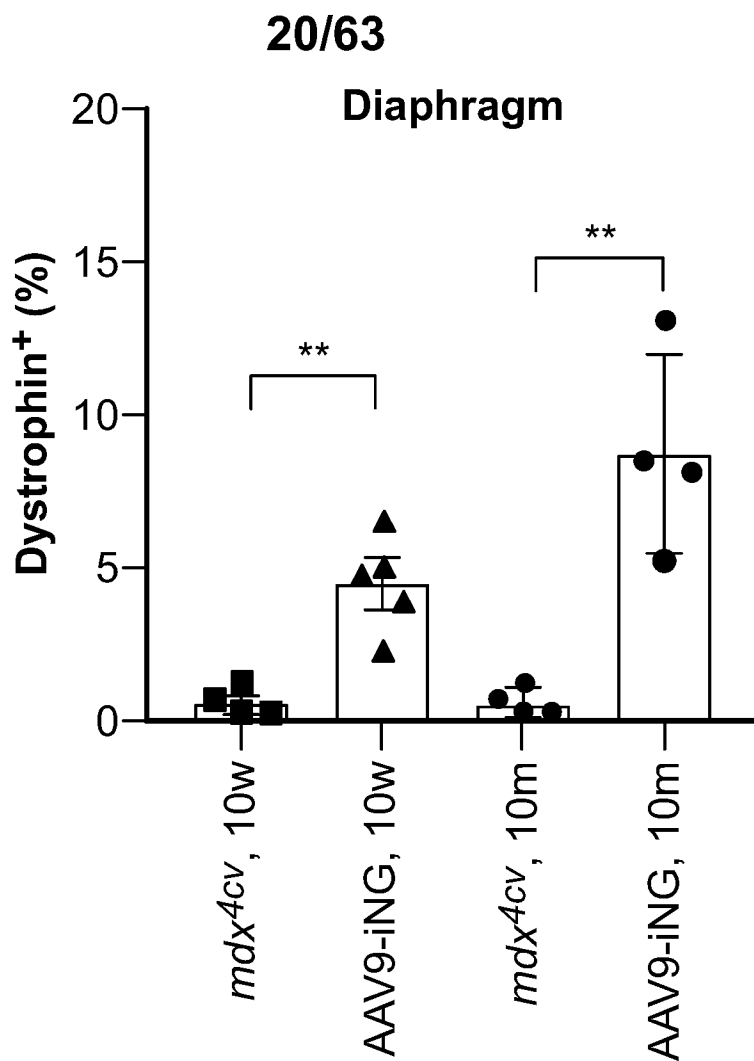
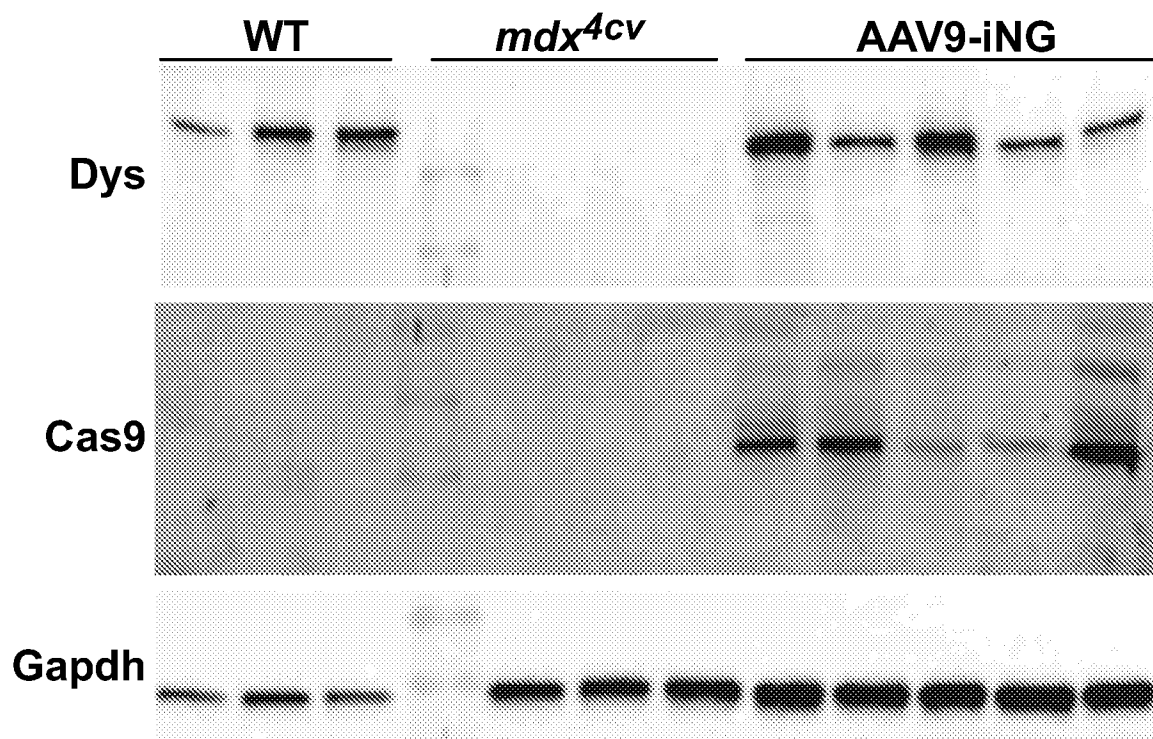
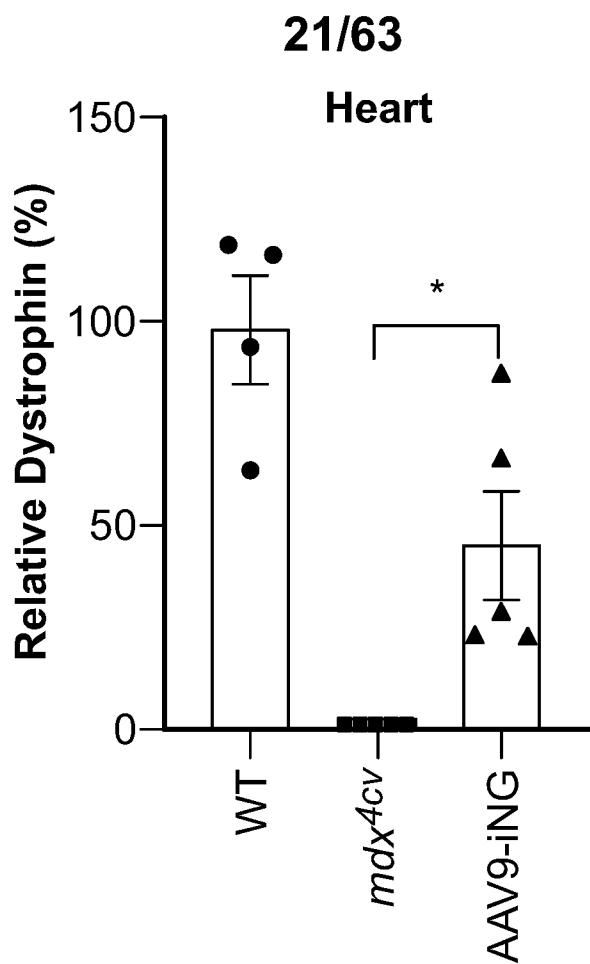
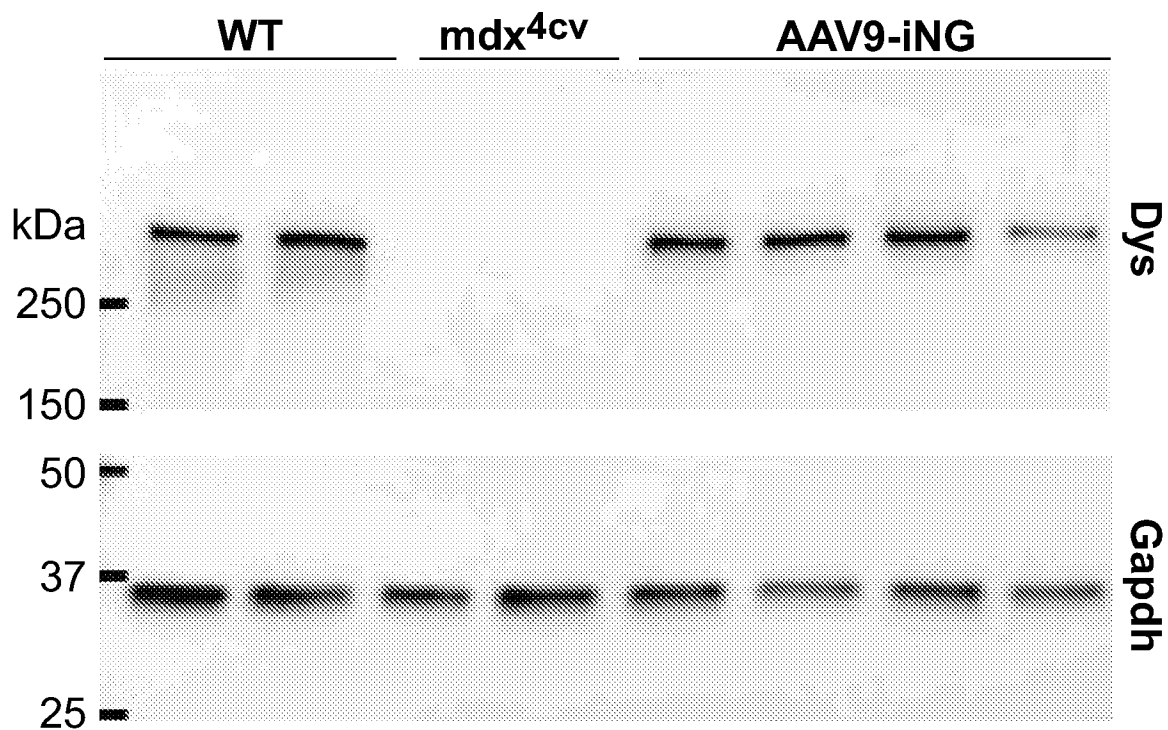


FIG. 8A



**FIG. 8D****FIG. 8E**

**FIG. 8F****FIG. 8G**

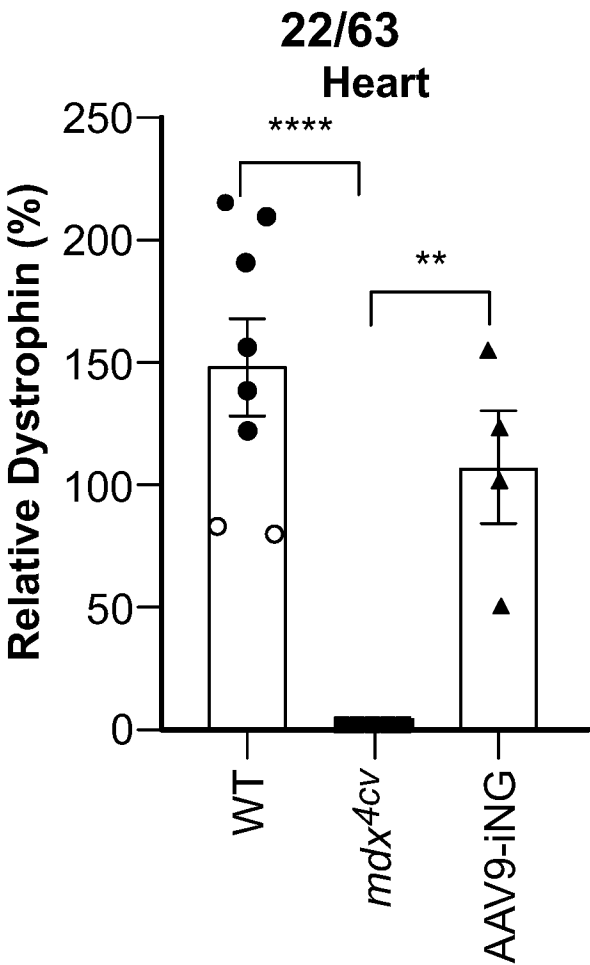


FIG. 8H

GATC

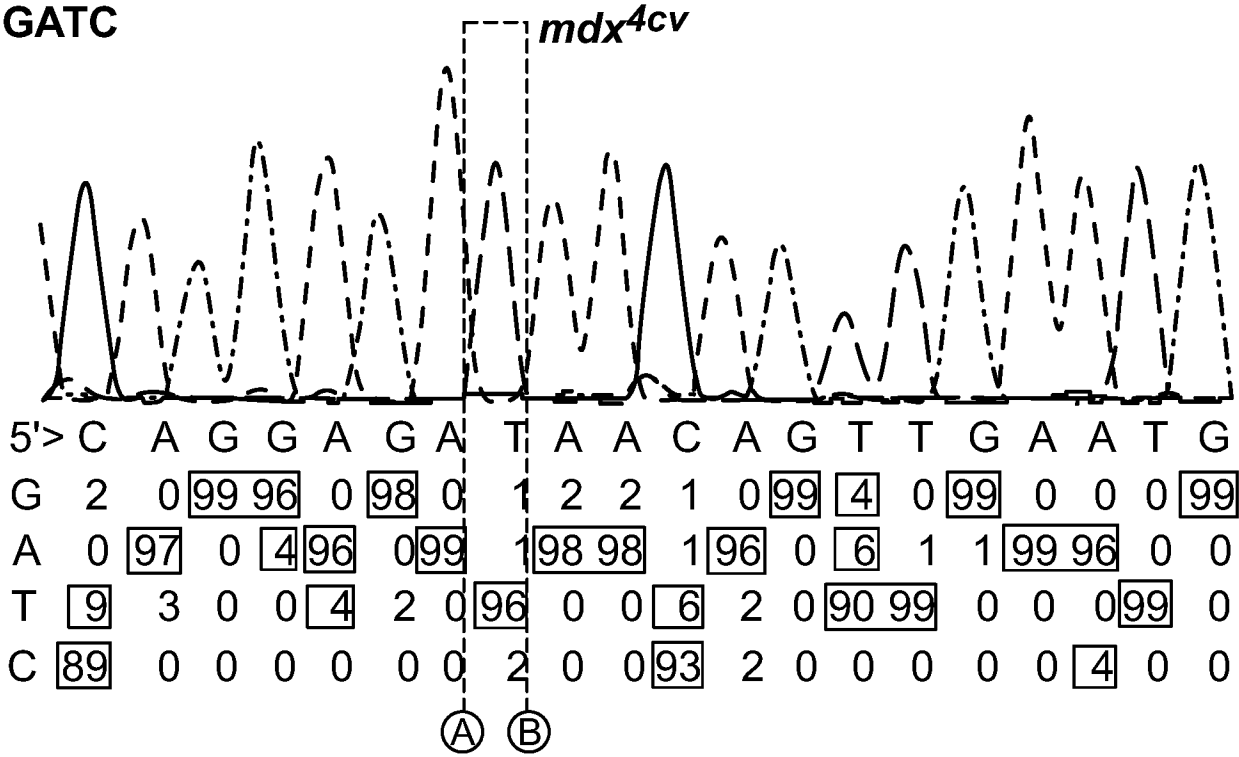
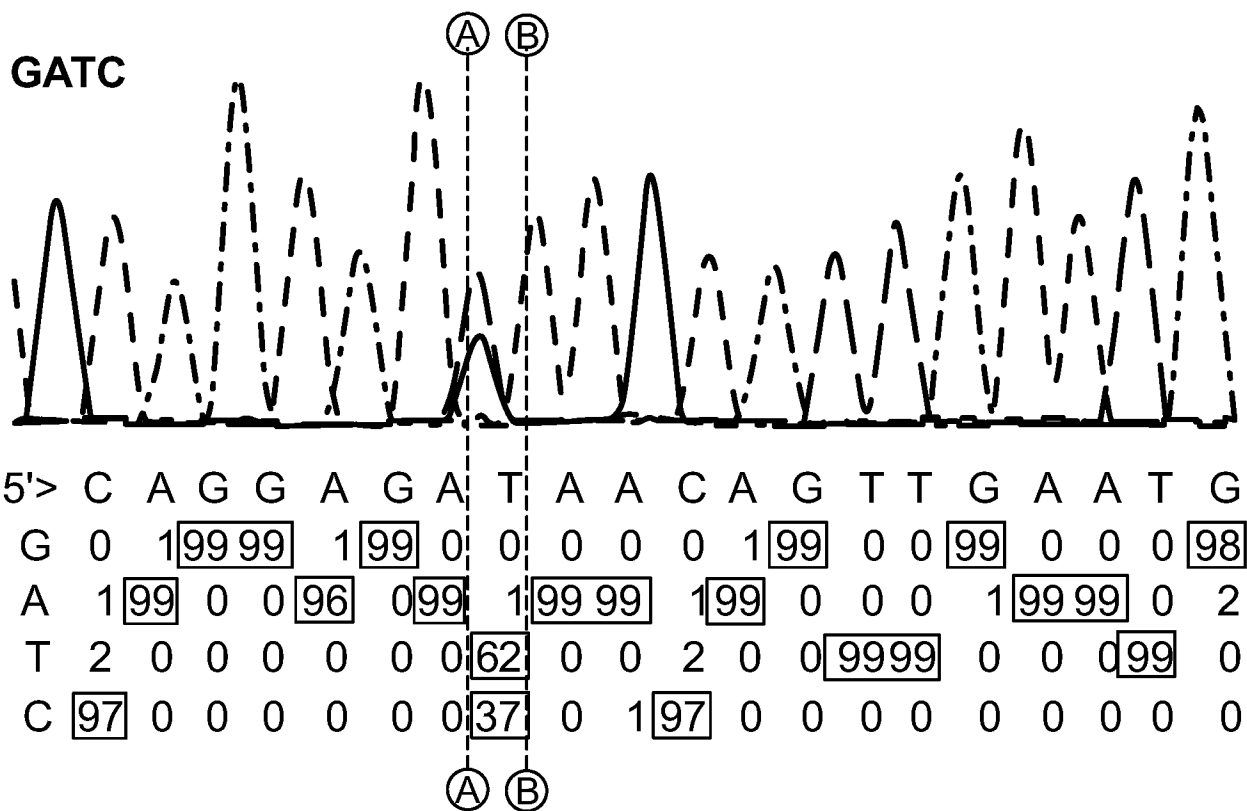
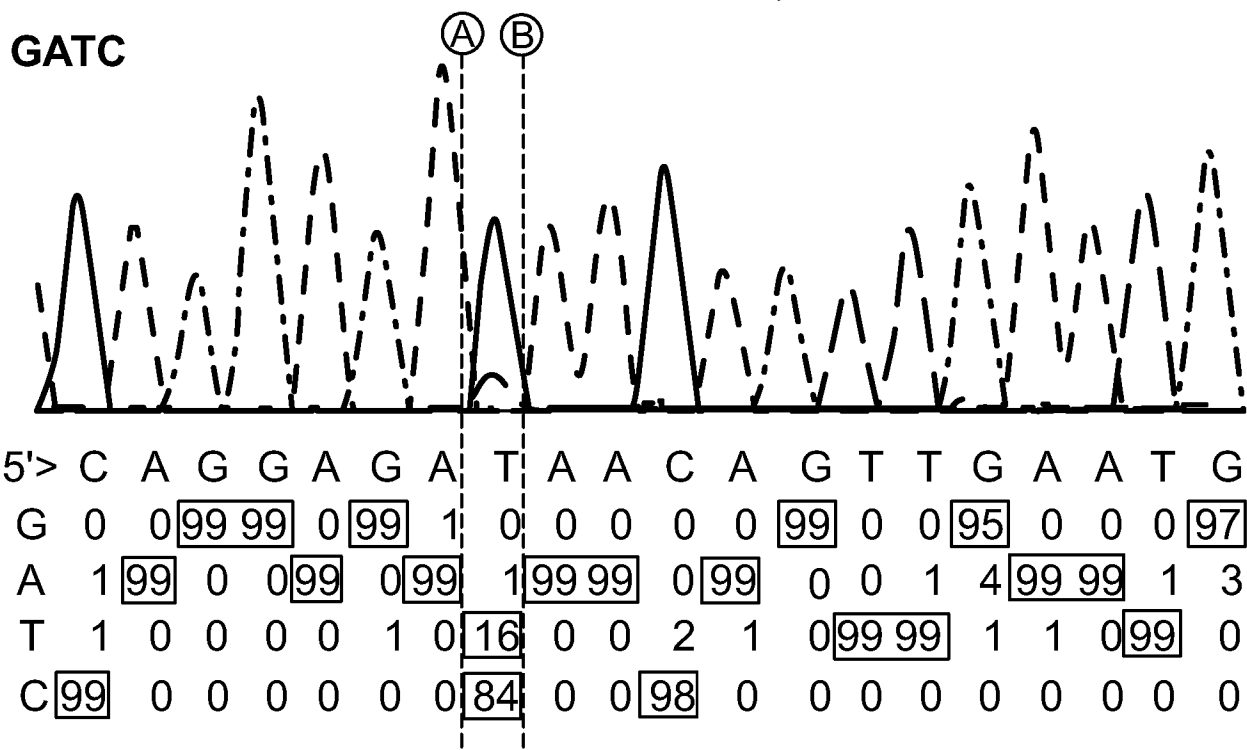


FIG. 8I

23/63

mdx^{4cv}+AAV9-iNG, 10w*mdx*^{4cv}+AAV9-iNG, 10mFIG. 8I
(Cont'd)

SUBSTITUTE SHEET (RULE 26)

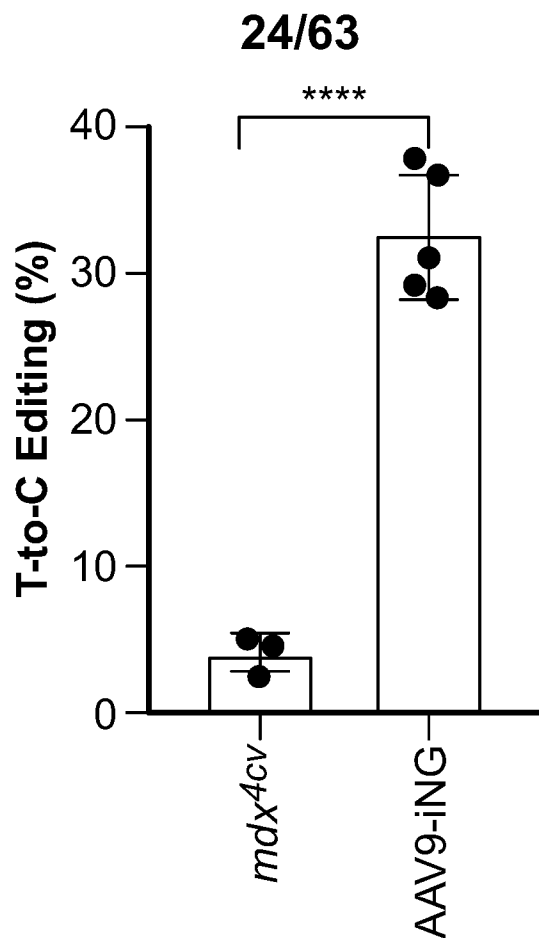


FIG. 8J

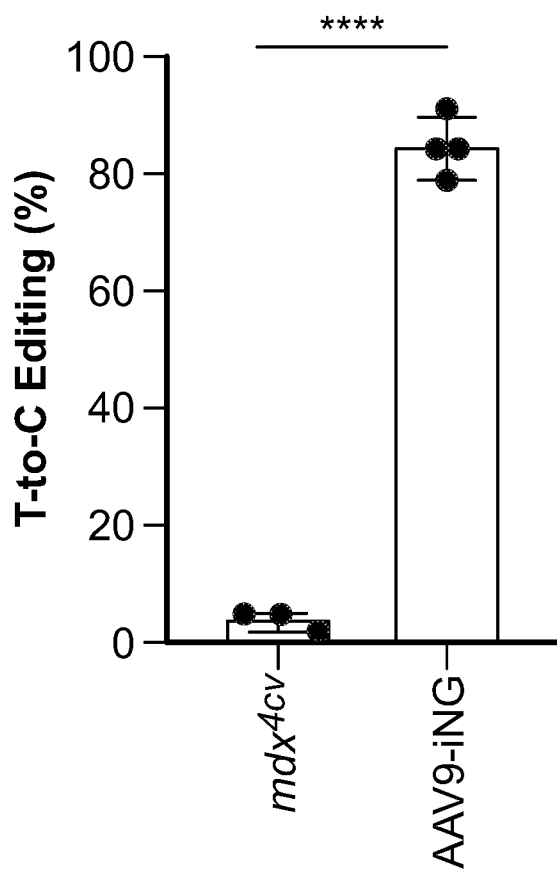


FIG. 8K

25/63

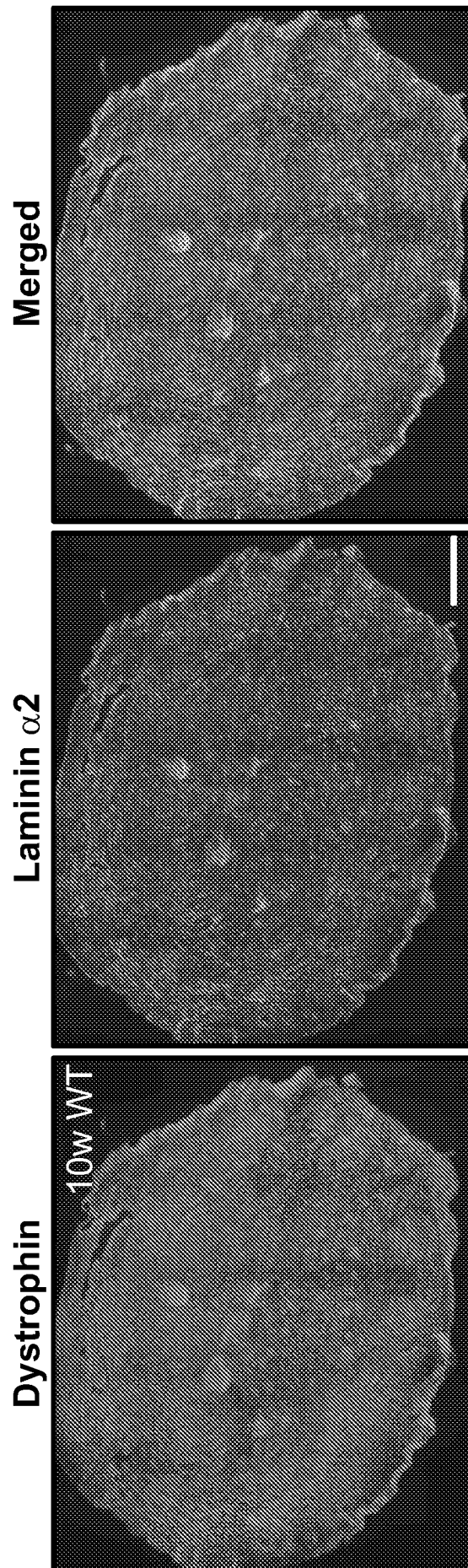
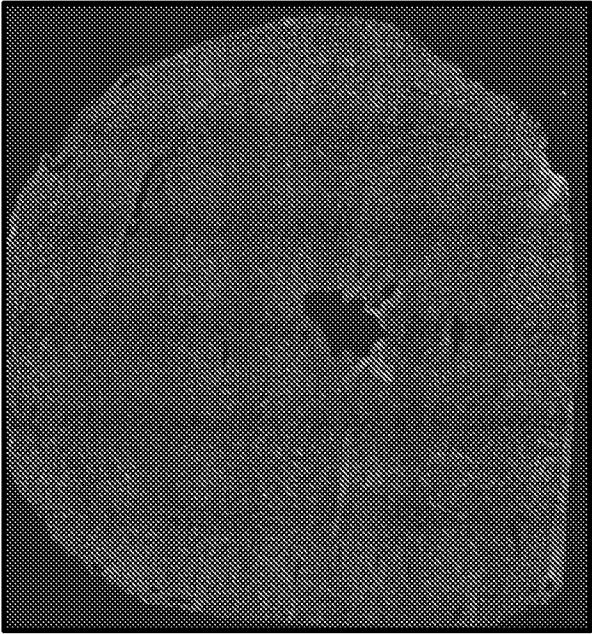


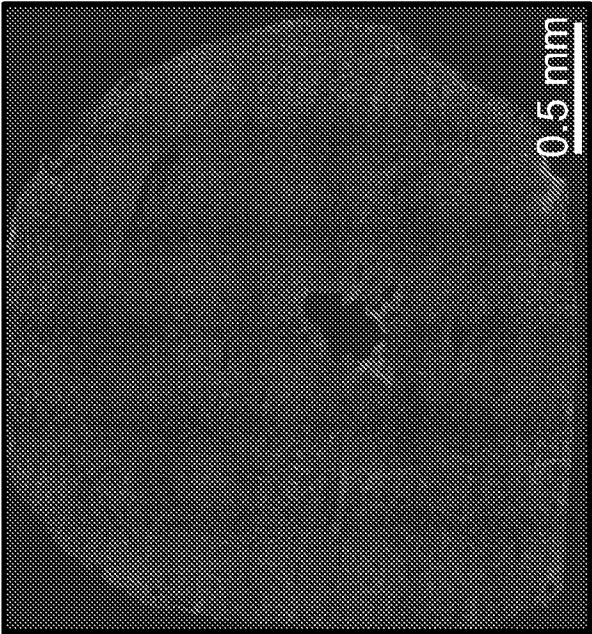
FIG. 9

26/63

Merged

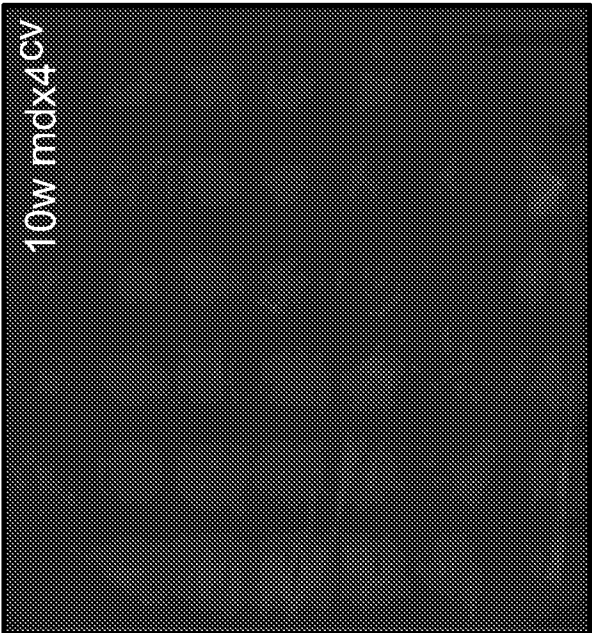


Laminin $\alpha 2$



0.5 mm

Dystrophin



10w mdx4cv

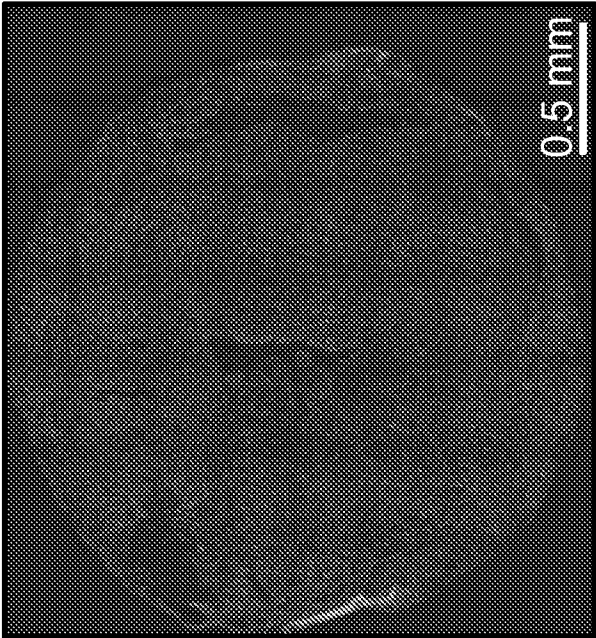
FIG. 10

27/63

Merged



Laminin $\alpha 2$



Dystrophin



FIG. 11

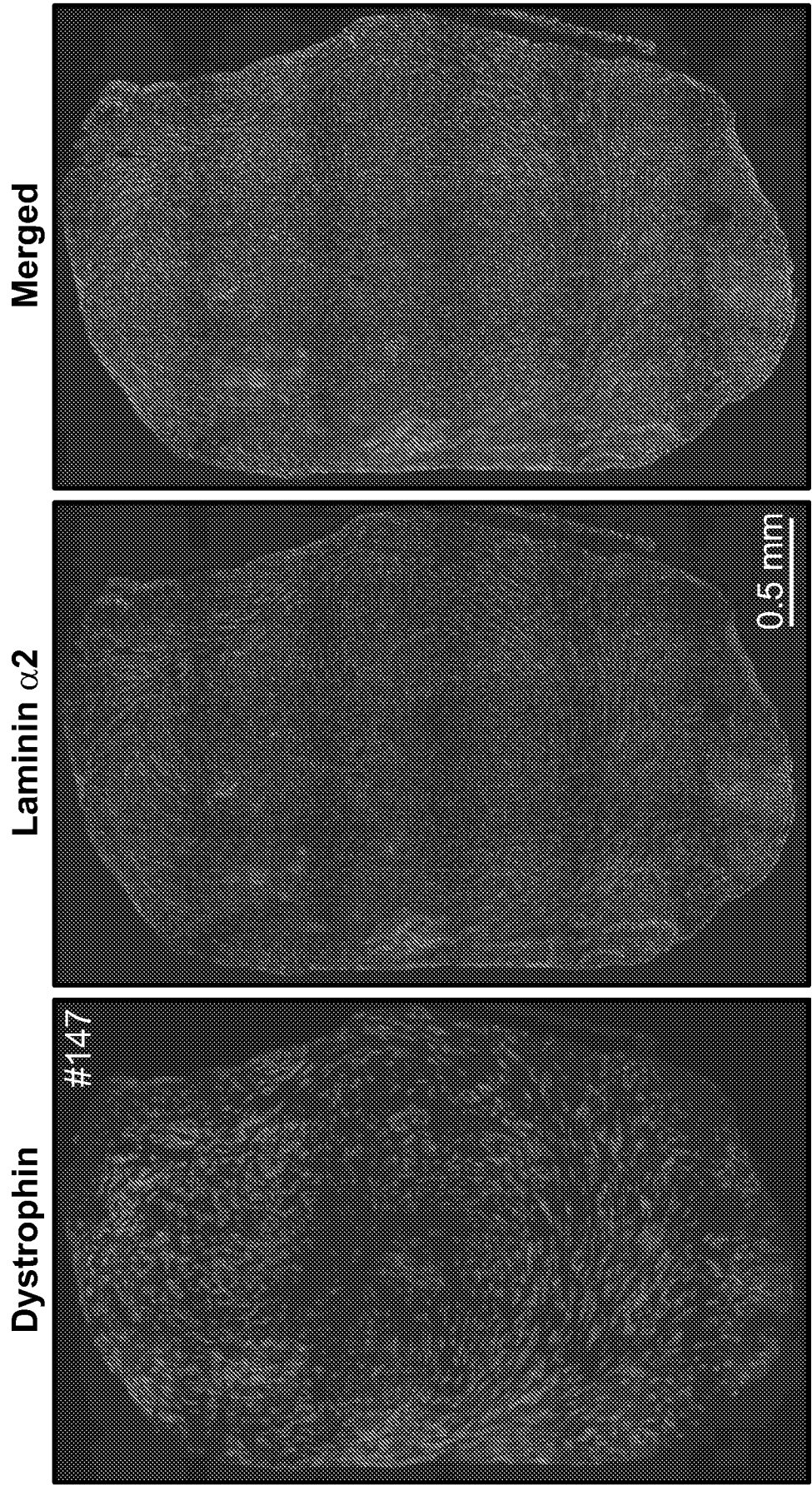


FIG. 12

29/63

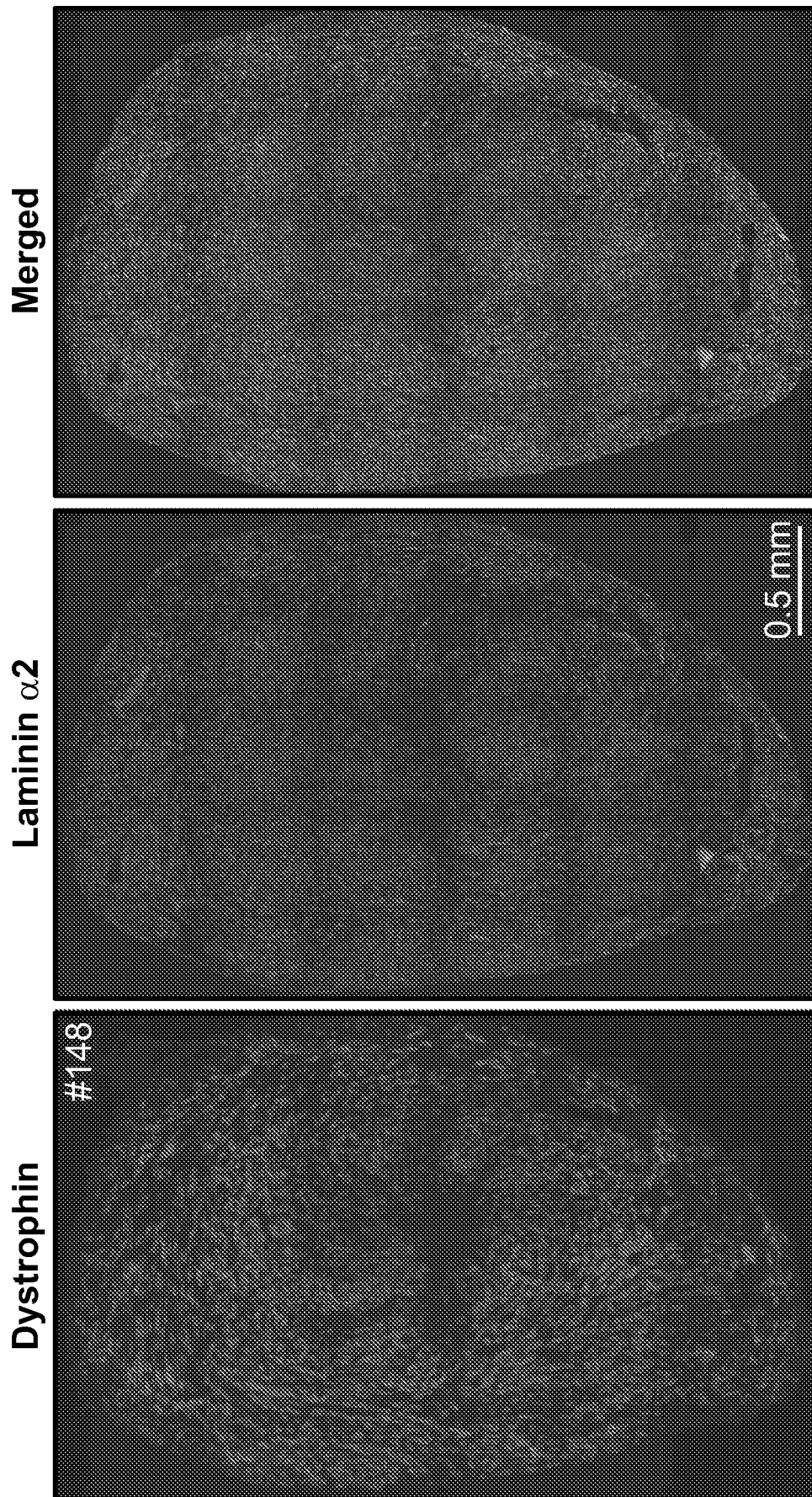


FIG. 13

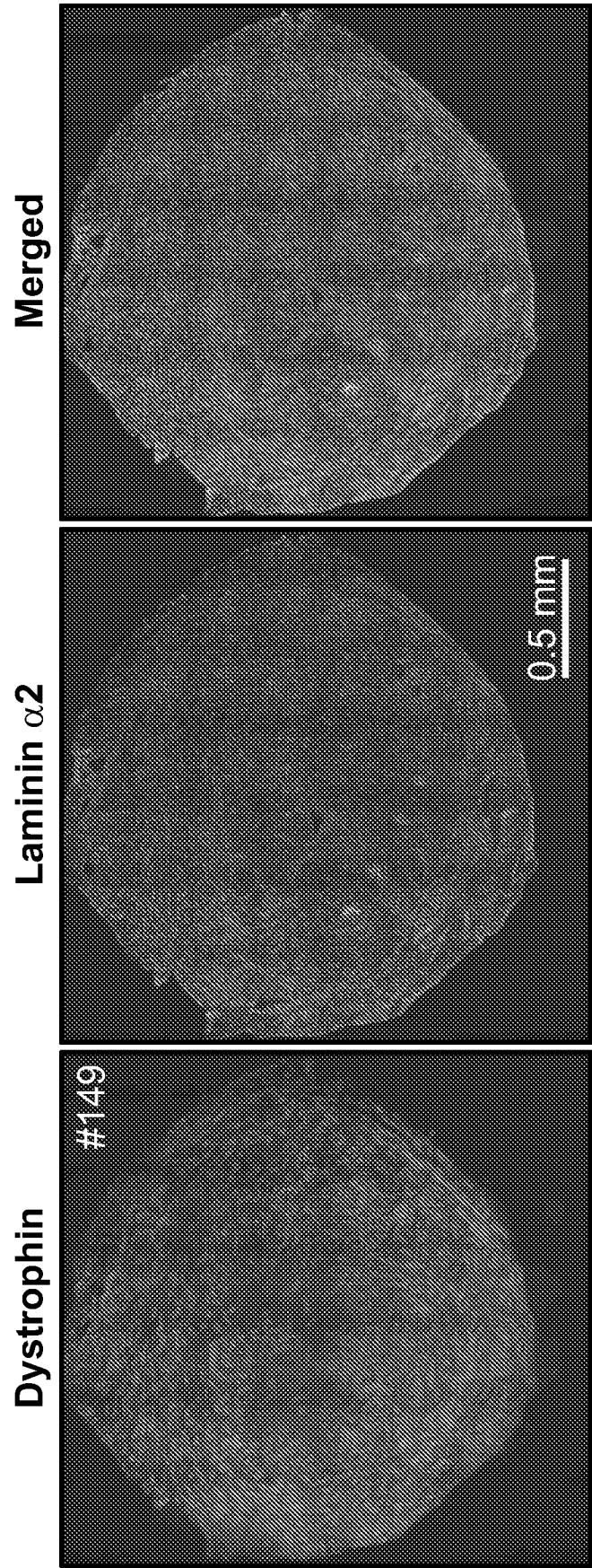


FIG. 14



FIG. 15

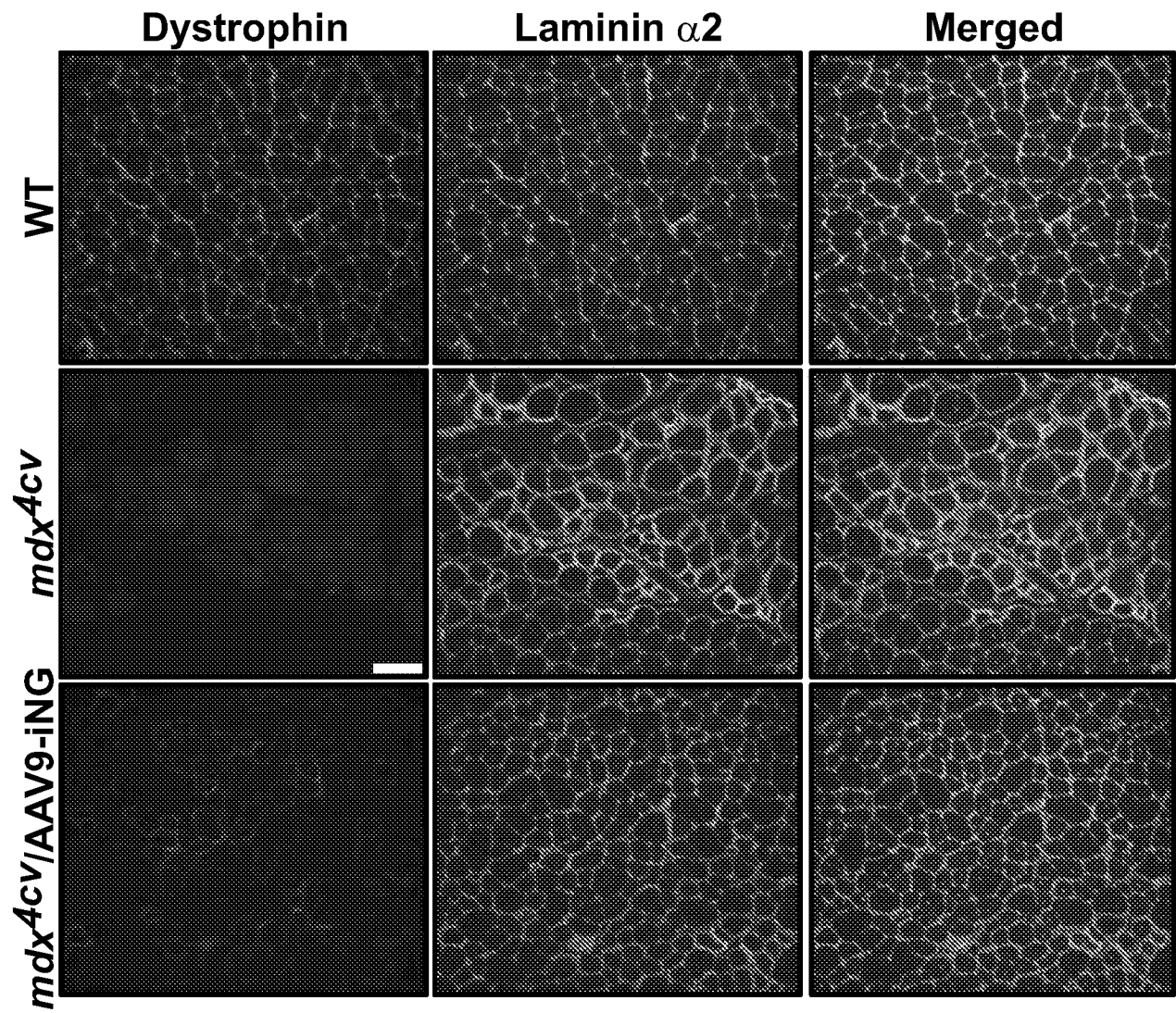


FIG. 16A

33/63

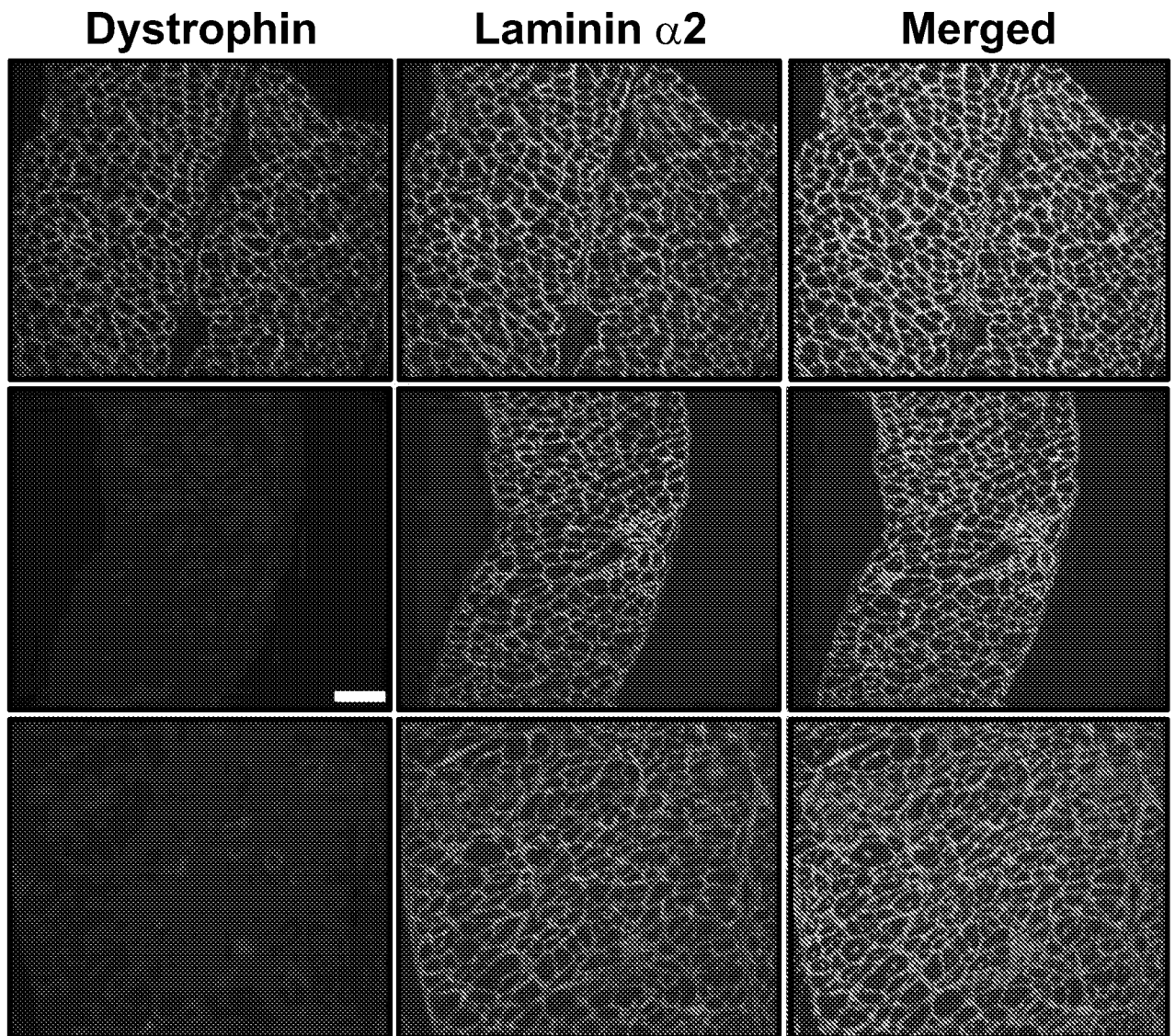


FIG. 16B

34/63

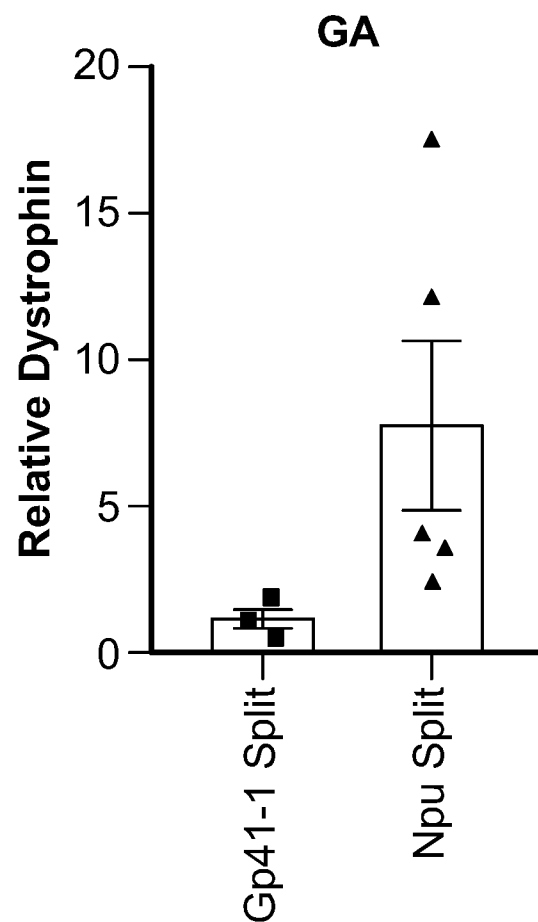
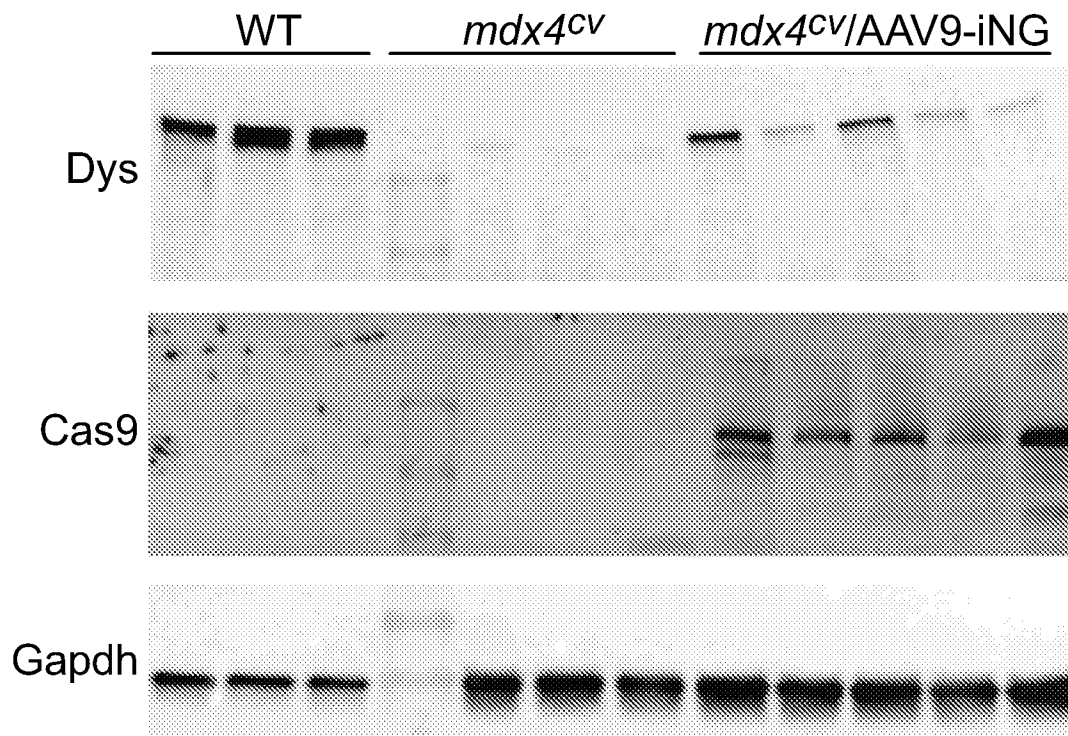


FIG. 17

SUBSTITUTE SHEET (RULE 26)

35/63

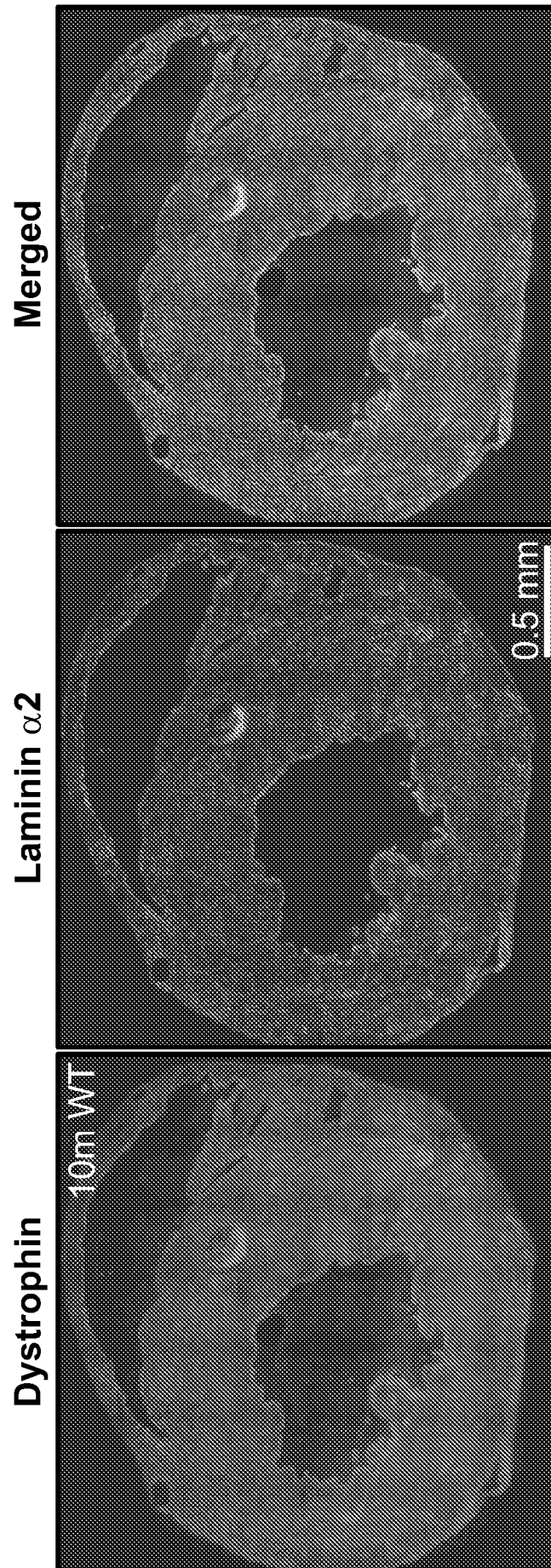


FIG. 18

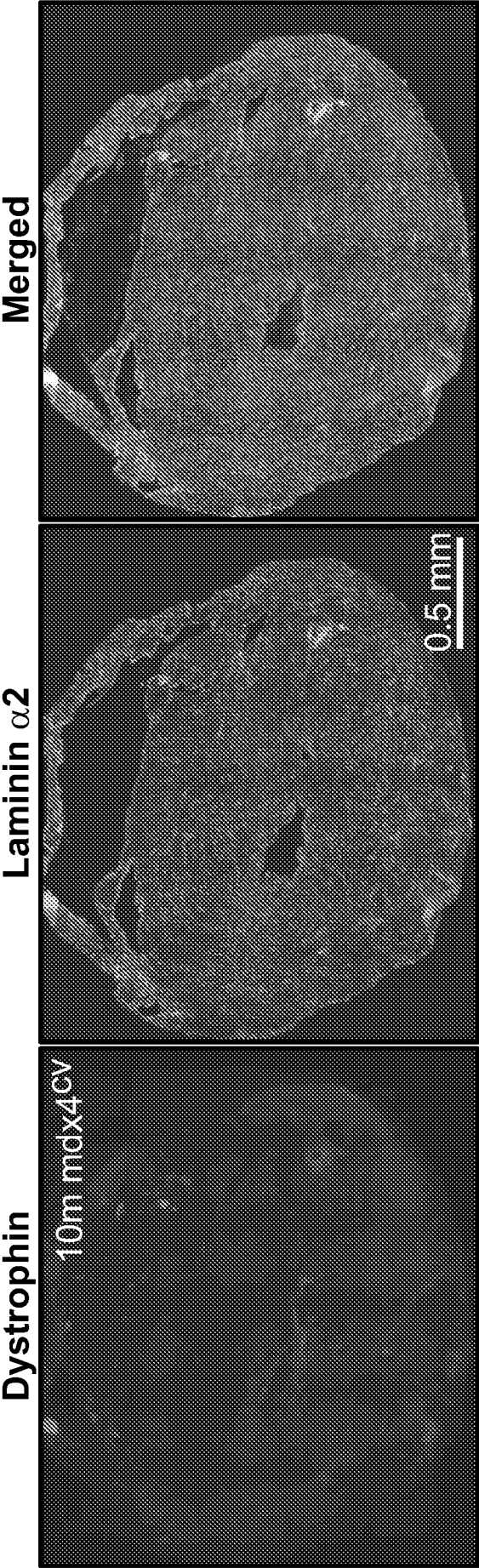


FIG. 19



FIG. 20

37/63

Merged



Laminin $\alpha 2$



0.5 mm

Dystrophin

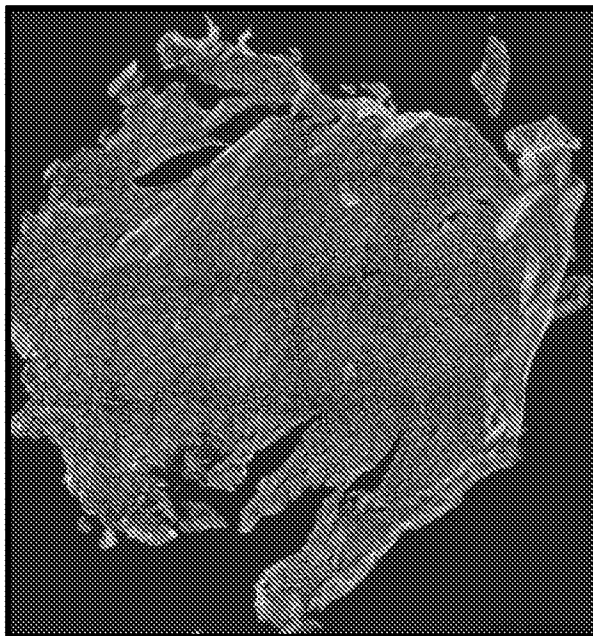


#1982

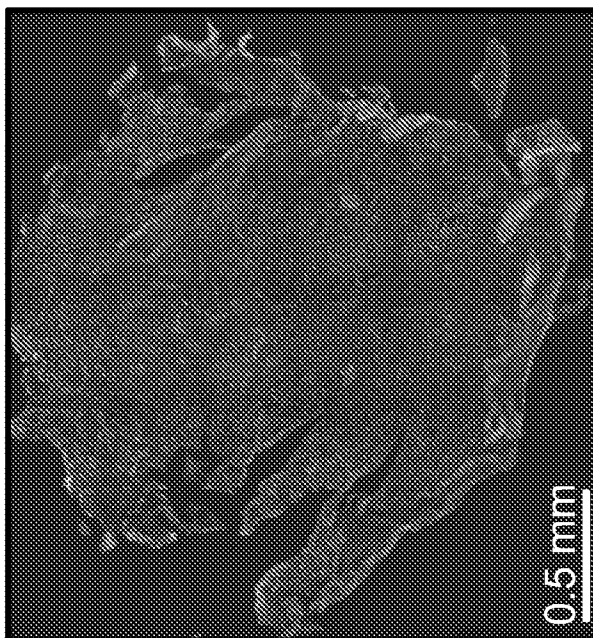
FIG. 21

38/63

Merged



Laminin $\alpha 2$



Dystrophin

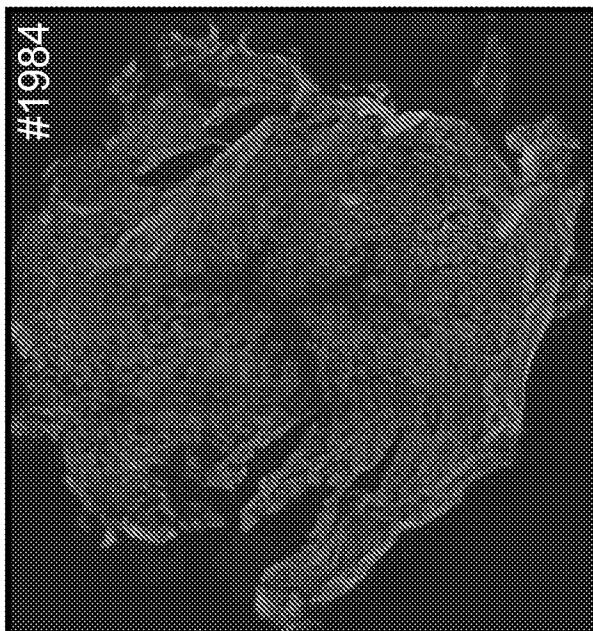


FIG. 22

39/63

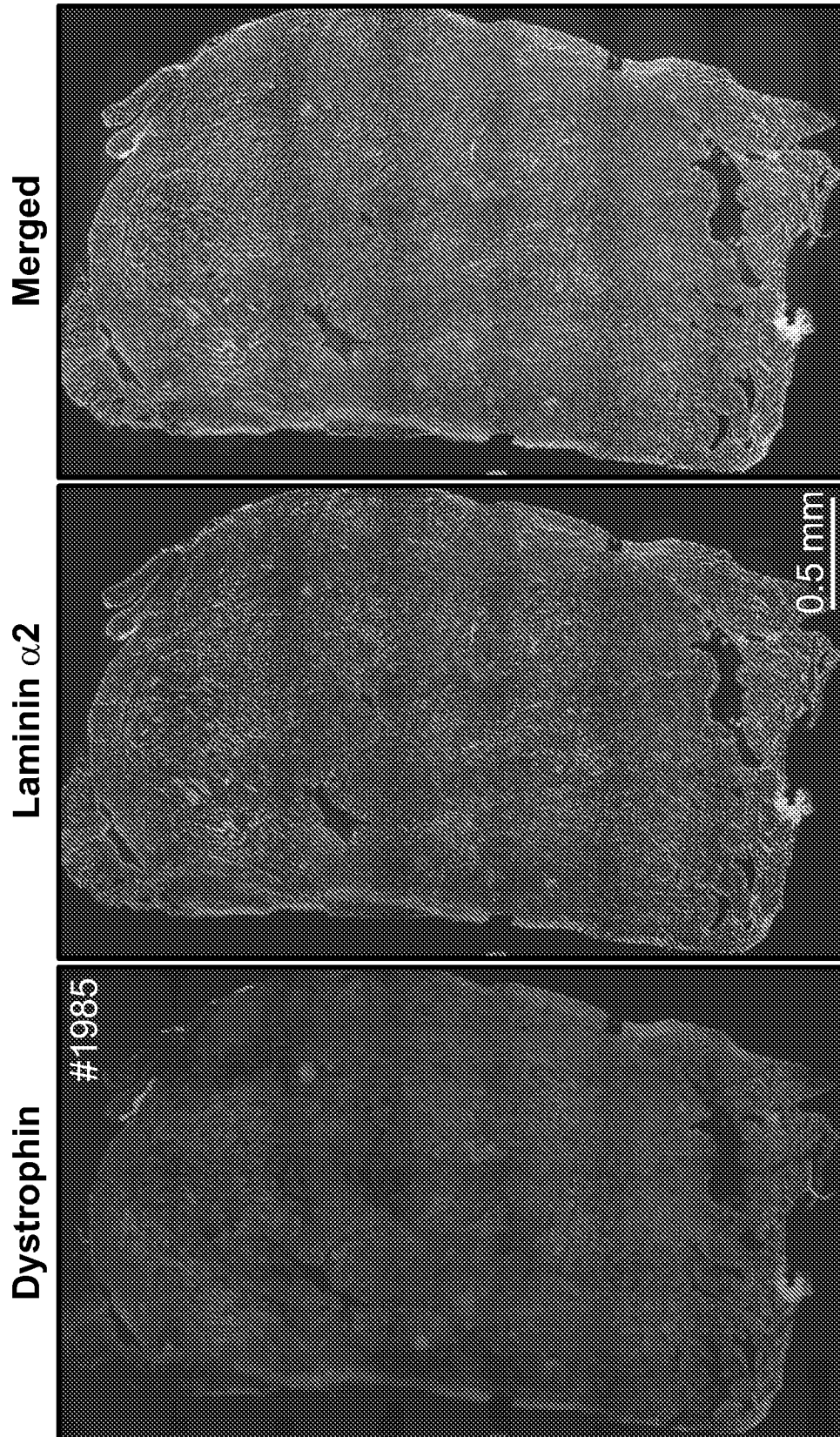


FIG. 23

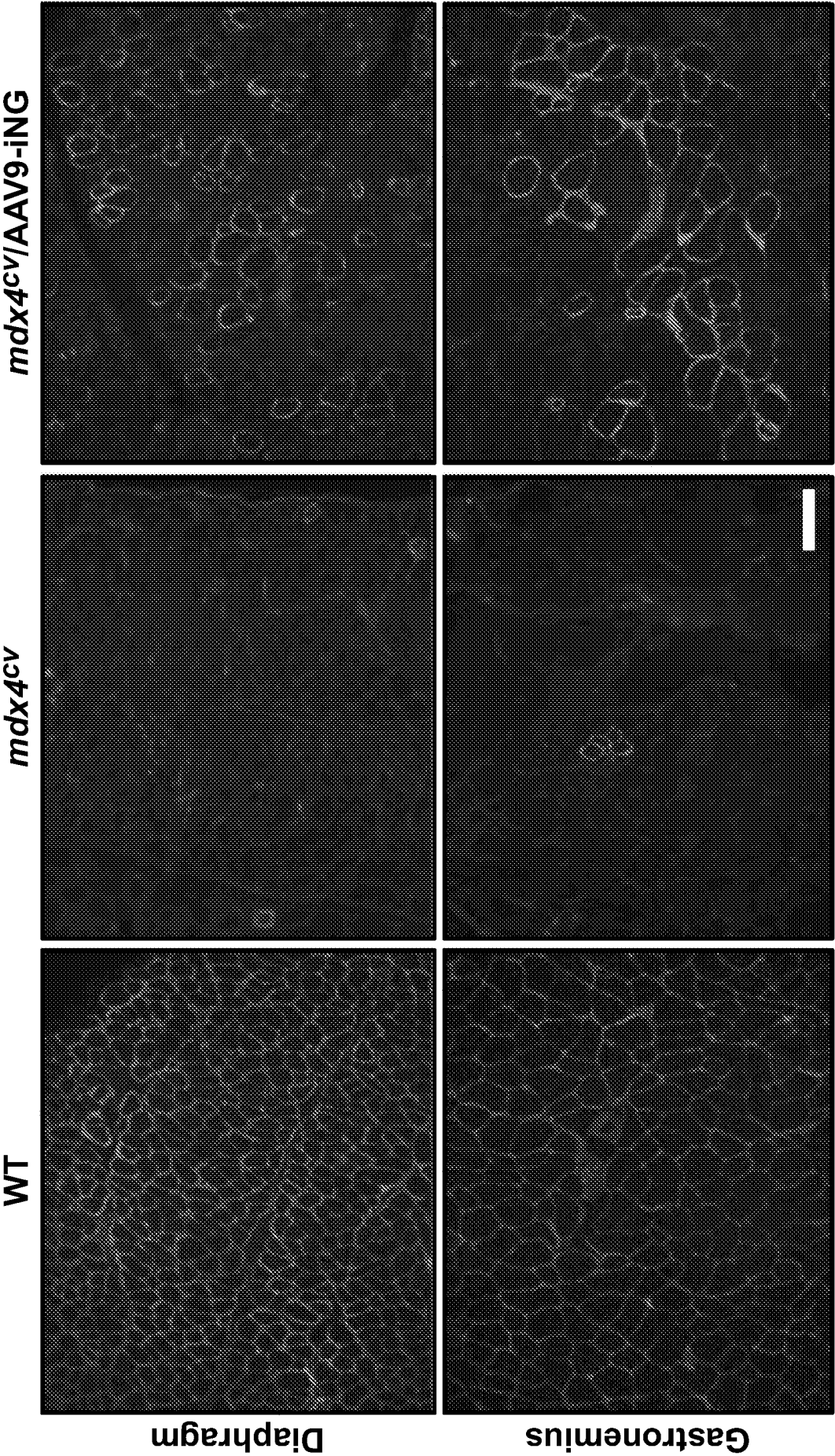


FIG. 24A

41/63

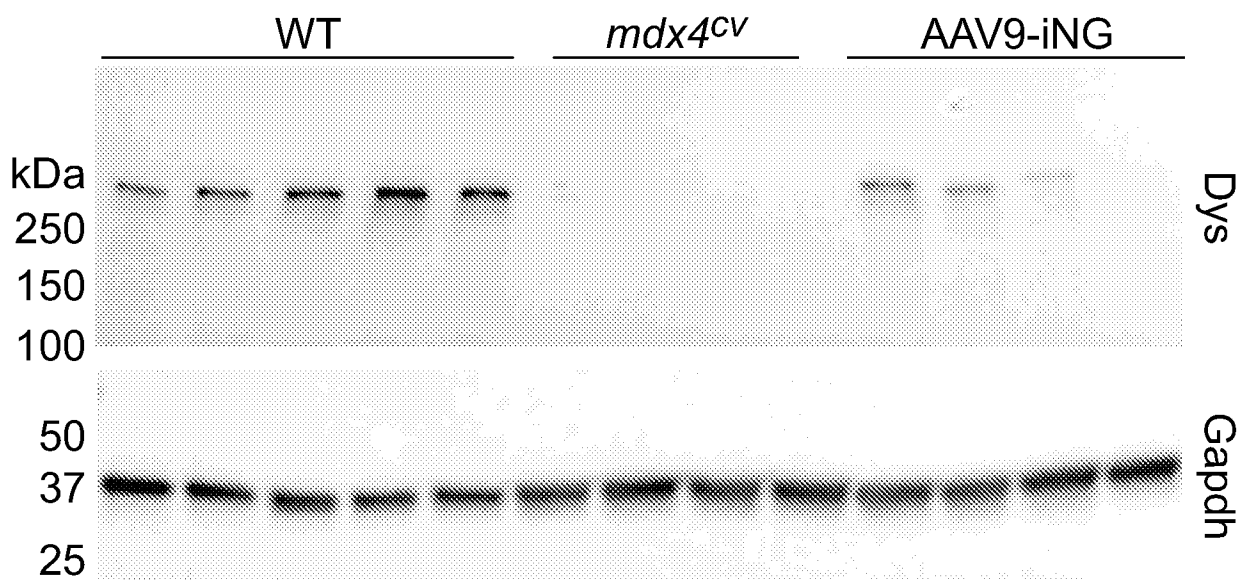


FIG. 24B

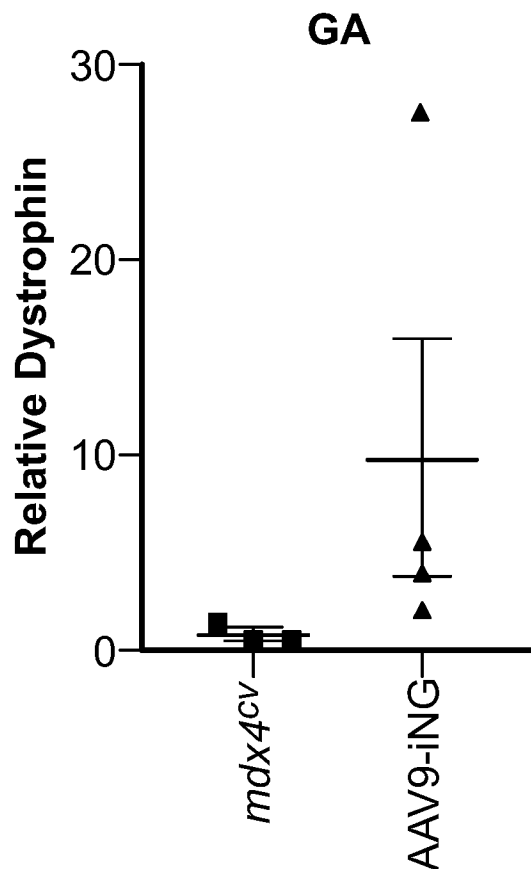


FIG. 24C

42/63

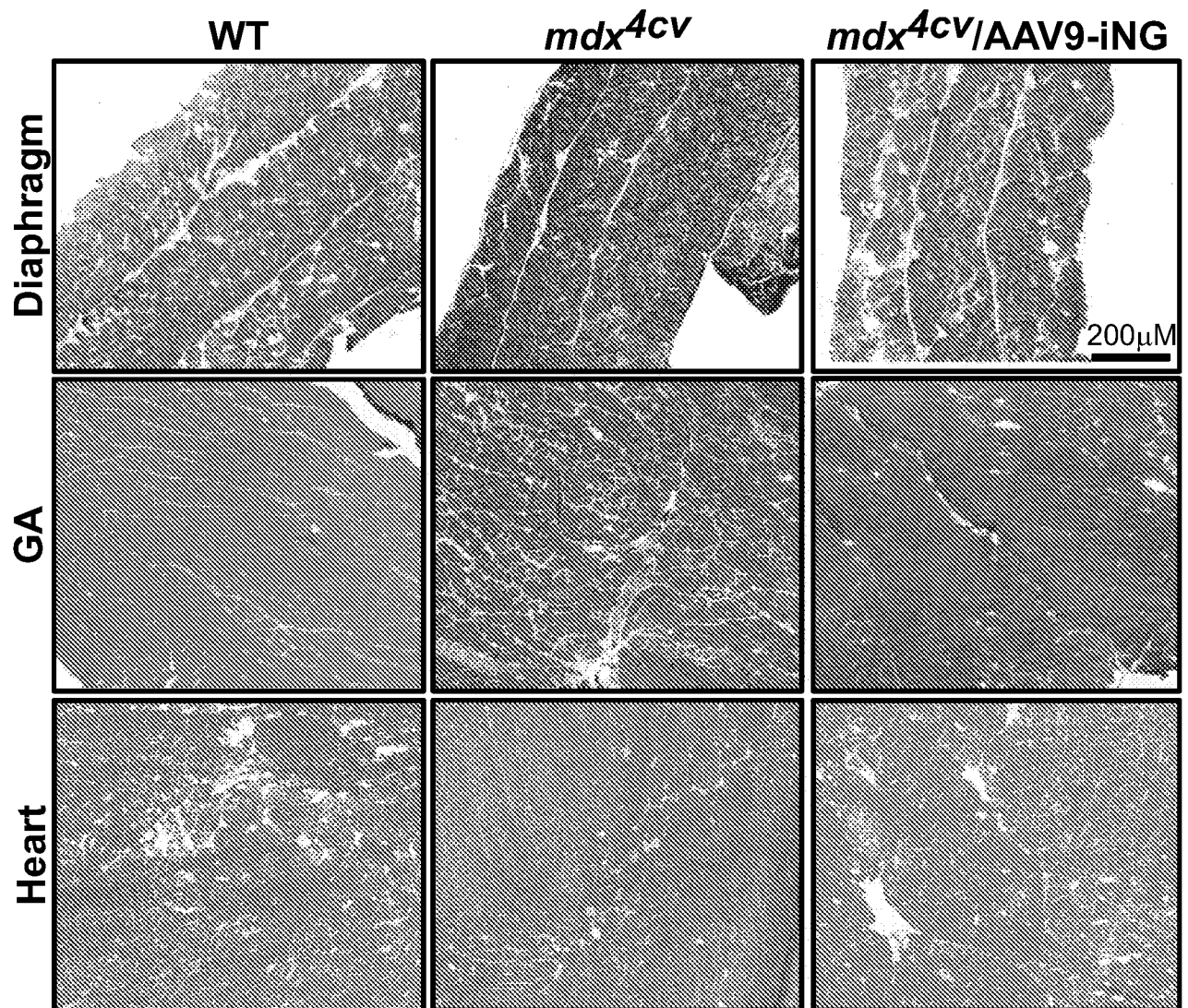
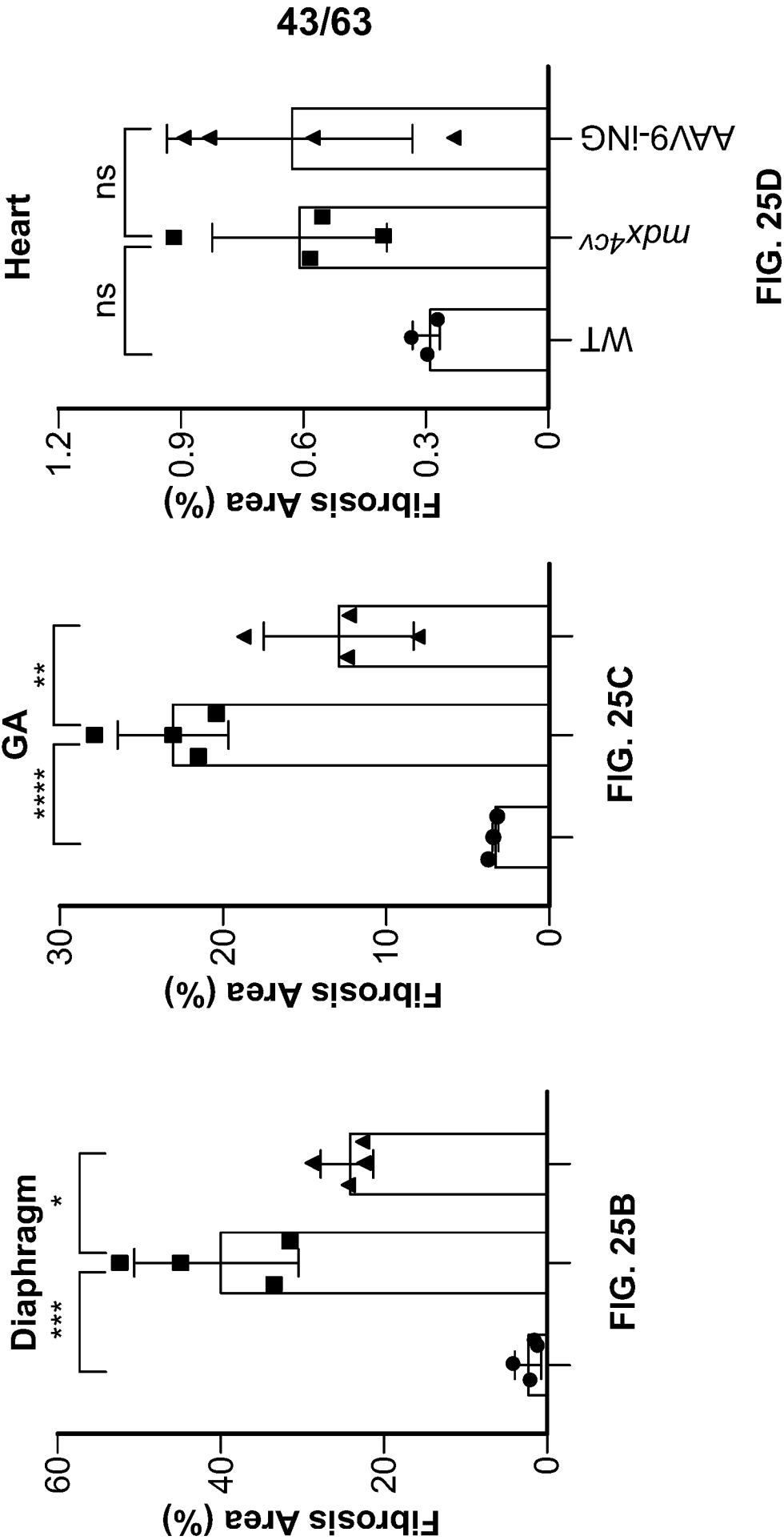
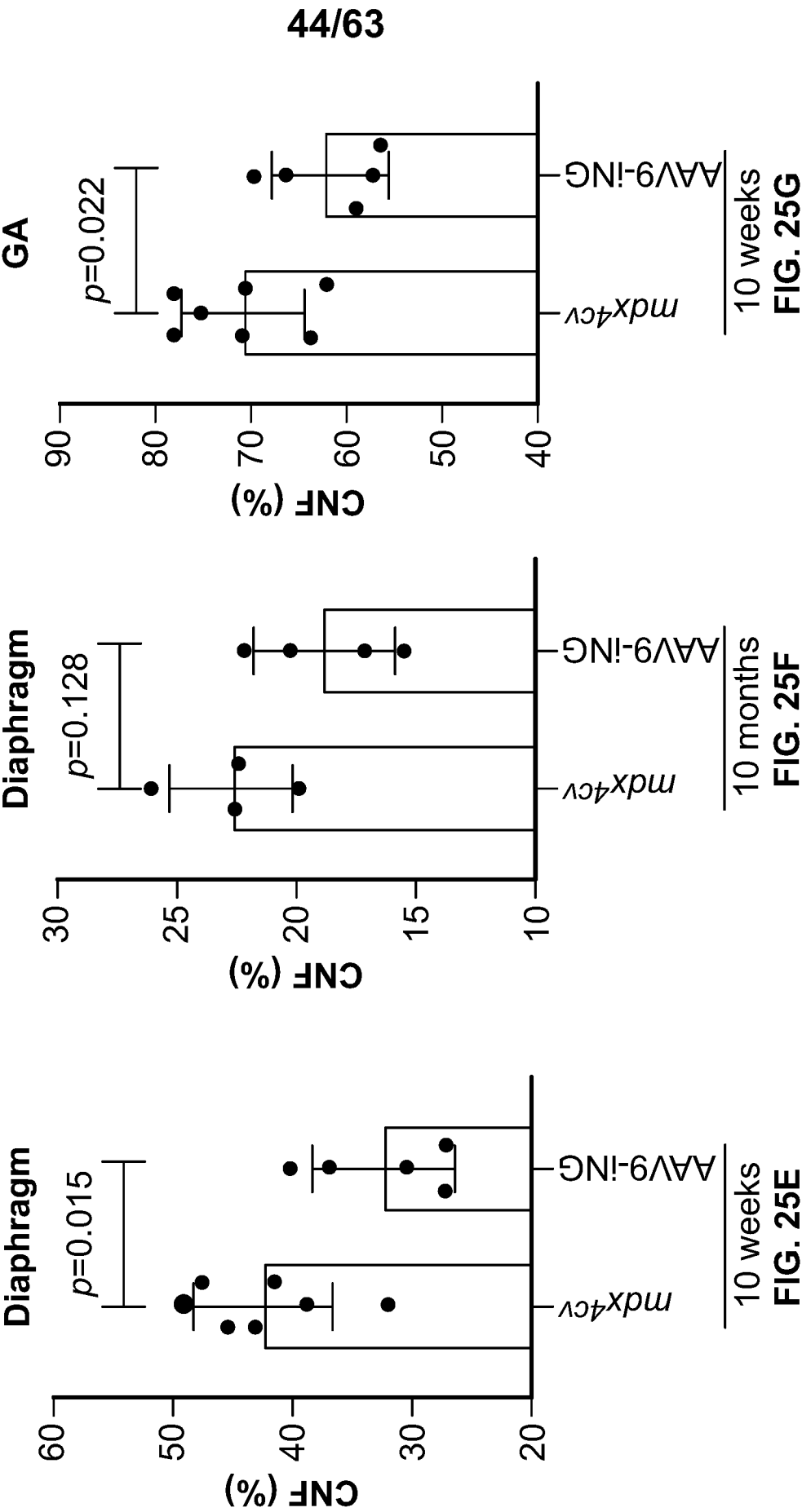


FIG. 25A





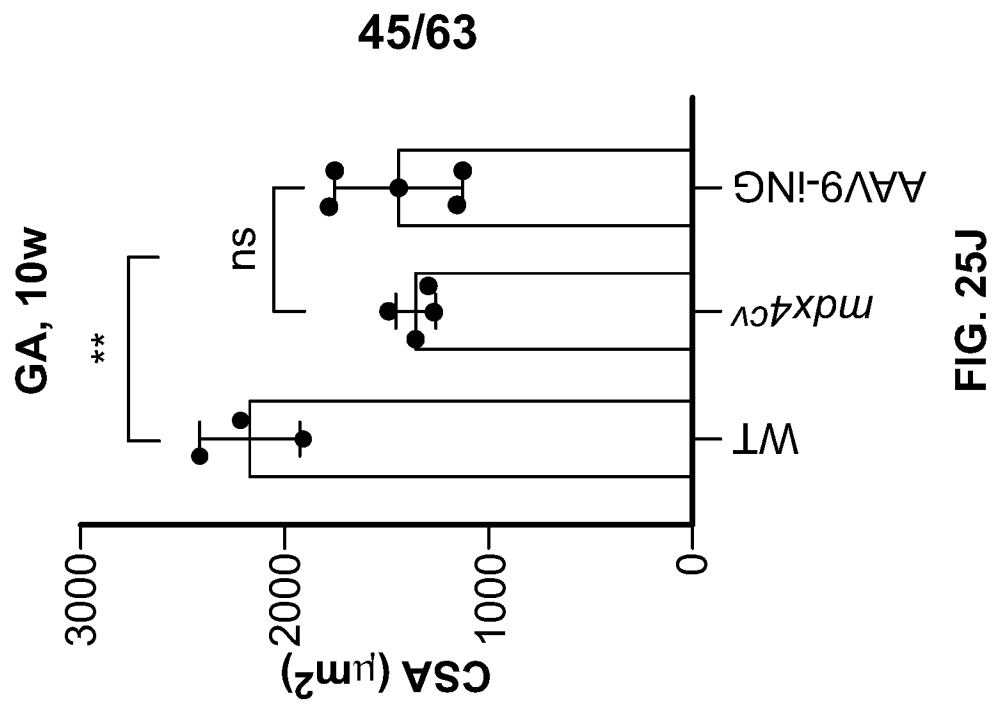


FIG. 25J

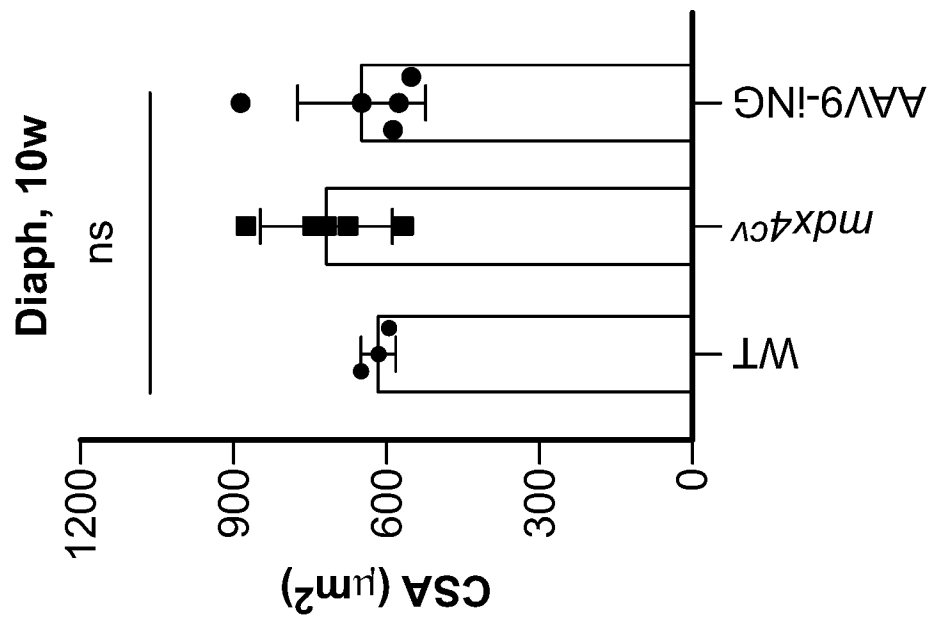
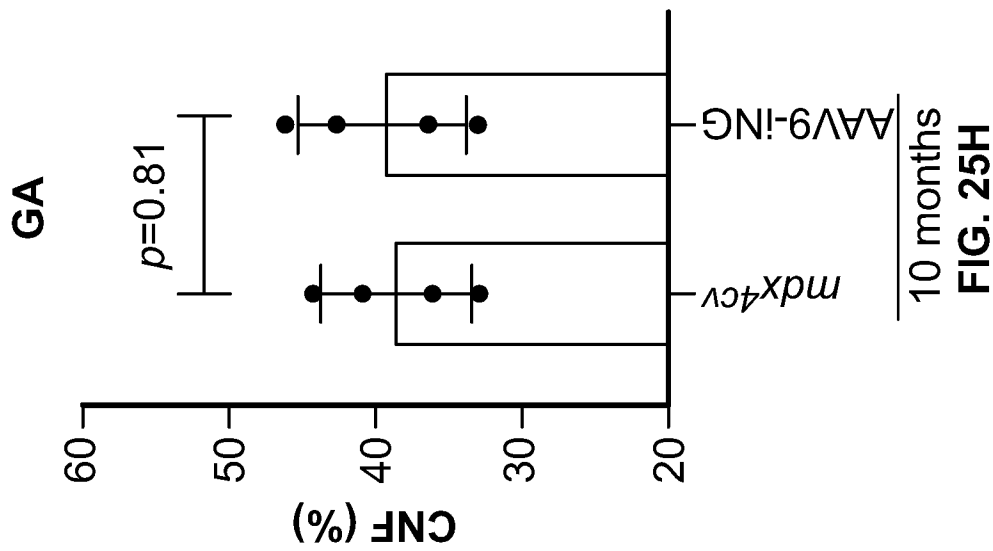


FIG. 25I



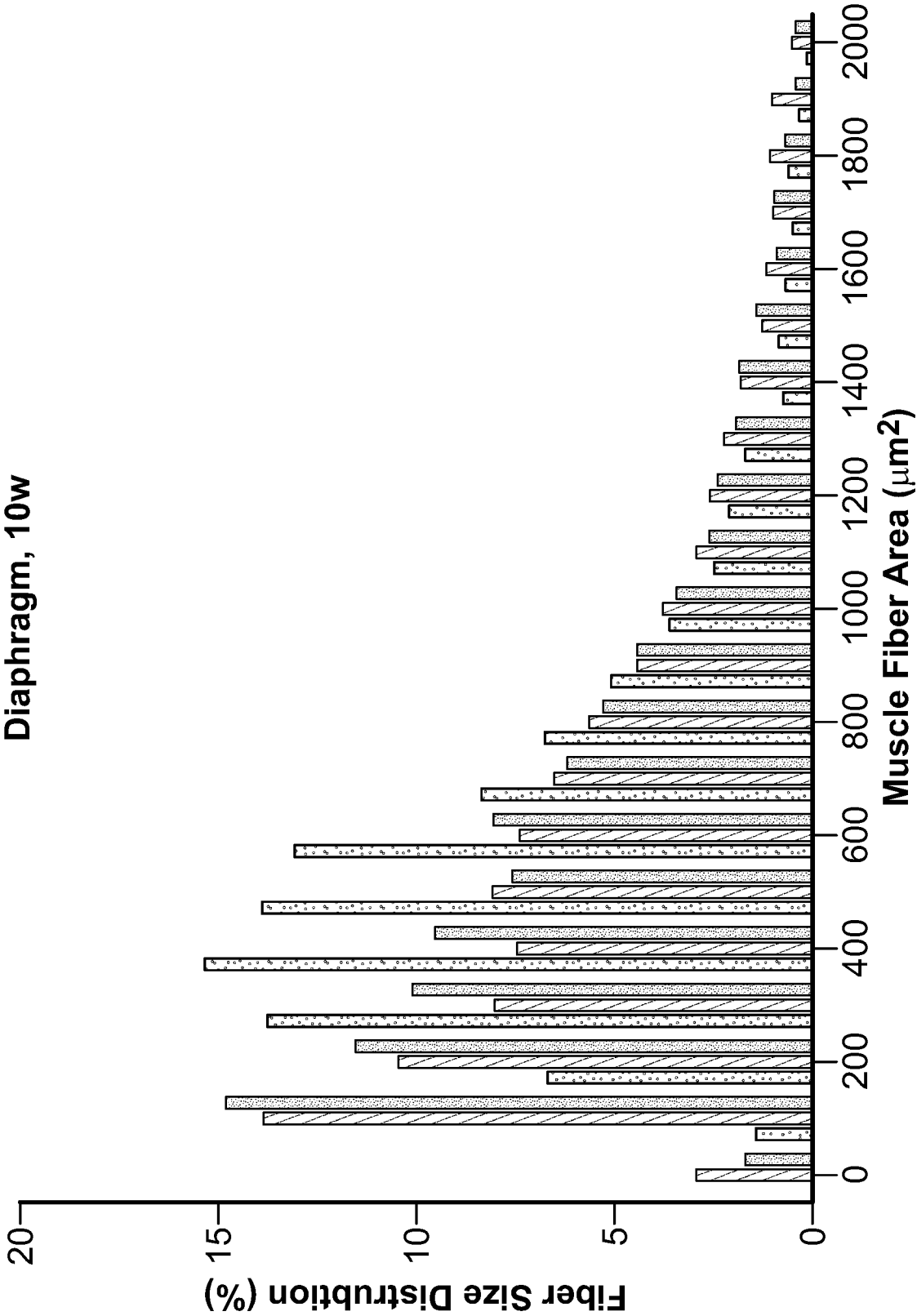
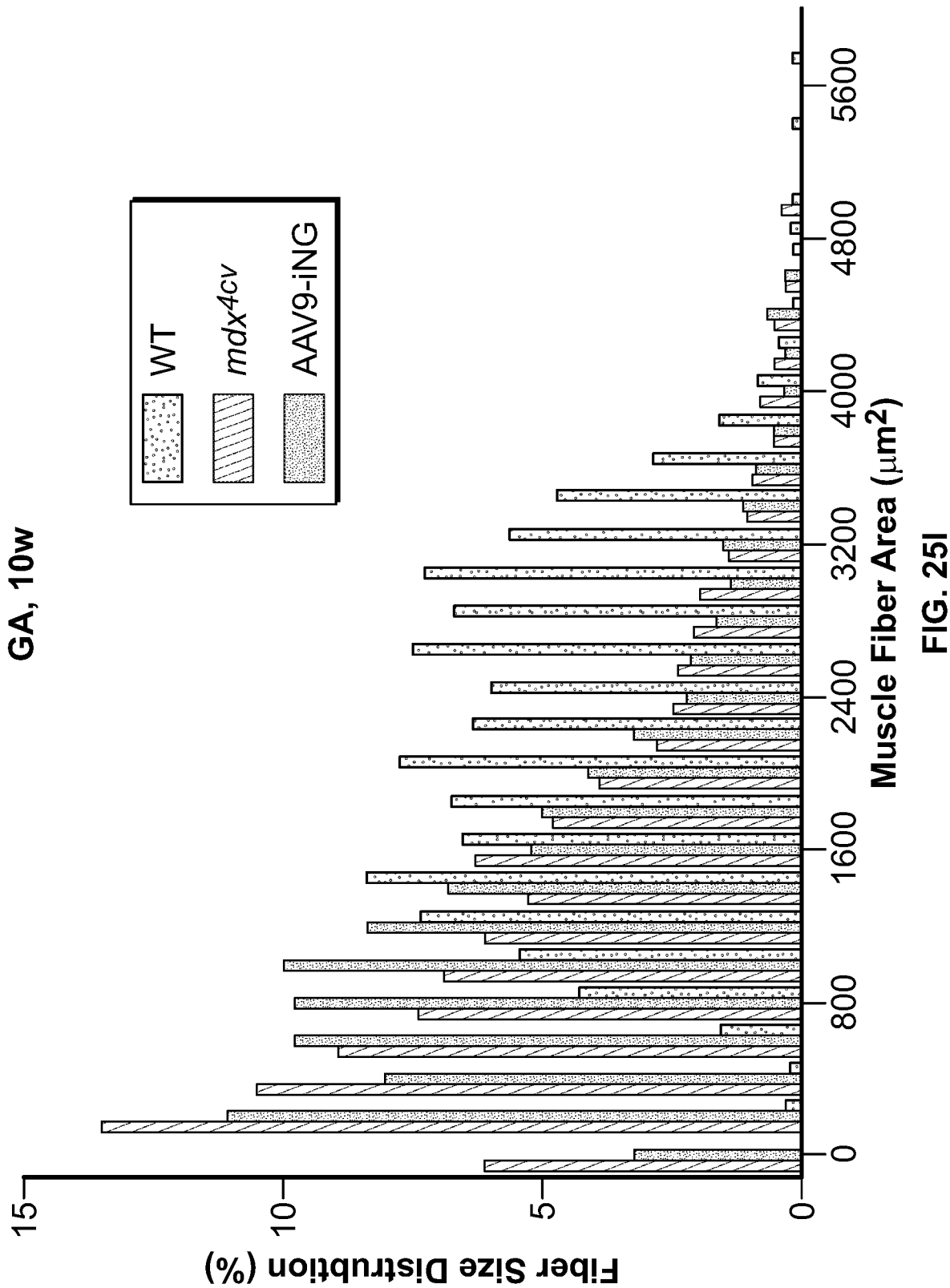


FIG. 25K



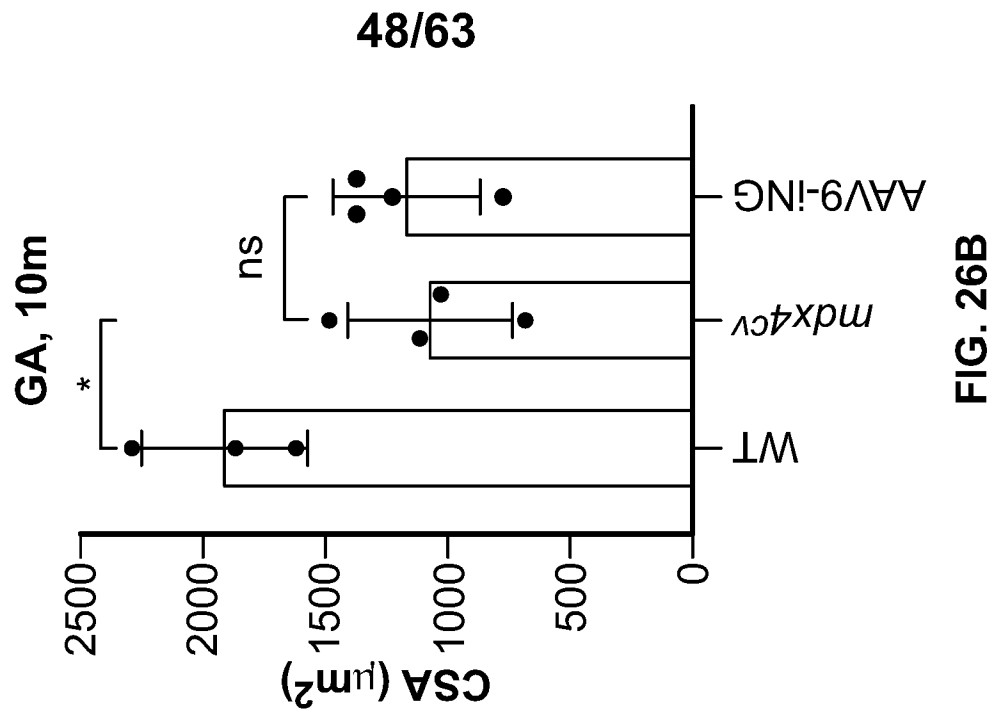


FIG. 26B

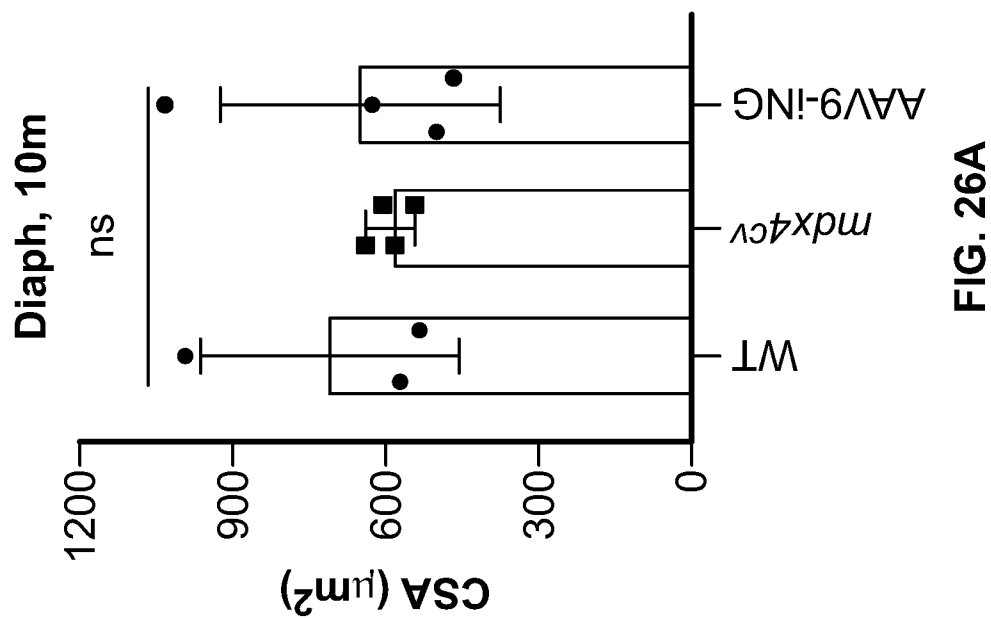


FIG. 26A

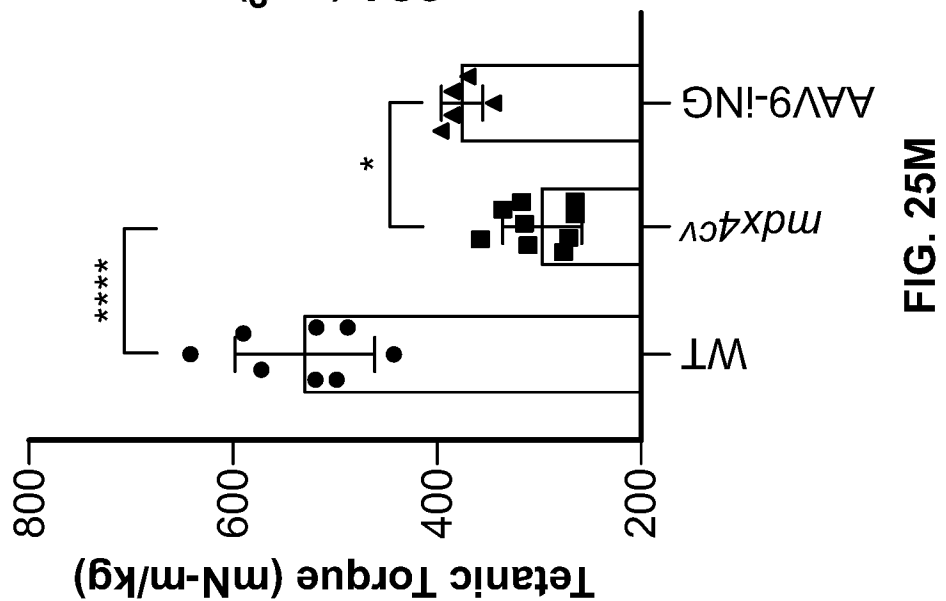


FIG. 25M

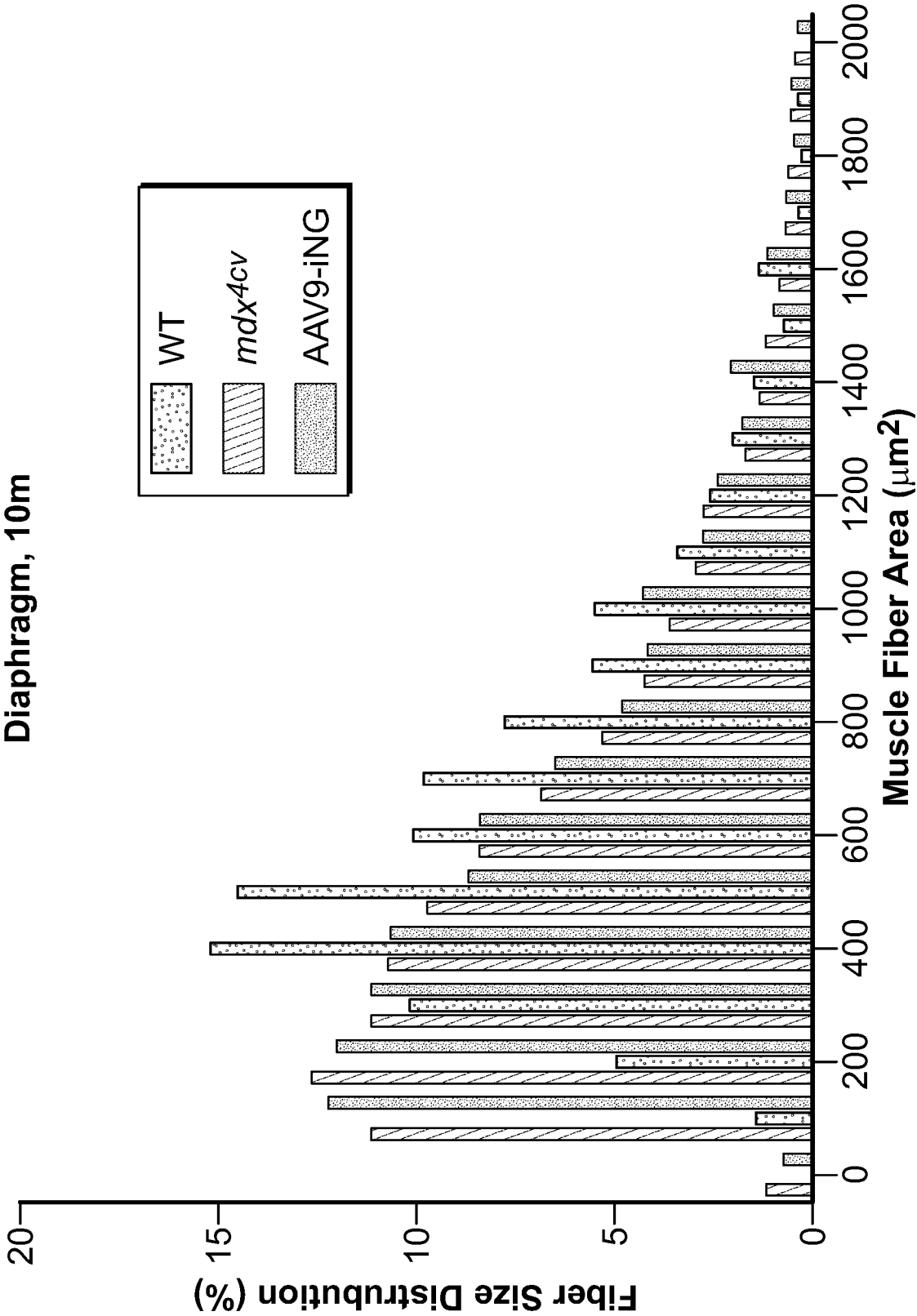
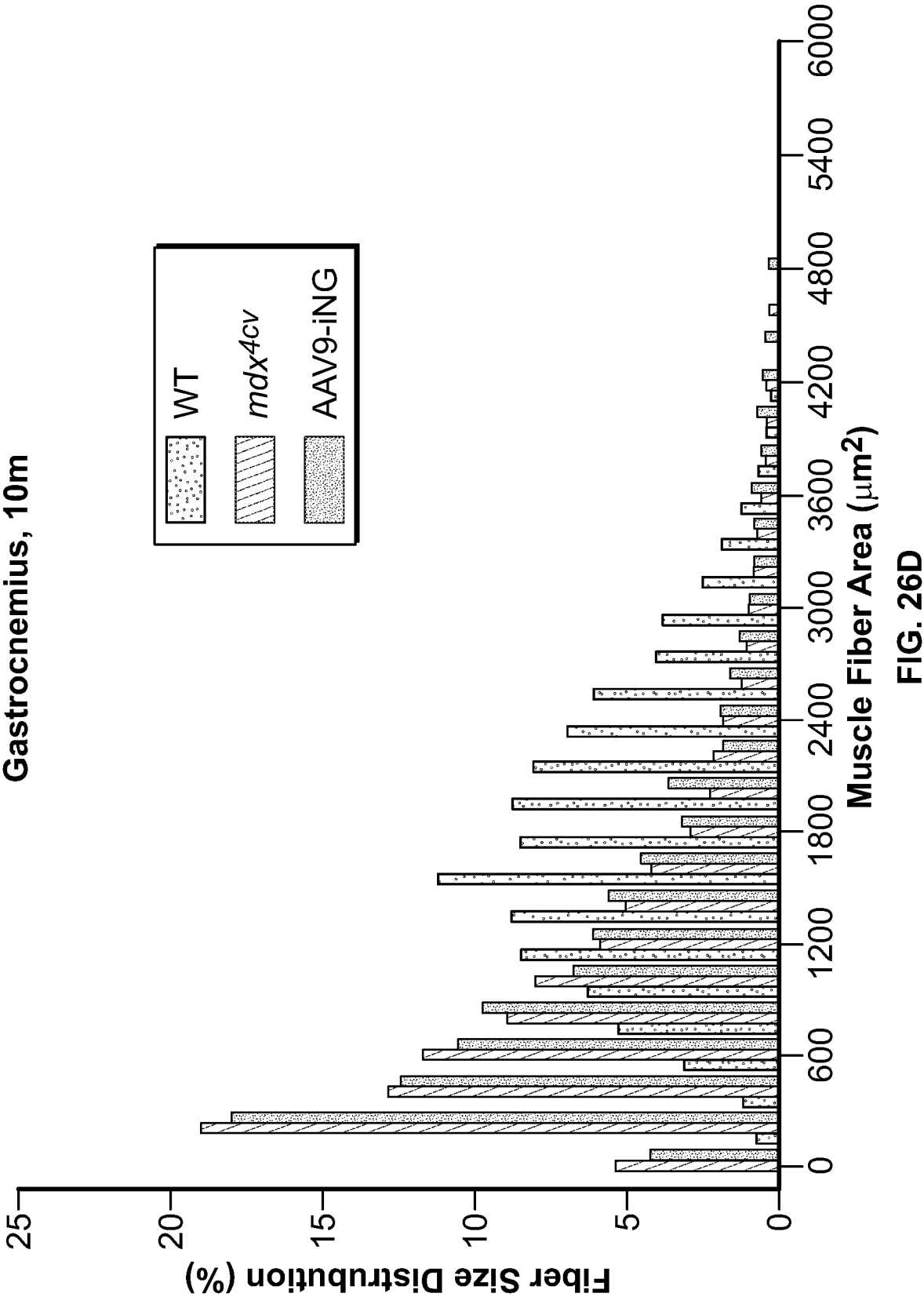
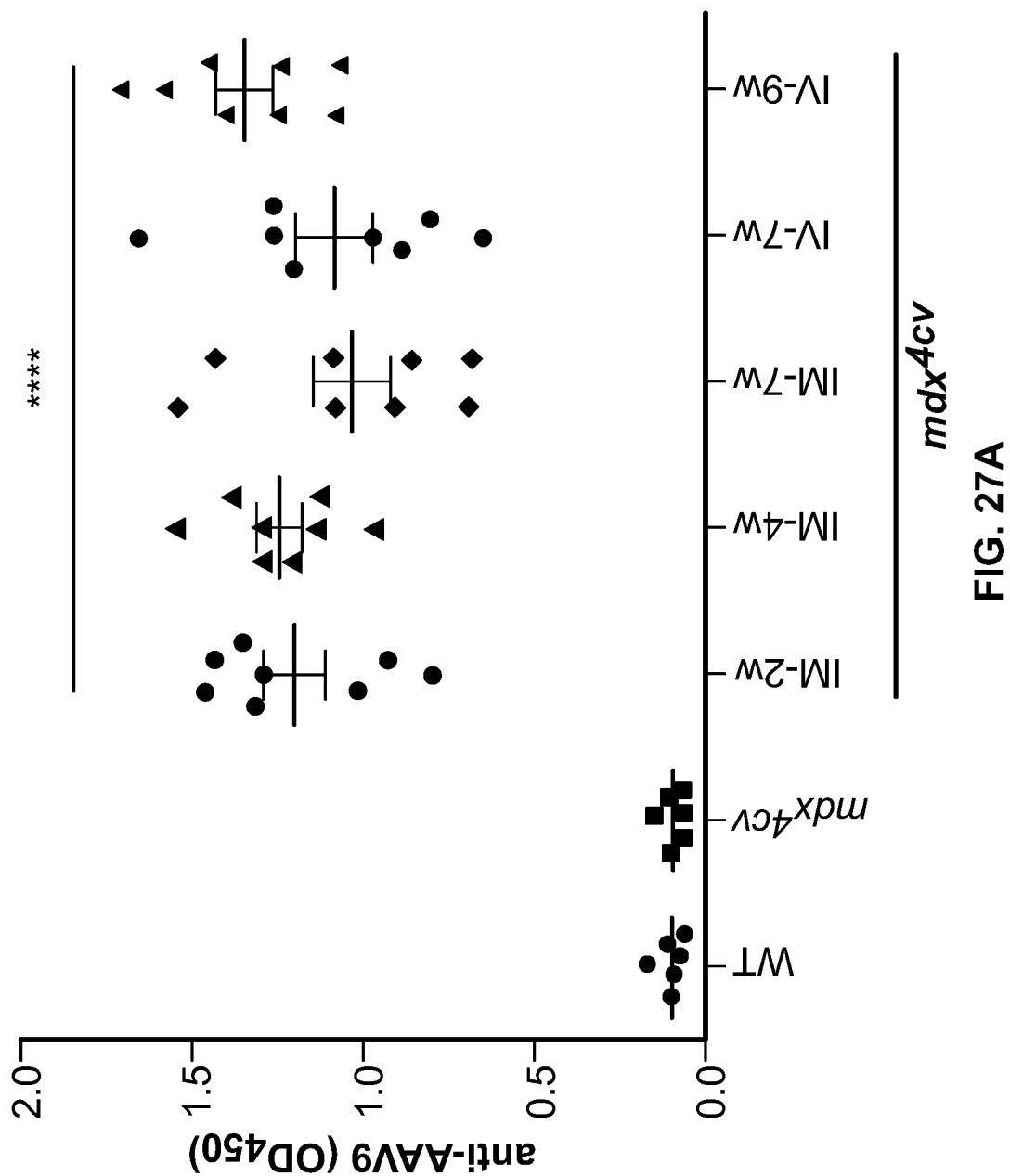


FIG. 26C



51/63



52/63

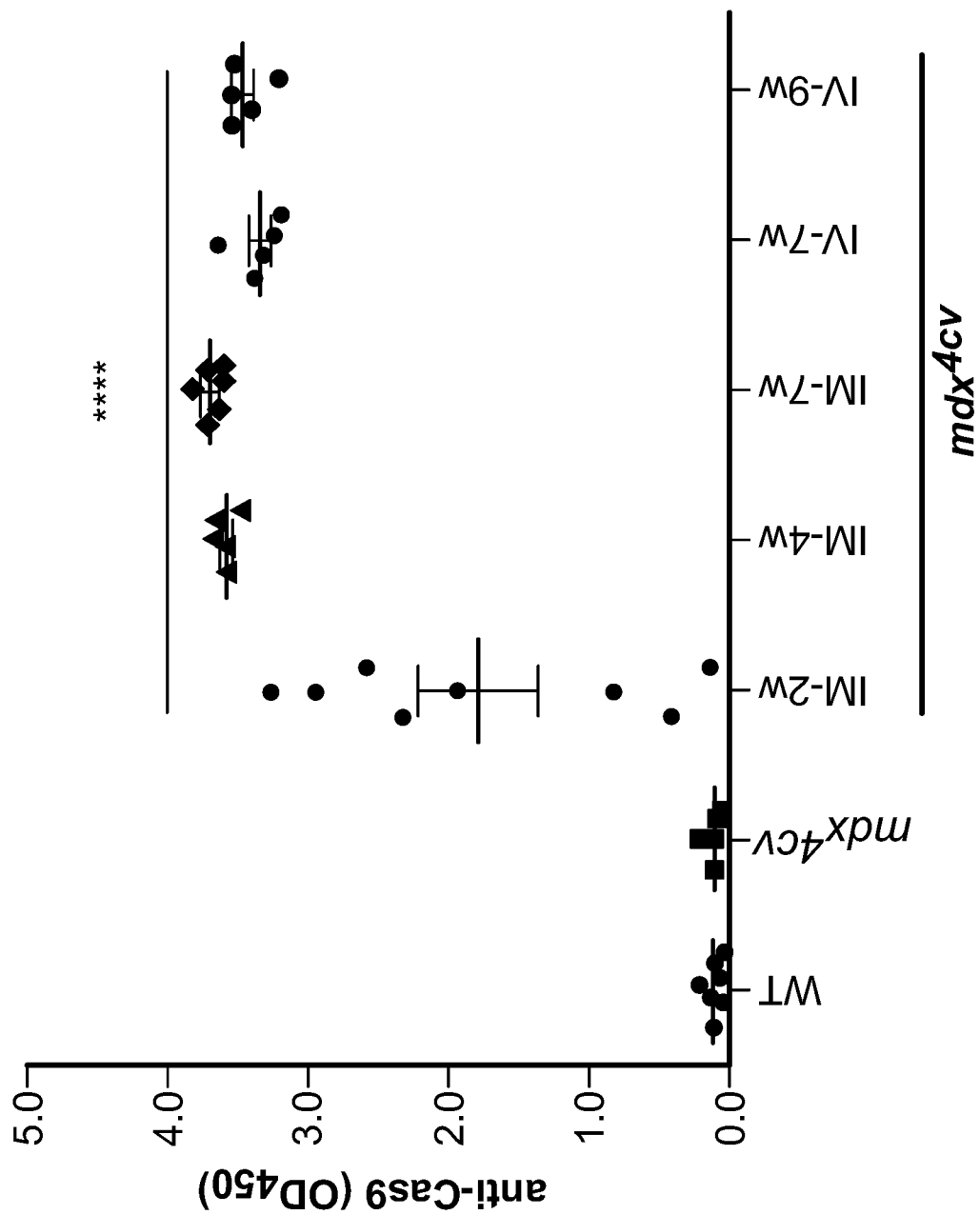
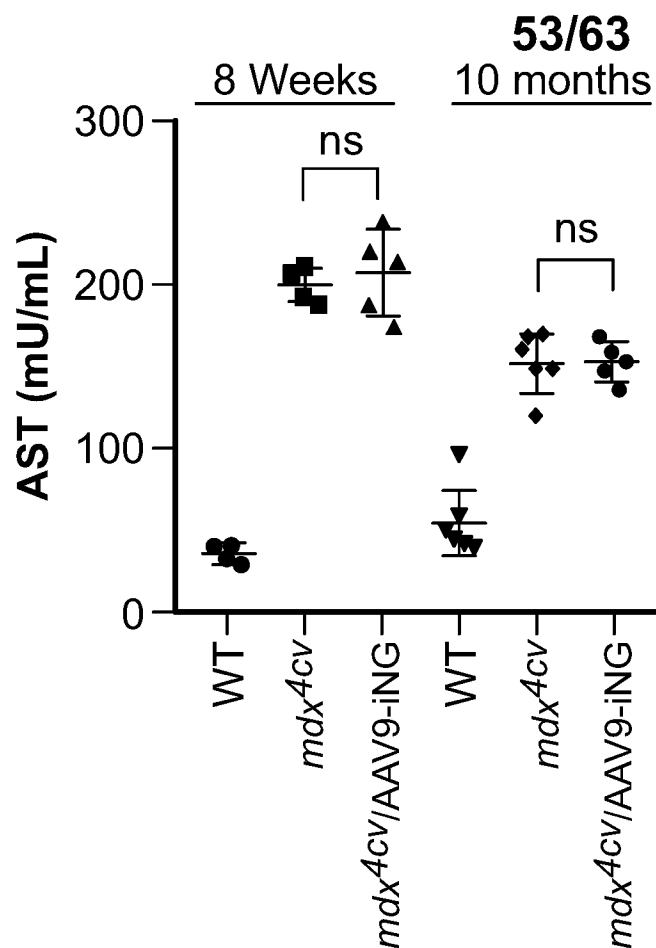
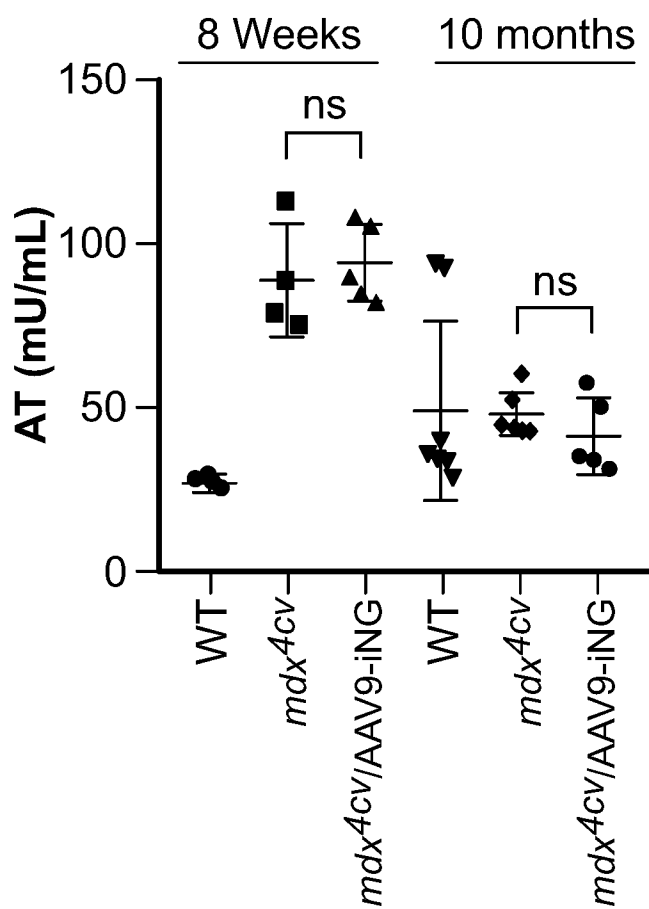


FIG. 27B

**FIG. 27C****FIG. 27D**

54/63

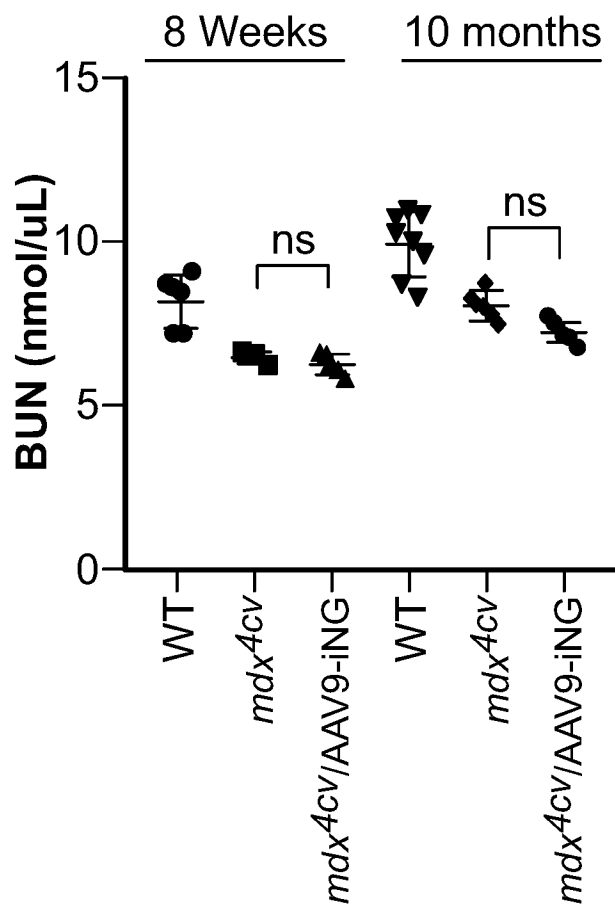


FIG. 27E

Mismatches	Number of Targets
1	1
2	2
3	55
4	562

ChrX: GTT**A**TCTCCTGTTCTGCAGC **TGT**
 Chr16: GTT**A**TCTCCTG**C**TCTGCAGC **AGA**
 Chr1: G**A**T**A**TCTCCTGTTCTGCAG**G** **AGA**

FIG. 27F

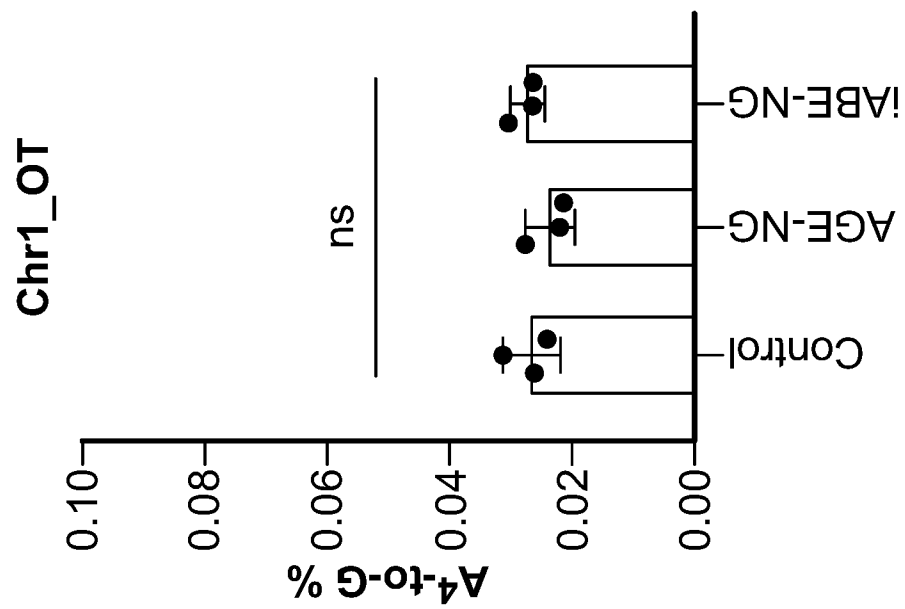


FIG. 27H

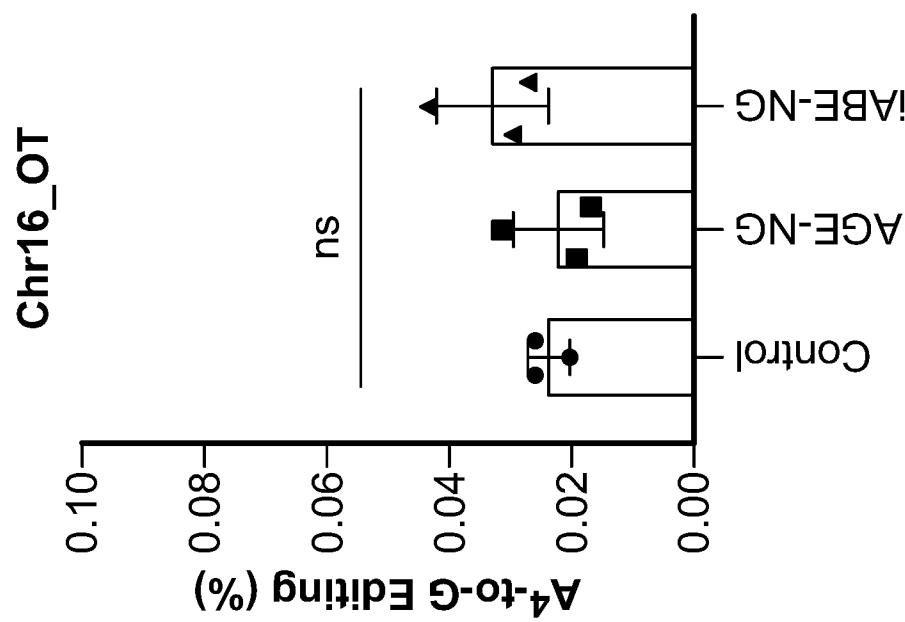
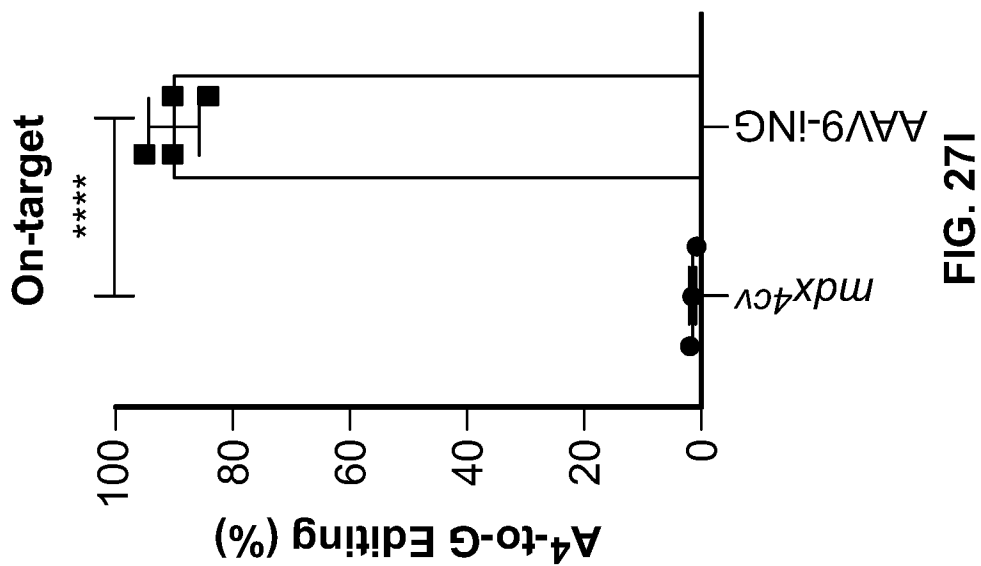
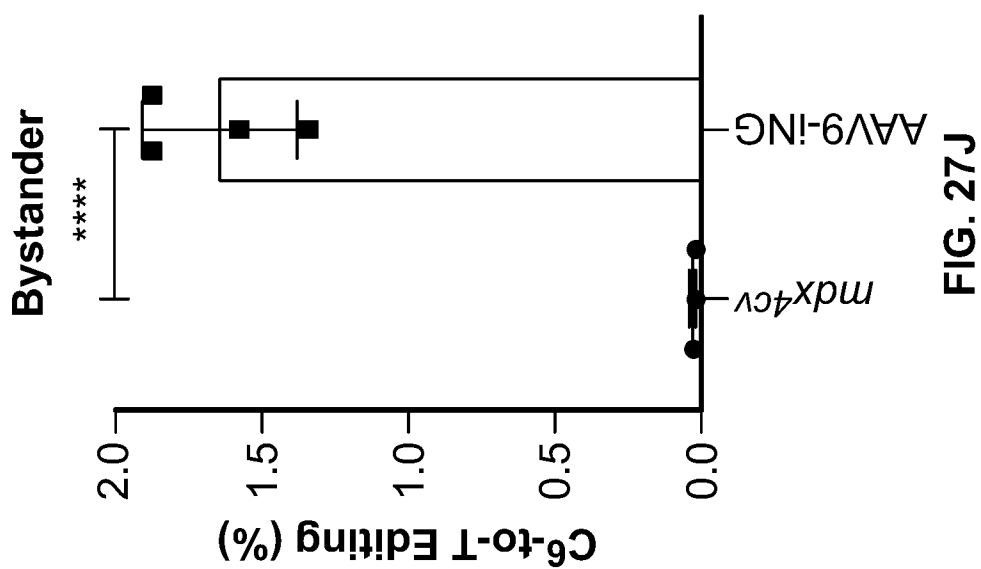


FIG. 27G

56/63



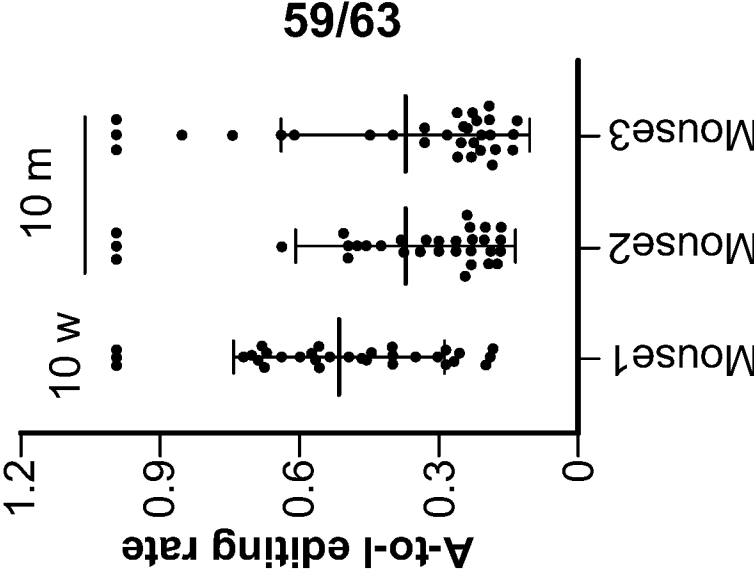
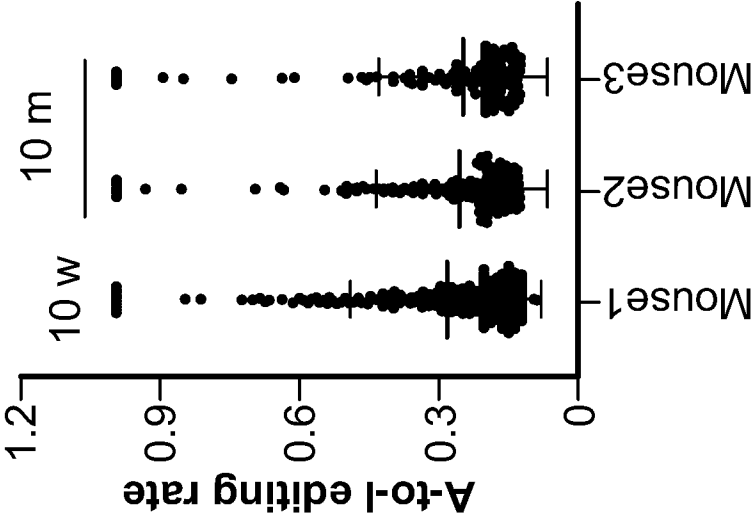
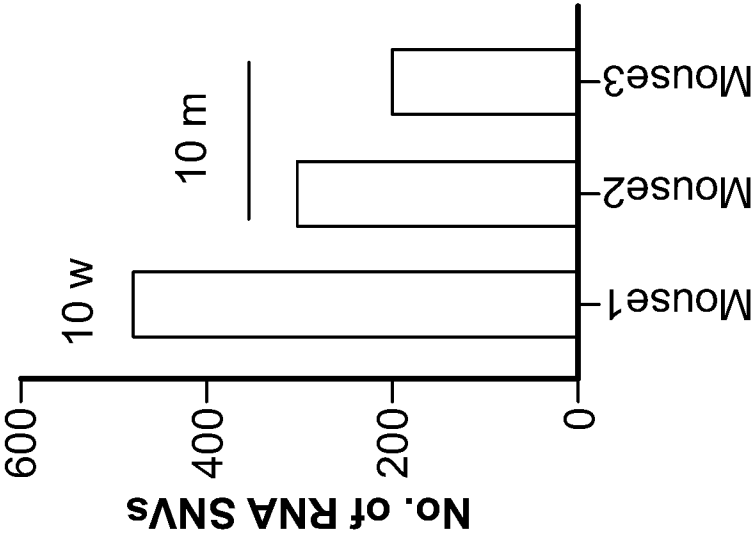
	G	T	T	T	A ⁴	T	C ⁶	T	C	C	T	G	T	T	C	T	G	C	A	G	C	T	G	T
Mouse1																								
A	0.05	0.01	0.00	0.00	13.64	0.01	0.03	0.03	0.05	0.02	0.02	0.02	0.01	0.01	0.00	0.03	0.08	0.01	99.90	0.02	0.01	0.01	0.03	0.01
C	0.00	0.03	0.07	0.01	0.01	0.03	97.93	0.04	99.86	99.96	0.09	0.00	0.02	0.03	99.96	0.11	0.00	99.96	0.00	0.00	99.97	0.07	0.00	0.02
G	99.95	0.00	0.00	0.00	86.34	0.00	0.19	0.00	0.02	0.00	0.01	99.97	0.00	0.00	0.00	0.00	99.92	0.00	0.08	99.97	0.00	0.05	99.96	0.00
T	0.00	99.96	99.92	0.02	0.02	99.95	1.84	99.93	0.06	0.02	99.89	0.00	99.97	99.97	0.03	99.86	0.00	0.03	0.01	0.02	0.02	99.87	0.01	99.97
Mouse2																								
A	0.07	0.01	0.00	0.00	14.04	0.02	0.04	0.03	0.04	0.03	0.02	0.04	0.01	0.01	0.01	0.03	0.07	0.01	99.90	0.04	0.01	0.01	0.03	0.01
C	0.00	0.03	0.06	0.00	0.00	0.02	98.27	0.04	99.90	99.92	0.09	0.00	0.02	0.03	99.97	0.14	0.00	99.94	0.01	0.00	99.97	0.09	0.00	0.02
G	99.92	0.00	0.00	0.00	85.93	0.00	0.21	0.00	0.00	0.03	0.01	99.96	0.00	0.00	0.00	0.01	99.93	0.00	0.08	99.94	0.00	0.01	99.96	0.00
T	0.00	99.95	99.94	0.02	0.02	99.96	1.48	99.92	0.05	0.02	99.88	0.00	99.97	99.96	0.02	99.82	0.00	0.04	0.02	0.02	0.02	99.89	0.01	99.97

FIG. 28

58/63

	G	T	T	A ⁴	T	C ⁶	T	C	C	T	G	T	C	T	G	C	A	G	C	T	G	T
Mouse3																						
A	0.08	0.01	0.01	7.94	0.02	0.04	0.03	0.02	0.04	0.02	0.03	0.02	0.01	0.01	0.04	0.07	0.00	99.88	0.02	0.01	0.01	0.01
C	0.00	0.06	0.06	0.01	0.05	97.74	0.07	99.97	99.94	0.09	0.00	0.02	0.02	0.03	0.15	0.00	99.95	0.01	0.00	99.96	0.09	0.02
G	99.91	0.00	0.00	92.03	0.00	0.36	0.01	0.00	0.01	0.01	99.97	0.00	0.00	0.00	0.01	99.92	0.00	0.10	99.97	0.00	0.01	99.96
T	0.00	99.93	99.93	0.02	99.93	1.86	99.89	0.01	0.02	99.89	0.00	99.96	99.95	0.02	99.81	0.00	0.05	0.01	0.01	0.02	99.89	0.01
Mouse4																						
A	0.10	0.01	0.00	19.63	0.02	0.03	0.04	0.03	0.04	0.02	0.05	0.01	0.01	0.01	0.03	0.08	0.01	99.87	0.03	0.01	0.01	0.01
C	0.00	0.04	0.06	0.01	0.02	98.51	0.06	99.93	99.94	0.10	0.00	0.03	0.04	99.97	0.17	0.00	99.93	0.01	0.00	99.96	0.12	0.02
G	99.89	0.00	0.00	80.32	0.01	0.19	0.00	0.00	0.00	0.01	99.95	0.00	0.00	0.00	0.01	99.91	0.00	0.10	99.95	0.00	0.02	0.00
T	0.00	99.95	99.94	0.04	99.96	1.28	99.89	0.04	0.02	99.87	0.00	99.96	99.95	0.02	99.79	0.01	0.06	0.02	0.02	0.03	99.84	0.01

FIG. 28
(Cont'd)



60/63

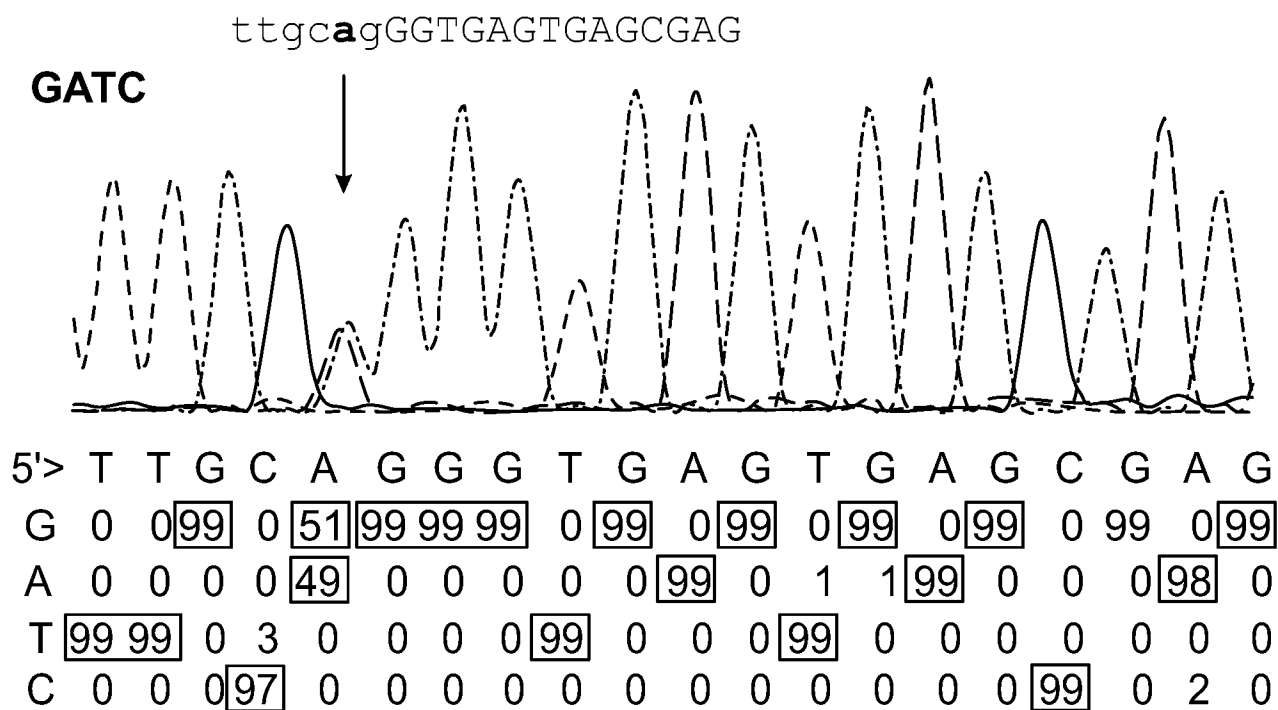


FIG. 30A

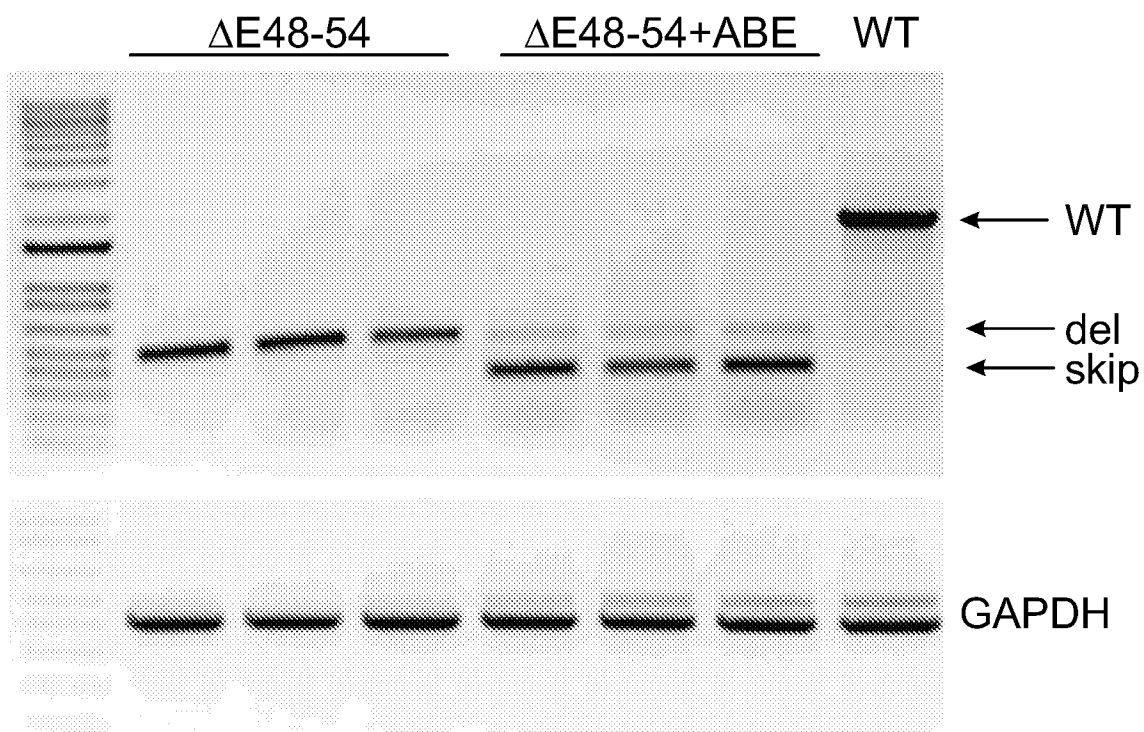


FIG. 30B

SMN1 Exon7: GGTTT⁶**C**AGACAAAAATCAAAAAGGAAGGTGCTC³⁶**A**CATTCCCTTAAATTAAAGGA
SMN2 Exon7: GGTTT**T**AGACAAAAATCAAAAAGGAAGGTGCTC**A**CATTCCCTTAAATTAAAGGA
gRNA1; GTGCTC**A**CATTCCCTTAAATTAA **AGG**
gRNA2; gGCTC**A**CATTCCCTTAAATTAA **GGA**

FIG. 31A

61/63

	ABEmax	iABE-NGA	ABEmax	iABE-NGA
Control				
	gRNA1	gRNA1	gRNA2	gRNA2

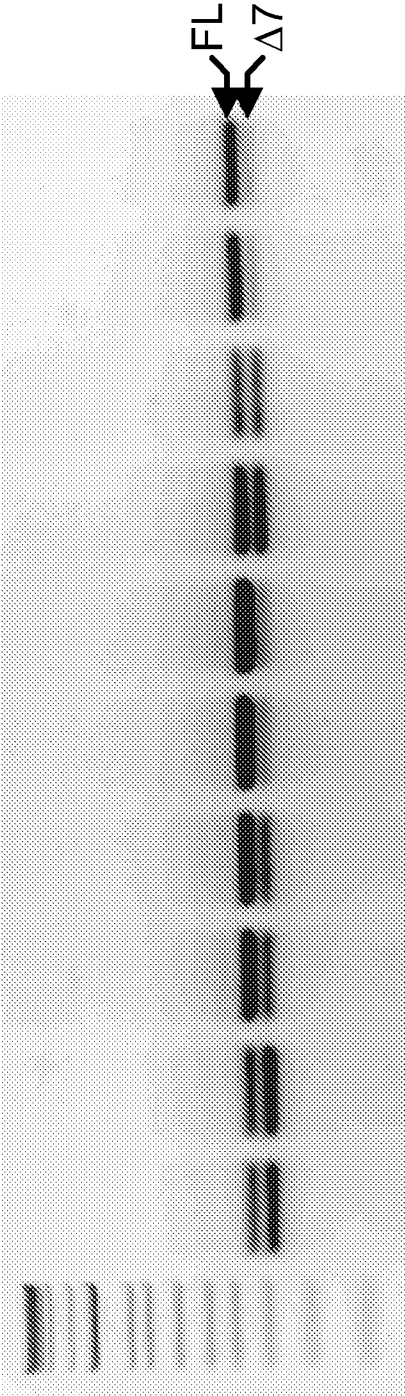


FIG. 31B

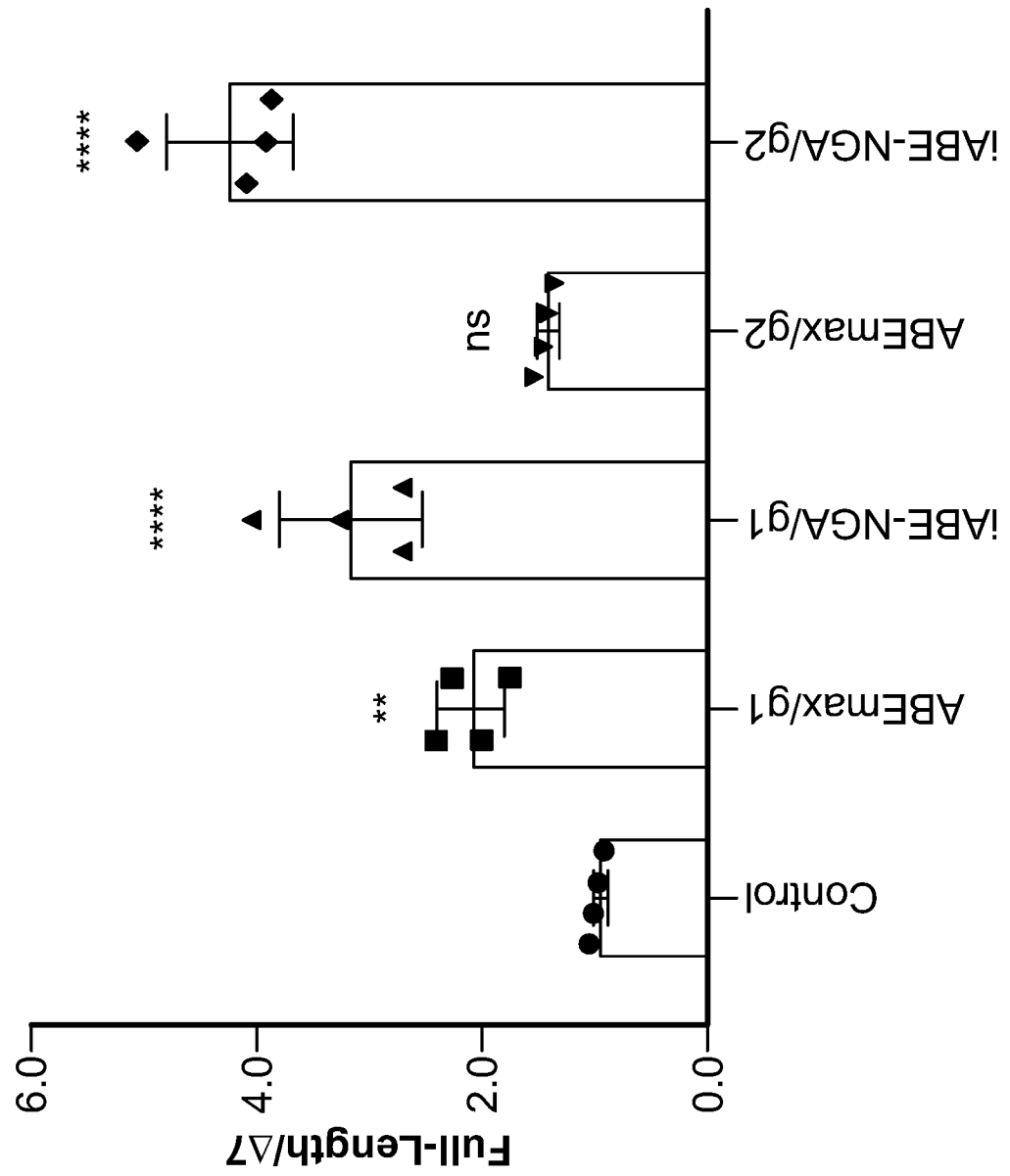
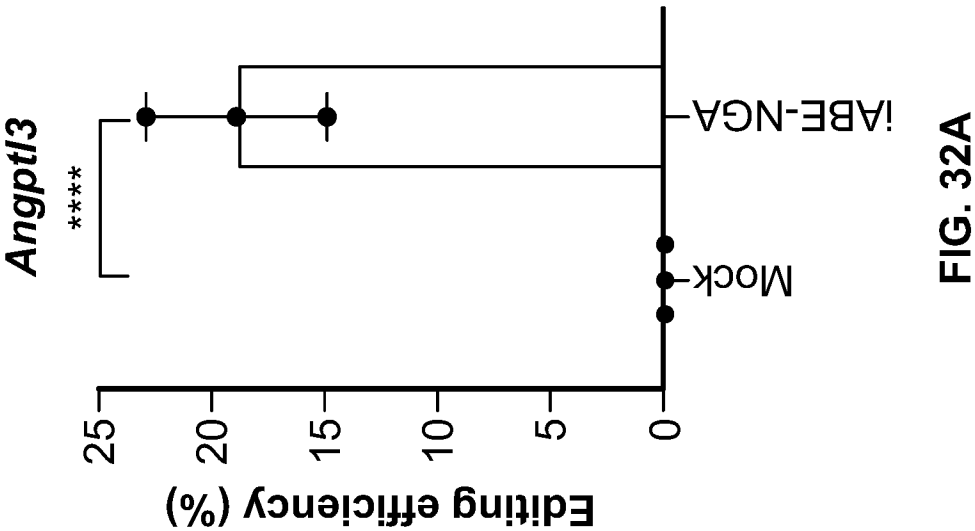
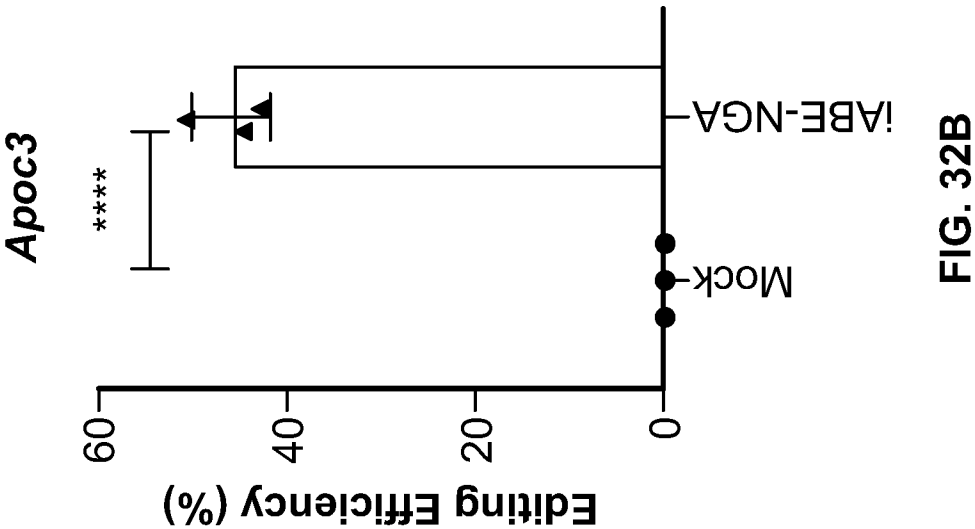
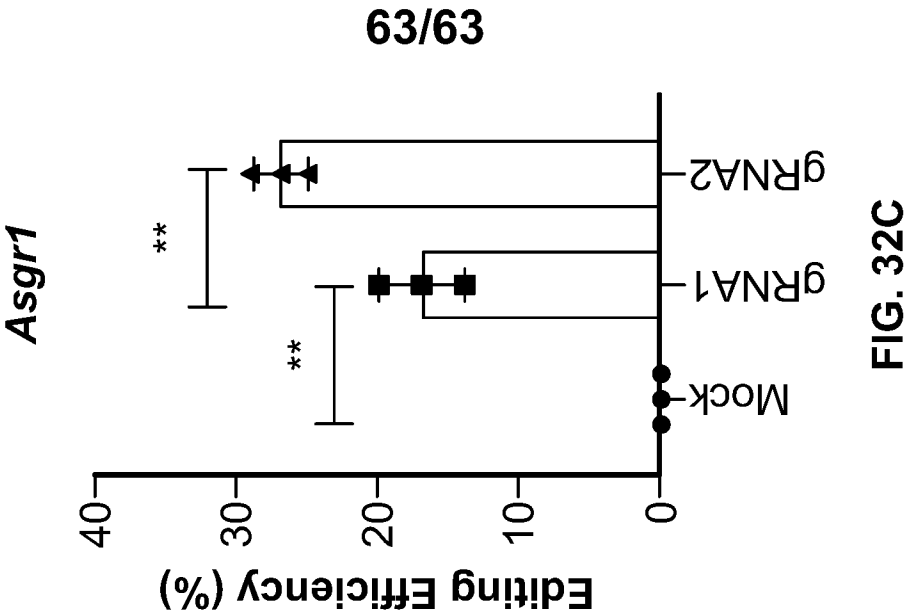


FIG. 31C



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/017868

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 9/22; C12N 9/78; C12N 15/86 (2021.01)

CPC - C07K 2319/80; C07K 2319/92; C12N 9/22; C12N 9/78; C12N 15/8645 (2021.02)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2019/217943 A1 (BEAM THERAPEUTICS INC.) 14 November 2019 (14.11.2019) entire document	1-3 ----- 11-14, 24-27
X -- Y	✓ CN 110029096 A (SHANGHAITECH UNIVERSITY) 19 July 2019 (19.07.2019) entire document; see machine translation	1, 3, 4 ----- 14, 27
Y	US 2018/0127780 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 10 May 2018 (10.05.2018) entire document	11-14, 24-27
A	WO 2019/139645 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 18 July 2019 (18.07.2019) entire document	1-4, 11-14, 24-27
A	✓ CN 110526993 A (SHANDONG SHUNFENG BIOTECHNOLOGY CO LTD) 03 December 2019 (03.12.2019) entire document	1-4, 11-14, 24-27



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

28 April 2021

Date of mailing of the international search report

MAY 07 2021

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/017868

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 5-10, 15-23, 28-42
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/017868

Box No. I **Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

- a. ☒ forming part of the international application as filed:
 ☒ in the form of an Annex C/ST.25 text file.
 ☐ on paper or in the form of an image file.
- b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
- c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 ☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 ☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs: 11-27 and 47 were searched.