

Miniaturized Dystrophins Having Spectrin Fusion Domains and Uses Thereof

5 CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims benefit of U.S. Patent Application Serial No. 63/017148, filed April 29, 2020, which is herein incorporated by reference in its entirety.

FIELD

10 The presently disclosed subject matter generally relates to polynucleotides, polypeptides, cells, vectors, uses, and kits relating to miniaturized dystrophin.

BACKGROUND OF THE DISCLOSURE

Duchenne muscular dystrophy (DMD) is a recessively-inherited muscle wasting disorder afflicting approximately 1 in 3,500 males. DMD is caused by mutations in the dystrophin gene, which is located on the X chromosome. Mutations in this gene lead to aberrant or absent expression of the dystrophin protein.

15 Dystrophin is a key component of a protein complex that is responsible for regulating muscle cell integrity and function. DMD patients typically lose the ability to physically support themselves during childhood and become progressively weaker over time. This progressive wasting of skeletal muscles and cardiac dysfunction typically leads to loss of ambulation and premature death, primarily due to cardiac or respiratory failure.

20 Some attempts have been made in the past to treat DMD. However, the available treatment options were significantly limited due to the large size of the wild type dystrophin cDNA (approximately 13.9 kb) which cannot be administered to and expressed in DMD patients using standard viral vectors, including Adeno-associated virus (AAV), which cannot transfer more than approximately 4.9 kb of heterologous DNA. Therefore, there is a need to develop a recombinant dystrophin gene that can be efficiently packaged into a vector for gene therapy.

25 Adeno-associated viral (AAV) vectors have been shown to be useful in gene therapeutic approaches aimed at correcting genetic deficiencies that result in reduced or completely abolished levels of protein expression (Nathwani *et al.*, *Human Gene Therapy*

28:1004-1012 (2017); Keeler A.M. *et al.*, *Clin. Transl. Sci.* 10:242-248 (2017)), and are potentially useful for gene knockdown, genome editing or modification, and non-coding RNA modulation (Valdmanis *et al.*, *Human Gene Therapy* 28(4):361-372 (2017 April)).

5 Packaging the entire cDNA of the muscle-specific isoform of dystrophin into a single rAAV capsid cannot be achieved easily because of the large size of the dystrophin cDNA. Previous studies have focused on the development of smaller genetic constructs that express only particular domains of dystrophin. *See* U.S. Patent Nos. 6,869,777 and 8,501,920, each of which is incorporated by reference. However, these approaches have had only limited success.

10 There remains a need for more precise and efficient gene therapy tools for treating patients with mutations in the dystrophin gene, and, in particular, a need to develop a recombinant dystrophin gene that can be efficiently packaged into a vector for gene therapy.

SUMMARY OF THE DISCLOSURE

15 The present disclosure provides a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising a modified spectrin repeat 16 (R16) domain, wherein a part of spectrin repeat 16 (R16) domain is replaced by a corresponding part of a different spectrin repeat domain. In some embodiments, the different spectrin repeat domain is spectrin repeat 2 (R2) domain. In
20 some embodiments the modified R16 domain comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to a sequence selected from the group consisting of SEQ ID NO: 68, 69, 70 and 71. In some
25 embodiments, the miniaturized dystrophin polypeptide comprises from N terminus to C terminus a hinge 1 (H1) domain, a spectrin repeat 1 (R1) domain, the modified R16 domain, a spectrin repeat 17 (R17) domain, a hinge 3 (H3) domain, a spectrin repeat 23 (R23) domain, a spectrin repeat 24 (R24) domain, and a hinge 4 (H4) domain of dystrophin. In
30 some embodiments, (i) the H1 domain and the R1 domain are fused directly, (ii) the R1 domain and the modified R16 domain are fused directly, (iii) the modified R16 domain and the R17 domain are fused directly, (iv) the R17 domain and the H3 domain are fused

directly, (v) the H3 domain and the R23 domain are fused directly, (vi) the R23 domain and the R24 domain are fused directly, or (vii) the R24 domain and the H4 domain are fused directly, or (vii) any combination thereof. In some embodiments, the miniaturized dystrophin polypeptide does not comprise a spectrin repeat 2 (R2) domain, spectrin repeat 3 (R3) domain, spectrin repeat 4 (R4) domain, spectrin repeat 5 (R5) domain, spectrin repeat 6 (R6) domain, spectrin repeat 7 (R7) domain, spectrin repeat 8 (R8) domain, spectrin repeat 9 (R9) domain, spectrin repeat 10 (R10) domain, spectrin repeat 11 (R11) domain, spectrin repeat 12 (R12) domain, spectrin repeat 13 (R13) domain, spectrin repeat 14 (R14) domain, spectrin repeat 15 (R15) domain, spectrin repeat 18 (R18) domain, spectrin repeat 19 (R19) domain, spectrin repeat 20 (R20) domain, spectrin repeat 21 (R21) domain, and/or spectrin repeat 22 (R22) domain. In some embodiments, the miniaturized dystrophin polypeptide further comprises an ABD1 domain and/or a CR domain. In some embodiments, the miniaturized dystrophin polypeptide consists essentially of or consists of, from N terminus to C terminus, the ABD1 domain, the H1 domain, the R1 domain, the modified R16 domain, the R17 domain, the H3 domain, the R23 domain, the R24 domain, the H4 domain, and the CR domain of dystrophin. In some embodiments, the H1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 74. In some embodiments, the R1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 75. In some embodiments, the modified R16 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 76. In some embodiments, the R17 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 77. In some embodiments, the H3 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 78. In some embodiments, the R23

domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 79. In some embodiments, the R24 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 80. In some embodiments, the H4 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 81. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the N terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 73. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the C terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 82. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 83. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 84. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 85. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 86. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about

85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 87. In some embodiments, the miniaturized dystrophin polypeptide exhibits a higher expression of the miniaturized dystrophin polypeptide than BXA-212372 (SEQ ID NO: 88). In some
5 embodiments, the miniaturized dystrophin polypeptide expression is at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about 1.9 fold, at least about 2 fold, at least about 2.1 fold, at least about 2.2 fold, at least about 2.3 fold, at least about 2.4 fold, at least about 2.5 fold, at least about 2.6 fold, at least about 2.7 fold, at least about 2.8 fold, at least about 2.9 fold or at least about 3 fold higher than the BXA-
10 212372 polypeptide (SEQ ID NO: 88) expression.

In some embodiments, the nucleic acid molecule disclosed herein further comprises a promoter. In some embodiments, the promoter is a tissue-specific promoter. In some embodiments, the promoter drives expression of the therapeutic protein in muscle cells, hepatocytes, endothelial cells, neuronal cells, sinusoidal cells, or any combination thereof.
15 In some embodiments, the promoter is selected from the group consisting of a C5-12(T) promoter, an MLC2v-cTNT455 promoter, a mouse thyretin promoter (mTTR), an endogenous human factor VIII promoter (F8), a human alpha-1-antitrypsin promoter (hAAT), a human albumin minimal promoter, a mouse albumin promoter, a tristetraprolin (TTP) promoter, a CASI promoter, a synapsin 1 gene promoter, a CAG promoter, a
20 cytomegalovirus (CMV) promoter, α 1-antitrypsin (AAT), muscle creatine kinase (MCK), myosin heavy chain alpha (α MHC), myoglobin (MB), desmin (DES), SPc5-12, 2R5Sc5-12, dMCK, tMCK, and a phosphoglycerate kinase (PGK) promoter. In some embodiments, the promoter is a C5-12(T) promoter. In some embodiments, the nucleic acid molecule disclosed herein further comprises an intronic sequence. In some embodiments, the intronic
25 sequence is positioned 5' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the intronic sequence is positioned 3' to the promoter. In some embodiments, the intronic sequence comprises a synthetic intronic sequence. In some embodiments, the nucleic acid molecule disclosed herein further comprises a post-transcriptional regulatory element. In some embodiments, the post-transcriptional
30 regulatory element is positioned 3' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the post-transcriptional regulatory element

comprises a mutated woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), a microRNA binding site, or a DNA nuclear targeting sequence, or any combination thereof. In some embodiments, the nucleic acid molecule disclosed further comprises a 3'UTR poly(A) tail sequence. In some embodiments, the 3'UTR poly(A) tail sequence is selected from the group consisting of dystrophin poly(A), bGH poly(A), actin poly(A), hemoglobin poly(A), and any combination thereof. In some embodiments, the 3'UTR poly(A) tail sequence comprises dystrophin poly(A). In some embodiments, the nucleic acid molecule disclosed further comprises an enhancer sequence. In some embodiments, the nucleic acid molecule disclosed herein further comprises a first ITR and/or a second ITR. In some embodiments, the first ITR and the second ITR are identical. In some embodiments, the first ITR and/or the second ITR are derived from adeno-associated virus. In some embodiments, the nucleic acid molecule disclosed herein comprises a sequence encoding a heterologous moiety. In some embodiments, the heterologous moiety is selected from the group consisting of albumin or a fragment thereof, an immunoglobulin Fc region, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, a PAS sequence, a HAP sequence, a transferrin or a fragment thereof, an albumin-binding moiety or a derivative thereof, and any combination thereof.

In some embodiments, provided is a vector comprising a nucleic acid molecule disclosed herein. In some embodiments, the vector is selected from the group consisting of a adenoviral vector, a retroviral vector, poxvirus vector, a baculovirus vector, a herpes viral vector. In some embodiments, the vector is an adeno-associated virus (AAV) vector. In some embodiments, the AAV vector is selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and AAV11. In some embodiments, the AAV vector is AAV8 or AAV9. In some embodiments, the AAV vector is AAV9. In some embodiments, the AAV vector is AAV8.

In some embodiments, the nucleic acid molecule or vector disclosed herein is formulated with a delivery agent. In some embodiments, the delivery agent comprises a lipid nanoparticle. In some embodiments, the delivery agent is selected from the group consisting of liposomes, non-lipid polymeric molecules, endosomes, and any combination thereof. In some embodiments, the nucleic acid molecule or vector disclosed herein is formulated for intravenous, transdermal, intradermal, subcutaneous, pulmonary, or oral

delivery, or any combination thereof. In some embodiments, the nucleic acid molecule or vector disclosed herein is formulated for intravenous delivery.

In some embodiments, provided is a polypeptide encoded by the nucleic acid molecule or vector disclosed herein.

5 In some embodiments, provided is a host cell comprising the nucleic acid molecule or vector disclosed herein. In some embodiments, the cell is a CHO cell, a HEK293 cell, a HBK cell, a COS cell, a NSO cell, or a HT1080 cell.

In some embodiments, provided is a pharmaceutical composition comprising (a) the nucleic, the vector, the polypeptide, or the host cell disclosed herein; and (b) a
10 pharmaceutically acceptable excipient.

In some embodiments, provided is a kit, comprising the nucleic, the vector, the polypeptide, the host cell, or the pharmaceutical composition disclosed herein, and instructions for administering the nucleic, the vector, the polypeptide, the host cell, or the pharmaceutical composition to a subject in need thereof.

15 In some embodiments, provided is a method of producing a miniaturized dystrophin polypeptide, comprising: culturing the host cell disclosed herein under suitable conditions and recovering the miniaturized dystrophin polypeptide.

In some embodiments, provided is a method of expressing a miniaturized dystrophin polypeptide in a subject in need thereof, comprising administering to the subject
20 the nucleic acid, the vector, the host cell, or the pharmaceutical composition disclosed herein.

In some embodiments, provided is a method of treating a subject having a disease or condition comprising administering to the subject the nucleic acid, the vector, the polypeptide, the host cell, or the pharmaceutical composition disclosed herein. In some
25 embodiments, the disease or condition is a disease caused by dystrophin deficiency. In some embodiments, the disease is Sarcopenia, a heart disease, cachexia, Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), X-linked dilated cardiomyopathy (XLDC), facioscapulohumeral muscular dystrophy, myotonic muscular dystrophy, limb-girdle muscular dystrophy, oculopharyngeal muscular dystrophy, Emery-
30 Dreifuss muscular dystrophy, distal muscular dystrophy, and/or congenital muscular dystrophy. In some embodiments, the nucleic acid molecule, the vector, the polypeptide,

the host cell, or the pharmaceutical composition is administered intravenously, transdermally, intradermally, subcutaneously, orally, or pulmonarily, or any combination thereof. In some embodiments, the method disclosed herein further comprises administering to the subject a second agent. In some embodiments, the subject is a human.

5 In some embodiments, the administration of the nucleic acid molecule, the vector, the polypeptide, the host cell, or the pharmaceutical composition to the subject results in increased dystrophin protein expression, relative to dystrophin protein expression in the subject prior to the administration, wherein the dystrophin protein expression is increased by at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at
10 least about 6-fold, at least about 7-fold, at least about 8-fold, at least about 9-fold, at least about 10-fold, at least about 11-fold, at least about 12-fold, at least about 13-fold, at least about 14-fold, at least about 15-fold, at least about 20-fold, at least about 25-fold, at least about 30-fold, at least about 35-fold, at least about 40-fold, at least about 50-fold, at least about 60-fold, at least about 70-fold, at least about 80-fold, at least about 90-fold, or at least
15 about 100-fold. In some embodiments, provided is a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising an amino acid sequence of SEQ ID NO: 83. In some embodiments, provided is a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide consisting of the amino acid sequence of SEQ ID NO: 83.

20 In some embodiments, the nucleic acid molecule, the vector, the polypeptide, the host cell, the pharmaceutical composition, the kit, or the method disclosed herein, comprises a nucleotide sequence encoding a miniaturized dystrophin polypeptide comprising an amino acid sequence of SEQ ID NO: 83.

In some embodiments, the nucleic acid molecule encodes miniaturized dystrophin
25 polypeptide BXA-220931 (SEQ ID NO: 83).

In some embodiments, provided is a nucleic acid molecule comprising a nucleotide sequence comprising in order a C5-12(T) promoter of SEQ ID NO: 109, an SV40 intron of SEQ ID NO: 110, a coding sequence for miniaturized dystrophin BXA-220931 of SEQ ID NO: 111, a 3' UTR of SEQ ID NO: 112, and a polyA sequence of SEQ ID NO: 113.

30 In some embodiments, the nucleic acid molecule, the vector, the polypeptide, the host cell, the pharmaceutical composition, the kit, or the method disclosed herein,

comprises a nucleotide sequence comprising in order a C5-12(T) promoter of SEQ ID NO: 109, an SV40 intron of SEQ ID NO: 110, a coding sequence for miniaturized dystrophin BXA-220931 of SEQ ID NO: 111, a 3' UTR of SEQ ID NO: 112, and a polyA sequence of SEQ ID NO: 113.

5

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a schematic diagram of the full length human Dystrophin protein. ABD1: actin-binding domain-1; H# (e.g., H1): hinge region; R# (e.g., R1): spectrin-like repeat domains; ABD2: actin-binding domain-2; CR: cysteine-rich domain; C-term: C-terminal domain of the protein.

10

FIG. 2 shows schematic diagrams of miniaturized dystrophin polypeptides BXA-212372, BXA-212372-J4, BXA-212372-J4V4, BXA-212372-J4V11, BXA-212372-J4V12, and BXA-212372-J4V13 (BXA-220931).

15

FIG. 3 shows miniaturized dystrophin polypeptide expression in human isogenic induced-pluripotent stem cell (iPSC)-derived cardiac myocytes (iCMs) (carrying an E2035X premature stop codon in the dystrophin gene that prevented endogenous dystrophin expression) after transfection of plasmids expressing the indicated miniaturized dystrophin polypeptides. Polypeptide expression was quantitated by ELISA. Significance: ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (one-way ANOVA with post-hoc Tukey test). Bar graphs reflect the means +/- standard deviations.

20

FIG. 4 shows a stack-plot of the relative binding potential for MHC class I compared to all other peptides in the EIDB database for the miniaturized dystrophin junction BXA-212372 J4 variants. The original non-natural junction 4 (J4) (R1-R16) is labelled as junction 1 in FIG. 4A and junction 0 in FIG. 4B and has a moderate risk for binding MHC class I. The other numbers on the x-axis indicate the J4 variants (e.g., 13 = J4V13 etc.). Modifications to the junction sequence (J4V4, J4V11, J4V12, J4V13) showed reduced MHC class I binding potential. Both J4V12 and J4V13 had the lowest predicted binding affinity.

25

30

FIG. 5 shows the immunogenic risk profile of miniaturized dystrophin polypeptide junctions. FIG. 5A shows a histogram indicating the proportion of samples, among the 40-samples cell panel tested, that were pulsed with various junction peptides as indicated and

had CD4⁺ proliferating cells (each square represents one patient sample). FIG. 5B shows a histogram indicating the proportion of samples, among the 40-samples cell panel tested, that were pulsed with various junction peptides as indicated and had CD8⁺ proliferating cells (each square represents one patient sample).

5 FIG. 6 shows a histogram indicating increased protein expression of miniaturized dystrophin in tissue culture cells transfected with an expression construct with an SV40 intron and a newly codon-optimized BXA-220931 (SEQ ID NO: 100) by comparison to protein expression in tissue culture cells transfected with comparable amounts of the corresponding expression construct without the SV40 intron and an older codon-optimized
10 coding sequence BXA-212372-J4V13 (SEQ ID NO: 101), as determined by ELISA (AU, arbitrary units). Significance was determined by one-way ANOVA with post-hoc Tukey test. Bar graphs reflect the means +/- standard deviations.

 FIG. 7 shows histograms indicating the effect a variety of promoters and introns/5'UTR coupled to a GFP reporter construct have on expression of GFP in tissue
15 culture. FIG. 7A shows the effect of the indicated promoters on GFP expression. Expression is relative to the use of a CMV promoter (left-most data point). FIG. 7B shows the effect of the indicated introns/5'UTR on GFP expression. Expression is relative to expression resulting from the lack of an intron (left-most data point) Significance with respect to CMV promoter (FIG. 7A) and no intron (FIG. 7B): ** $P < 0.01$, *** $P < 0.001$
20 (one-way ANOVA with post-hoc Tukey test). Bar graphs reflect the means +/- standard deviations.

 FIG. 8 shows expression of miniaturized dystrophin polypeptides in mice and lack of dystrophin protein aggregates. FIG. 8A shows immuno-fluorescence visualization of expression of miniaturized dystrophin polypeptide and wheat germ-agglutinin (WGA) in
25 muscle tissue of *mdx*^{SCSH} mice treated with AAV9-BXA-212372-J4V4. FIG. 8B shows immuno-fluorescence visualization of expression of miniaturized dystrophin polypeptide and wheat germ-agglutinin (WGA) in muscle tissue of *mdx*^{SCSH} mice treated with AAV9-BXA-212372-J4V11. FIG. 8C shows immuno-fluorescence visualization of expression of miniaturized dystrophin polypeptide and wheat germ-agglutinin (WGA) in muscle tissue
30 of *mdx*^{SCSH} mice treated with AAV9-BXA-212372-J4V12. FIG. 8D shows immuno-fluorescence visualization of expression of miniaturized dystrophin polypeptide and wheat

germ-agglutinin (WGA) in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-212372-J4V13 (BXA-220931). No dystrophin protein aggregates are detectable.

FIG. 9 shows nNOS restoration on the muscle sarcolemma of *mdx^{scsn}* mice treated with the indicated AAV9 constructs. Samples were stained with an anti-nNOS antibody or WGA/DAPI as indicated.

FIG. 10 illustrates the effect of miniaturized dystrophin BXA-220931 on the physiology of human isogenic induced-pluripotent stem cell (iPSC)-derived induced cardiomyocytes (iCMs) that carry an E2035X premature stop codon in the dystrophin gene that prevents endogenous dystrophin expression. iCMs were infected with AAV8-BXA-220931 virus to achieve expression. FIG. 10A shows a schematic illustrating the experimental setup and impulse conduction across a microelectrode array in tissue culture. FIG. 10B shows a graphic wherein the conduction velocity of the tested iCMs expressing miniaturized dystrophin polypeptide of BXA-220931 is plotted as a function of time post transfection. BXA-220931 increased conduction velocity of the tested iCMs. Untreated iCMs served as controls. Significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (one-way ANOVA with post-hoc Tukey test). FIG. 10C shows a histogram indicating the expression of miniaturized dystrophin polypeptide BXA-220931 in cells in which conduction velocity was measured. Untreated iCMs served as controls. Bar graphs reflect the means +/- standard deviations.

FIG. 11 shows target engagement and expression of AAV9-BXA-220931 and AAV9-BXA-212374 determined in *mdx^{scsn}* mice at 4 weeks of age. FIG. 11A shows a histogram indicating the relative amount of vector genomes (VG) per μg genomic DNA in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. FIG. 11B shows a histogram indicating the relative amount of miniaturized dystrophin mRNA in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. FIG. 11C shows a histogram indicating the relative amount of miniaturized dystrophin protein in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. Miniaturized dystrophin mRNA and protein expression remained above wild-type dystrophin levels. Wild-type mice and untreated *mdx^{scsn}* mice served as controls. Bar graphs reflect the means +/- standard deviations.

FIG. 12 shows target engagement of AAV9-BXA-220931 and AAV9-BXA-212374 and biodistribution of the corresponding miniaturized dystrophins determined in *mdx^{scsn}* mice at 4 weeks of age. FIG. 12A shows immuno-fluorescence visualization of expression of miniaturized dystrophin polypeptides and α 2-Laminin in diaphragm muscle tissue of mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. Nuclei were visualized with DAPI. The miniaturized dystrophin co-localized with α 2-Laminin, a general marker for muscle sarcolemma. FIG. 12B shows a histogram indicating the relative number of cells in various muscles positive for miniaturized dystrophin (+ve = positive). Wild-type mice and untreated *mdx^{scsn}* mice stained for dystrophin and α 2-Laminin served as controls. Bar graphs reflect the means +/- standard deviations.

FIG. 13 shows an H&E histological analysis of striated muscle from wild-type mice, *mdx^{scsn}* mice and *mdx^{scsn}* mice treated with AAV9-BXA-220931 at 12 weeks of age. Treatment with AAV9-BXA-220931 prevents the *mdx^{scsn}* dystrophic phenotype.

FIG. 14 shows target engagement and expression of AAV9-BXA-220931 and AAV9-BXA-212374 determined in *mdx^{scsn}* mice 12 weeks of age. FIG. 14A shows a histogram indicating the relative amount of vector genomes (VG) per μ g genomic DNA in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. FIG. 14B shows a histogram indicating the relative amount of miniaturized dystrophin mRNA in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. FIG. 14C shows a histogram indicating the relative amount of miniaturized dystrophin protein in muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. Miniaturized dystrophin mRNA and protein expression remained above wild-type dystrophin levels. Wild-type mice and untreated *mdx^{scsn}* mice served as controls. Bar graphs reflect the means +/- standard deviations.

FIG. 15 shows expression of both miniaturized dystrophin BXA-220931 and BXA-212374 is maintained in nearly every muscle fiber and prevention of central nucleation similar to wild-type muscle in muscles of *mdx^{scsn}* mice at 12 weeks of age that had been treated with AAV9-BXA-220931 or AAV9-BXA-212374. FIG. 15A shows immunofluorescence visualization of miniaturized dystrophin and laminin in the tibialis anterior muscle of 12 weeks old *mdx^{scsn}* mice treated with AAV9-BXA-220931. Nuclei were visualized with DAPI. Miniaturized dystrophin remained on the sarcolemma of nearly

every muscle fiber similar to dystrophin in wild-type mice. Untreated *mdx^{scsn}* mice stained for dystrophin and laminin served as controls. FIG. 15B shows a histogram indicating the proportion of cells in various muscles positive for miniaturized dystrophin in *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374 (+ve = positive). FIG. 15C shows a histogram indicating the proportion of muscle cells with central nuclei in *mdx^{scsn}* mice treated with AAV9-BXA-220931 or AAV9-BXA-212374. Wild-type mice and untreated *mdx^{scsn}* mice served as controls. Bar graphs reflect the means +/- standard deviations.

FIG. 16 shows target engagement of AAV9-BXA-220931 and biodistribution of the corresponding miniaturized dystrophin determined in the heart of *mdx^{scsn}* mice at 12 weeks of age. Miniaturized dystrophin polypeptide and laminin were visualized by immunofluorescence in heart muscle tissue of *mdx^{scsn}* mice treated with AAV9-BXA-220931. Nuclei were visualized with DAPI. Expression of miniaturized dystrophin BXA-220931 is seen in nearly every cardiomyocyte in the heart. Wild-type mice and untreated *mdx^{scsn}* mice stained for dystrophin and laminin served as controls.

FIG. 17 shows restoration of the dystrophin glycoprotein complex to the sarcolemma of *mdx^{scsn}* mice treated with AAV9-BXA-220931 at 12 weeks of age. The indicated markers for the muscle sarcolemma, α -syntrophin and β -sarcoglycan, and nNOS were visualized by immunofluorescence. Co-localization of nNOS with α -syntrophin and β -sarcoglycan is seen in *mdx^{scsn}* mice treated with AAV9-BXA-220931, but not in mice treated with AAV9-BXA-212374 or in untreated mice. Untreated *mdx^{scsn}* mice served as controls.

FIG. 18 shows an assessment of muscle mass in treated and untreated *mdx^{scsn}* mice at 12 weeks of age. Tibialis anterior muscle mass is heavier in untreated *mdx^{scsn}* mice due to the significant muscle degeneration and regeneration. Treatment with AAV9-BXA-220931 and AAV9-BXA-212374 prevented this phenotype and resulted in normal muscle mass. Wild-type mice and untreated *mdx^{scsn}* mice served as controls. Bar graphs reflect the means +/- standard deviations.

FIG. 19 shows co-localization of miniaturized dystrophins with ankyrin G in costameres within the sarcolemma of *mdx^{scsn}* mice treated with AAV9-BXA-220931 and AAV9-BXA-212374 at 12 weeks of age. Miniaturized dystrophins and ankyrin G were

visualized by immunofluorescence. Both BXA-220931 and BXA-212374 miniaturized dystrophins localize to both the Z-disks and M bands of costameres similar to dystrophin in wild-type muscles. Wild-type mice and untreated *mdx^{scsn}* mice stained for dystrophin and ankyrin G served as controls.

5 FIG. 20 shows an analysis of the postsynaptic endplate of the 3rd EDL muscle in treated and untreated *mdx^{scsn}* mice at 12 weeks of age. Neuromuscular junctions were labelled with α -bungarotoxin. The postsynaptic endplate is continuous in wild-type muscles, but fragments upon muscle degeneration in muscles of *mdx^{scsn}* mice. Treatment with AAV9-BXA-220931 and AAV9-BXA-212374 prevented the fragmentation of
10 neuromuscular junctions in *mdx^{scsn}* mice.

DETAILED DESCRIPTION OF THE DISCLOSURE

Overview

The present disclosure relates to novel miniaturized dystrophins or the genes encoding the same. The miniaturized dystrophins can be operatively linked to a regulatory
15 cassette. The present disclosure also relates to methods of treating a subject having muscular dystrophy, sarcopenia, heart failure, or cachexia. Further, the present disclosure relates to methods of prophylactically treating a subject at risk of developing muscular dystrophy, sarcopenia, heart failure, or cachexia. The methods for treating a subject having, or at risk of developing, muscular dystrophy, sarcopenia, heart failure, or cachexia can
20 comprise administering a pharmaceutical composition including a miniaturized dystrophin gene and a delivery vehicle to the subject.

Definitions

In order that the present disclosure can be more readily understood, certain terms are first defined. As used in this application, except as otherwise expressly provided herein,
25 each of the following terms shall have the meaning set forth below. Additional definitions are set forth throughout the application.

The term "and/or" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term "and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B,"

"A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following aspects: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone). It is understood that wherever aspects are described herein with the language "comprising,"
5 otherwise analogous aspects described in terms of "consisting of" and/or "consisting essentially of" are also provided.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure is related. For example, the Concise Dictionary of Biomedicine and Molecular
10 Biology, Juo, Pei-Show, 2nd ed., 2002, CRC Press; The Dictionary of Cell and Molecular Biology, 3rd ed., 1999, Academic Press; and the Oxford Dictionary Of Biochemistry And Molecular Biology, Revised, 2000, Oxford University Press, provide one of skill with a general dictionary of many of the terms used in this disclosure.

Units, prefixes, and symbols are denoted in their Système International de Unites
15 (SI) accepted form. Numeric ranges are inclusive of the numbers defining the range. The headings provided herein are not limitations of the various aspects of the disclosure, which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification in its entirety.

Dystrophin (DMD) is a large human X-linked gene that encodes Dystrophin. The
20 protein Dystrophin is a 427 kDa cytoskeletal protein that localizes to the cytoplasmic face of the sarcolemma and is enriched at costameres in muscle fibers. The Dystrophin protein has four main functional domains: an actin-binding amino-terminal domain (ABD1); a central rod domain comprising a series of rods, called "spectrin repeat domains" and hinges; a cysteine-rich domain; and a carboxyl-terminus.

25 As used herein, the term "miniaturized dystrophin polypeptide" or "miniaturized dystrophin peptide" refers to a polypeptide that is smaller in size than the full-length wild-type dystrophin polypeptide. In some embodiments, the miniaturized dystrophin polypeptide is capable of altering (increasing or decreasing, as the case may be) a measurable value of muscle physiology or anatomy in a DMD animal model by at least
30 approximately 10 or 20% of the wild type value, such that the value is closer to the wild-type value (e.g., a *mdx* mouse has a measurable value of muscle physiology or anatomy

that is 50% of the wild-type value, and this value is increased to at least 60% of the wild-type value; or a *mdx* mouse has a measurable value of muscle physiology or anatomy that is 150% of the wild-type value, and this value is decreased to at most 140% of the wild-type value). In certain embodiments, the miniaturized dystrophin polypeptide is capable of altering a measurable value of muscle physiology or anatomy in a DMD animal model by at least approximately 30% of the wild type value. In some embodiments, the miniaturized dystrophin polypeptide is capable of altering a measurable value of muscle physiology or anatomy in a DMD animal model to a level similar to the wild-type value (e.g., $\pm 4\%$).

As used herein, the term "spectrin repeats" or "spectrin-like repeats" refers to peptides composed of approximately 100 amino acids that are responsible for the rod-like shape of many structural proteins including, but not limited to, dystrophin, wherein the spectrin repeats are typically present in multiple copies. Spectrin repeats can include mutations of the natural peptide sequences, such as conservative and/or non-conservative changes in amino acid sequence, as well as the addition or deletion of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids to/from the end of a spectrin repeat or within the spectrin repeat. In some embodiments, each spectrin repeat (each of R1 to R24) has at least 70%, at least 75%, 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 100% sequence identity to the naturally occurring spectrin repeat (each of the naturally occurring R1 to R24).

As used herein, the term "spectrin repeat encoding sequences" refers to nucleic acid sequences encoding spectrin repeat peptides. This term includes natural and synthetic nucleic acid sequences encoding the spectrin repeats (e.g., both the naturally occurring and mutated spectrin repeat peptides).

As used herein, the term "spectrin repeat domain" refers to the region in a miniaturized dystrophin polypeptide that contains the spectrin repeats of the miniaturized dystrophin polypeptide.

The term "fused" refers to a first amino acid sequence that is linked in frame to a second amino acid sequence with which it is not normally linked in nature, forming a "fusion" protein/polypeptide. These fused amino acid sequences which normally exist in separate proteins can be brought together in the fusion polypeptide, or the amino acid sequences which normally exist in the same protein can be placed in a new arrangement in

the fusion polypeptide. A fusion protein is created, for example, by chemical peptide synthesis, or by recombinant DNA technology whereby a polynucleotide is created, and then translated, in which the peptide regions are encoded in the desired relationship. A fusion protein can also comprise a second amino acid sequence associated with the first amino acid sequence by a covalent, non-peptide bond or by a non-covalent bond. In some embodiments, "fusion" between two polypeptides is achieved by a linker. Linkers can be amino acids or other chemical structures. In some embodiments, linkers can be synthetic. In some embodiments, "fusion" between two polypeptides is a direct fusion, i.e., without intervening linker. The term "fused directly" or "direct fusion" refers to a linkage between two polypeptide chains by a peptide bond. For example, a first amino acid is "fused directly" to a second amino acid when the first amino acid is "fused" to a second amino acid by a peptide bond.

"Heterologous" and "heterologous moiety" in reference to a polypeptide moiety or polynucleotide moiety that is part of a larger polypeptide or polynucleotide, respectively, describes a polypeptide or polynucleotide that originates from a different polypeptide or polynucleotide than the remaining part of the polypeptide or polynucleotide molecule. The additional heterologous component of the polypeptide or polynucleotide can originate from the same organism as the remaining polypeptide or polynucleotide, respectively, described herein, or the additional components can be from a different organism. For instance, a heterologous polypeptide can be synthetic, or derived from a different species, different cell type of an individual, or the same or different type of cell of distinct individuals. In one aspect, a heterologous moiety is a polypeptide fused to another polypeptide to produce a polypeptide. In another aspect, a heterologous moiety is a non-polypeptide such as PEG conjugated to a polypeptide or protein.

As used herein, the terms "muscle cell" refers to a cell derived from muscle tissue, including, but not limited to, cells derived from skeletal muscle, smooth muscle (e.g. from the digestive tract, urinary bladder, and blood vessels), and cardiac muscle. The term includes muscle cells *in vitro*, *ex vivo*, and *in vivo*. Thus, for example, an isolated cardiomyocyte would constitute a muscle cell, as would a cell as it exists in muscle tissue present in a subject *in vivo*. This term also encompasses both terminally differentiated and

nondifferentiated muscle cells, such as myocytes, myotubes, myoblasts, cardiomyocytes, and cardiomyoblasts.

As used herein, the term "muscle-specific" in reference to a gene regulatory element (e.g., enhancer sequence, promoter sequence) means that the regulatory element drives transcriptional activity primarily in muscle cells or tissue (e.g., 20:1) compared to the transcriptional activity driven by the regulatory element in other tissues. Assays to determine the muscle-specificity of a regulatory element are known in the art (e.g., *in vitro* assay using murine muscle cells and liver cells transfected with an expression vector comprising the regulatory element to be tested driving expression of a beta-galactoside reporter).

As used herein, the term "adeno-associated virus" or "AAV" includes but is not limited to, AAV type 1, AAV type 2, AAV type 3 (including types 3A and 3B), AAV type 4, AAV type 5, AAV type 6, AAV type 7, AAV type 8, AAV type 9, AAV type 10, AAV type 11, AAV type 12, AAV type 13, snake AAV, avian AAV, bovine AAV, canine AAV, equine AAV, ovine AAV, goat AAV, shrimp AAV, primate AAV, non-primate AAV, and ovine AAV, those AAV serotypes and clades disclosed by Gao *et al.* (J. Virol. 78:6381 (2004)) and Moris *et al.* (Virol. 33:375 (2004)), and any other AAV now known or later discovered. *See, e.g.*, Fields *et al.* VIROLOGY, volume 2, chapter 69 (4th ed., Lippincott-Raven Publishers). AAV refers to a *Dependoparvovirus* within the *Parvoviridae* family of viruses. For example, the AAV can be an AAV derived from a naturally occurring "wild-type" virus, an AAV derived from a recombinant AAV (rAAV) genome packaged into a capsid derived from capsid proteins encoded by a naturally occurring *cap* gene and/or a rAAV genome packaged into a capsid derived from capsid proteins encoded by a non-natural capsid *cap* gene. As used herein, "AAV" can be used to refer to the virus itself or derivatives thereof. The term covers all subtypes and both naturally occurring and recombinant forms, except where expressly indicated otherwise. "Primate AAV" refers to AAV that infect primates, "non-primate AAV" refers to AAV that infects animals other than primates, "bovine AAV" refers to AAV that infect bovine mammals, etc. *See, e.g.*, BERNARD N. FIELDS *et al.*, VIROLOGY, volume 2, chapter 69 (3 d ed., Lippincott-Raven Publishers).

The term "rAAV" refers to a "recombinant AAV." In some embodiments, a recombinant AAV has an AAV genome in which part or all of the *rep* and *cap* genes have been replaced with heterologous polynucleotide sequences.

5 An "AAV vector" or "adeno-associated virus vector" as used herein refers to an rAAV comprising a polynucleotide sequence not of AAV origin (i.e., a polynucleotide heterologous to AAV), typically a sequence of interest for the genetic transformation of a cell. In general, the heterologous polynucleotide is flanked by at least one, and generally by two, AAV inverted terminal repeat sequences (ITRs).

10 A "capsid-free" or "capsid-less" (or variations thereof) viral (e.g., AAV) genome or nucleic acid molecule refers to a genome or nucleic acid molecule free from a capsid. In some embodiments, the capsid-less genome or nucleic acid molecule does not contain sequences encoding, for example, an AAV Rep protein.

15 An "AAV" or "AAV viral particle" or "AAV vector" or "rAAV vector particle" refers to a viral particle composed of at least one AAV capsid protein (typically of all of the capsid proteins of a wild-type AAV) and an encapsidated polynucleotide. If the particle comprises a heterologous polynucleotide (i.e. a polynucleotide other than a wild-type AAV genome, such as a transgene to be delivered to a mammalian cell), it is typically referred to as an "rAAV vector particle" or simply an "AAV vector."

20 A "helper virus" for AAV refers to a virus that allows AAV (e.g., wild-type AAV) to be replicated and packaged by a mammalian cell. A variety of such helper viruses for AAV are known in the art, including adenoviruses, herpesviruses and poxviruses such as vaccinia. The adenoviruses encompass a number of different subgroups, although Adenovirus type 5 of subgroup C is most commonly used. Numerous adenoviruses of human, non-human mammalian and avian origin are known and available from depositories
25 such as the ATCC. Viruses of the herpes family include, for example, herpes simplex viruses (HSV) and Epstein-Barr viruses (EBV), as well as cytomegaloviruses (CMV) and pseudorabies viruses (PRV), all of which are also available from depositories such as ATCC.

30 As used herein, the term "inverted terminal repeat" (or "ITR") refers to a single stranded sequence of nucleotides followed downstream by its reverse complement. The intervening sequence of nucleotides between the initial sequence and the reverse

complement can be any length including zero. The AAV genome typically comprises inverted terminal repeats (ITRs) at both ends, wherein each end typically is palindromic and can form a hairpin.

5 The terms "polynucleotide" and "nucleic acid" are used interchangeably herein and refer to a biopolymer composed of a plurality of nucleotide monomers covalently bonded in a chain

The term "tropism" as used herein refers to a virus's (e.g., AAV's) ability to infect only one or more particular cell types and its ability to interact only with specific cell surface moieties to achieve cell entry, optionally and preferably followed by expression
10 (e.g., transcription and, optionally, translation) of sequences carried by the virus (e.g., AAV) into the cell (e.g., for a recombinant virus, expression of the heterologous nucleotide sequence(s)).

As used herein, the term "transduction" refers to the entry of the virus (e.g., AAV) into the cell and the transfer of genetic material contained within the virus into the cell to
15 obtain expression from the virus genome. Typically, a virus (e.g., AAV) enters cells in accordance with its tropism.

"Administering" refers to the physical introduction of a therapeutic agent to a subject, using any of the various methods and delivery systems known to those skilled in the art. Exemplary routes of administration, *e.g.*, for an AAV therapy, include intravenous,
20 intramuscular, intraarterial, intrathecal, intralymphatic, intralesional, intracapsular, intraorbital, intracardiac, intradermal, intraperitoneal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural, intrasterna, oral, rectal, topical, epidermal, mucosal, intranasal, vaginal, rectal, and sublingual administration. Administering can also be performed, for example, once, a plurality of
25 times, and/or over one or more extended periods.

"Treatment" or "therapy" of a subject refers to any type of intervention or process performed on, or the administration of an active agent to, a subject with the objective of reversing, alleviating, ameliorating, inhibiting, slowing down, or preventing the onset, progression, development, severity, or recurrence of a symptom, complication, condition,
30 or biochemical indicia associated with a disease.

A "therapeutically effective amount," "therapeutic dose," "effective dose," or "effective dosage," as used herein, means an amount or a dose that achieves a therapeutic goal, as described herein. One of ordinary skill in the art will further understand that a therapeutically effective amount etc. can be administered in a single dose, or can be
5 achieved by administration of multiple doses (*i.e.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more doses). The ability of a therapeutic agent to promote disease regression or inhibit the development or recurrence of the disease can be evaluated using a variety of methods known to the skilled practitioner, such as in human subjects during clinical trials, in animal model systems predictive of efficacy in humans, or by assaying the activity of the agent in *in vitro*
10 assays.

A "subject" includes any human or non-human animal. The term "nonhuman animal" includes, but is not limited to, vertebrates such as nonhuman primates, sheep, dogs, and rodents such as mice, rats, and guinea pigs. In some embodiments, the subject is a human. The terms "subject" and "patient" are used interchangeably herein.
15 As used herein, the terms "ug" and "uM" are used interchangeably with "μg" and "μM," respectively.

The use of the alternative (*e.g.*, "or") should be understood to mean either one, both, or any combination thereof of the alternatives. As used herein, the indefinite articles "a" or "an" should be understood to refer to "one or more" of any recited or enumerated
20 component or entity.

Approximately or about: As used herein, the term "approximately" or "about," as applied to one or more values of interest, refers to a value that is similar to a stated reference value and within a range of values that fall within 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either
25 direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value). When the term "approximately" or "about" is applied herein to a particular value, the value without the term "approximately" or "about" is also disclosed herein. As described herein, any concentration range, percentage range, ratio range, or integer
30 range is to be understood to include the value of any integer within the recited range and,

when appropriate, fractions thereof (such as one tenth and one hundredth of an integer), unless otherwise indicated.

Various aspects of the disclosure are described in further detail in the following subsections.

5 **Polynucleotides and Polypeptides**

Miniaturized dystrophin

The present disclosure is directed to a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide. In some embodiments, the miniaturized dystrophin polypeptide comprises at least three hinge
10 domains of dystrophin and at least five Spectrin repeat domains.

Dystrophin is a rod-shaped cytoplasmic protein that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. This protein is located primarily in muscles used for movement (skeletal muscles) and in heart (cardiac) muscle. Small amounts of dystrophin are present in nerve cells in the brain. In
15 skeletal and cardiac muscles, dystrophin is part of a group of proteins (a protein complex) that work together to strengthen muscle fibers and protect them from injury as muscles contract and relax. The dystrophin complex acts as an anchor, connecting each muscle cell's structural framework (cytoskeleton) with the lattice of proteins and other molecules outside the cell (extracellular matrix). The dystrophin complex can also play a role in cell signaling
20 by interacting with proteins that send and receive chemical signals.

The DMD gene, encoding the full length dystrophin protein, is one of the longest human genes known, covering 2.3 megabases (0.08% of the human genome) at locus Xp21. The primary transcript in muscle measures about 2,100 kilobases and takes 16 hours to transcribe; the mature mRNA measures 14.0 kilobases. The 79-exon muscle transcript
25 codes for a protein of 3685 amino acid residues.

Disclosed herein are amino acid and nucleotide sequences for dystrophin. The amino acid sequence constituting human wild type dystrophin, isoform Dp427m, is known as UniProt identifier No. NP_003997.1 and shown in Table 1.

Table 1. Amino Acids sequence of full-length Dystrophin Protein (NP_003997.1).

SEQ ID NO: 1
MLWWEVEEDCYEREDVQKKTF TKWVNAQFSKFGKQH IENLFS DLQDGRRL LDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVD GNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYP QVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFNIAR YQLGIEKLLDPEDVDTTYDPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPP KVTKEEHFQLHHQM HYSQQITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTS DPTRSPFPSQHLEAPEDKSF GSSLMESVNLDRYQTAL EEVLSWLLSAEDTLQ AQGEISNDVEVVKDQFH THEGYMMDLTAHQGRVGNILQLGSKLIGTGKLSE DEETEVQEQMNLLNSRWECLRVASMEKQSNLHRVLM DLQNQKLKELNDW LTKTEERTRKMEEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVNSLTHMV VVVDESSGDHATAALEEQLKVLGDRWANICRWTE DRWVLLQDILLKWQRL TEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQ SMGKLYSLKQDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQIS QAVTTTQPSLTQTTVMETVTTVT TREQILVKHAEELPPPPQKKRQITVDSEI RKRLDVDITELHSWITRSEAVLQSPEFAIFRKEGNFSDLKEKVNAIEREKA EKF RKLQDASRSAQALVEQMVNEGVNADSIKQASEQLNSRWIEFCQLLSERL NW LEYQNNIIAFYNQLQQLEQMTTAAENWLKIQPTTPSEPTAIKSQLKICKDEVN RLSGLQPQIERLKIQSIALKEKGQGP MFLDADFVAFTNHFKQVFS DVQAREKE LQTIFDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQA LQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWKKLSSQLV EHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL KQCRLVSDIQTIPSLNSVNEG GQKIKNEAEPEFASRLETTELKELNTQWDH MCQQVYARKEALKGGLEKTVSLQKDLSEMHEWMTQAE EEEYLERDFEYKTP DELQKAVEEMKRAKEEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELE TLTTNYQWLCTRLNGKCKTLEEVWACWHELLSYLEKANKWLNEVEFKLKT TENIPGGAE EISEVLD SLENLMRHSEDNPNQIRILAQTLTDGGVMDELINEELE TFNSRWRELHEEA VRRQKLL EQSISQAQETEKSLHLIQESLTFIDKQLAAYIA DKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEAAQRVLSQIDVAQK KLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALET KSVEQEVVQS QLNHCVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLH YNELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRS AVE GMPSNLDSEVAWGKATQKEIEKQKVHLKSITEVGEALKT VLGKKETLVEDK LSLLNSNWIAVTSRAE EWLNLLEYQKHMETFDQNVDHITKWIIQADTLLDE SEKKKPQQKEDVLKRLKAELNDIRPKVDSTRDQAANLMANRGDHCRKLVEP QISELNHRFAAISHRIKTGKASIPLKELEQFN SDIQKLL EPLEAEIQQGVNLKEE DFNKM NEDNEGTVKELLQRGDNLQQRITDERKREEIKIKQQLLQTKHNAL KDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKLASLPEPRDERKIK EIDRELQKKKEELNAVRRQA EGLSEDGAAMAVEPTQIQLSKRWREIESKFAQ FRRLNFAQIHTVREETMMVMTE DMPLEISYVPSTYLTEITHVSQALLEVEQLL NAPDLCAKDFEDLFKQEESLKNIKDSLQSSGRIDIIH SKKTAALQSATPVER VKLQEALSQLDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLT EAEQFLRKTQIPENWEHAKYKWYKELQDGIGQRQT VVRTL NATGEEIIQQS

SKTDASILQEKLGSNLRWQEVCKQLSDRKKRLEEQKNILSEFQORDLNEFVL
WLEEADNIAISIPLEPGKEQQLKEKLEQVKLLVEELPLRQGILKQLNETGGPVL
VSAPISPEEQDKLENKLKQTNLQWIKVSRALPEKQGEIEAQIKDLGQLEKKLE
DLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFDVQETEIAVQAKQPDVEEILS
KGQHLYKEKPATQPVKRKLEDLSSEWKAVNRLLQELRAKQPD LAPGLTTIG
ASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWTELTDWLS
LLDQVIKSQRVMVGDLNEMIIKQKATMQDLEQRRPQLEELITAAQNLKN
KTSNQEARTIITDRIERIQNWDEVQEHLQNRQQLNEMLKDSTQWLEAKEE
AEQVLGQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVAN
DLALKLLRDYSADDTRKVHMITENINASWRSIHKRVSEREAALETHRLLOQ
FPLDLEKFLAWL TEAETTANVLQDATRKERLLED SKGVKELMKQWQDLQG
EIEAHTDVYHNLDENSQKILRSLEGSDDAVLLQRRLDNMNFKWSELKKSL
NIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPAVQKQN
DVHRAFKRELKTKPEVIMSTLETVRIFLTEQPLEGLEKLYQEPRELPPERAQ
NVTRLLRKQAEENVTEWEKLNLSADWQRKIDETLERLQELQEATDEL DLK
LRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQ
LTTLGIQLSPYNLSTLEDLNTRWKLQVAVEDRVRQLHEAHRDFGPASQHFL
STSVQGPWERAI SPNKVPYYINHETQTTCDWHPKMTLEYQSLADLNNVRFSA
YRTAMKLRRLQKALCLDLSLSAACDALDQHNKQNDQPM DILQINCLTTI
YDRLEQEHNNLVNVPCLVDMCLNWLNVYDTGRTGRIRVLSFKTGIISLCKA
HLEDKYRYLKFQVASTGFCDQRRLLGLLHDSIQIPRQLGEVASFGGSNIEPS
VRSCFQFANNKPEIEAALFLDWMRLPQSMVWLPVLHRVAAAE TAKHQAK
CNICKECPHGFYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTS
GEDVRDFAKVLKNKFR TKRYFAKHPRMGYLPVQTVLEGDNMETPVT LINF
WPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLND SISPNESIDDE
HLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGERILADLEENRN LQAE
YDRLKQQHEHKGLSPLSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEAR
MQILEDHNKQLESQ LHLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPM
LLRVVGSQTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKP
MREDTM

Various other dystrophin isoforms are known in the art that result from alternative splicing. In some embodiments, the constructs comprise the nucleotide sequences recited in Table 2, or parts thereof.

- 5 Also disclosed herein is a nucleotide sequence encoding the full-length dystrophin protein.

Table 2. Nucleotide sequence of full-length Dystrophin Protein (isoform Dp427m).

SEQ ID NO: 2
GGGATTCCCTCACTTTCCCCCTACAGGACTCAGATCTGGGAGGCAATTACCTTCGG AGAAAAACGAATAGGAAAAACTGAAGTGTTACTTTTTTTAAAGCTGCTGAAGTTT GTTGGTTTCTCATTGTTTTTAAGCCTACTGGAGCAATAAAGTTTGAAGAACTTTTA

CCAGGTTTTTTTTATCGCTGCCTTGATATACACTTTTCAAAATGCTTTGGTGGAAG
AAGTAGAGGACTGTTATGAAAGAGAAGATGTTCAAAAGAAAACATTCACAAAAT
GGGTAAATGCACAATTTTCTAAGTTTGGGAAGCAGCATATTGAGAACCTCTTCAGT
GACCTACAGGATGGGAGGCGCCTCCTAGACCTCCTCGAAGGCCTGACAGGGCAAA
AACTGCCAAAAGAAAAAGGATCCACAAGAGTTCATGCCCTGAACAATGTCAACA
AGGCACTGCGGGTTTTGCAGAACAATAATGTTGATTTAGTGAATATTGGAAGTAC
TGACATCGTAGATGGAAATCATAAACTGACTCTTGTTTGGTTTGGAAATATAATCC
TCCACTGGCAGGTCAAAAATGTAATGAAAAATATCATGGCTGGATTGCAACAAAC
CAACAGTGAAAAGATTCTCCTGAGCTGGGTCCGACAATCAACTCGTAATTATCCA
CAGGTTAATGTAATCAACTTCACCACCAGCTGGTCTGATGGCCTGGCTTTGAATGC
TCTCATCCATAGTCATAGGCCAGACCTATTTGACTGGAATAGTGTGGTTTGCCAGC
AGTCAGCCACACAACGACTGGAACATGCATTCAACATCGCCAGATATCAATTAGG
CATAGAGAACTACTCGATCCTGAAGATGTTGATACCACCTATCCAGATAAGAAG
TCCATCTTAATGTACATCACATCACTCTTCCAAGTTTTGCCTCAACAAGTGAGCAT
TGAAGCCATCCAGGAAGTGGAATGTTGCCAAGGCCACCTAAAGTGACTAAAGA
AGAACATTTTCAGTTACATCATCAAATGCACTATTCTCAACAGATCACGGTCAGTC
TAGCACAGGGATATGAGAGAACCTTCTCCCTAAGCCTCGATTCAAGAGCTATGC
CTACACACAGGCTGCTTATGTCACCACCTCTGACCCTACACGGAGCCCATTTCCTT
CACAGCATTTGGAAGCTCCTGAAGACAAGTCATTTGGCAGTTCATTGATGGAGAG
TGAAGTAAACCTGGACCGTTATCAAACAGCTTTAGAAGAAGTATTATCGTGGCTT
CTTTCTGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAG
TGGTGAAGACCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCA
TCAGGGCCGGGTTGGTAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGA
AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATT
CAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATA
GAGTTTTAATGGATCTCCAGAATCAGAACTGAAAGAGTTGAATGACTGGCTAAC
AAAAACAGAAGAAAGAACAAAGGAAAATGGAGGAAGAGCCTCTTGACCTGATCT
TGAAGACCTAAAACGCCAAGTACAACAACATAAGGTGCTTCAAGAAGATCTAGA
ACAAGAACAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTGGTAGTTGATGAA
TCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGGAG
ATCGATGGGCAAACATCTGTAGATGGACAGAAGACCGCTGGGTTCTTTTACAAGA
CATCCTTCTCAAATGGCAACGTCTTACTGAAGAACAGTGCCTTTTTAGTGCATGGC
TTTCAGAAAAAGAAGATGCAGTGAACAAGATTACACAACCTGGCTTTAAAGATCA
AAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTTAAAGCGGATCTAGAA
AAGAAAAAGCAATCCATGGGCAAACCTGTATTCCTCAACAAGATCTTCTTTCAA
CACTGAAGAATAAGTCAGTGACCCAGAAGACGGAAGCATGGCTGGATAACTTTGC
CCGGTGTTGGGATAATTTAGTCCAAAAACCTTGAAAAGAGTACAGCACAGATTTCA
CAGGCTGTCACCACCCTCAGCCATCACTAACACAGACAACCTGTAATGGAAACAG
TAACTACGGTGACCACAAGGGAACAGATCCTGGTAAAGCATGCTCAAGAGGAACT
TCCACCACCACCTCCCCAAAAGAAGAGGCAGATTACTGTGGATTCTGAAATTAGG
AAAAGGTTGGATGTTGATATAACTGAACTTCACAGCTGGATTACTCGCTCAGAAG
CTGTGTTGCAGAGTCCTGAATTTGCAATCTTTCGGAAGGAAGGCAACTTCTCAGAC
TTAAAAGAAAAAGTCAATGCCATAGAGCGAGAAAAAGCTGAGAAGTTCAGAAAA
CTGCAAGATGCCAGCAGATCAGCTCAGGCCCTGGTGGAACAGATGGTGAATGAGG
GTGTTAATGCAGATAGCATCAAACAAGCCTCAGAACAACCTGAACAGCCGGTGGAT
CGAATTCTGCCAGTTGCTAAGTGAGAGACTTAACTGGCTGGAGTATCAGAACAAC

ATCATCGCTTTCTATAATCAGCTACAACAATTGGAGCAGATGACAACTACTGCTGA
AAACTGGTTGAAAATCCAACCCACCACCCCATCAGAGCCAACAGCAATTAAGT
CAGTTAAAAATTTGTAAGGATGAAGTCAACCGGCTATCAGGTCTTCAACCTCAA
TTGAACGATTAAAAATTCAAAGCATAGCCCTGAAAGAGAAAGGACAAGGACCCA
TGTTCTGGATGCAGACTTTGTGGCCTTTACAAATCATTTTAAGCAAGTCTTTCTG
ATGTGCAGGCCAGAGAGAAAGAGCTACAGACAATTTTGACACTTTGCCACCAAT
GCGCTATCAGGAGACCATGAGTGCCATCAGGACATGGGTCCAGCAGTCAGAAACC
AAACTCTCCATACCTCAACTTAGTGTACCGACTATGAAATCATGGAGCAGAGAC
TCGGGGAATTGCAGGCTTTACAAAGTTCTCTGCAAGAGCAACAAAGTGGCCTATA
CTATCTCAGCACCCTGTGAAAGAGATGTGCAAGAAAGCGCCCTCTGAAATTAGC
CGGAAATATCAATCAGAATTTGAAGAAATTGAGGGACGCTGGAAGAAGCTCTCCT
CCCAGCTGGTTGAGCATTGTCAAAGCTAGAGGAGCAAATGAATAAACTCCGAAA
AATTCAGAATCACATACAAACCTGAAGAAATGGATGGCTGAAGTTGATGTTTTT
CTGAAGGAGGAATGGCCTGCCCTGGGGATTGAGAAATTCTAAAAAGCAGCTGA
AACAGTGCAGACTTTTAGTCAGTGATATTCAGACAATTCAGCCCAGTCTAAACAG
TGTCAATGAAGGTGGGCAGAAGATAAAGAATGAAGCAGAGCCAGAGTTTGCTTC
GAGACTTGAGACAGAACTCAAAGAACTTAACACTCAGTGGGATCACATGTGCCAA
CAGGTCTATGCCAGAAAGGAGGCCTTGAAGGGAGGTTTGAGAGAAACTGTAAGC
CTCCAGAAAGATCTATCAGAGATGCACGAATGGATGACACAAGCTGAAGAAGAG
TATCTTGAGAGAGATTTTGAATATAAACTCCAGATGAATTACAGAAAGCAGTTG
AAGAGATGAAGAGAGCTAAAGAAGAGGCCCAACAAAAAGAGCGAAAGTGAAA
CTCCTTACTGAGTCTGTAAATAGTGTATAGCTCAAGCTCCACCTGTAGCACAGA
GGCCTTAAAAAGGAACTTGAACTCTAACCACCACTACCAGTGGCTCTGCACT
AGGCTGAATGGGAAATGCAAGACTTTGGAAGAAGTTTGGGCATGTTGGCATGAGT
TATTGTCATACTTGAGAGAAAGCAAACAAGTGGCTAAATGAAGTAGAATTTAACT
TAAAACCACTGAAAACATTCTCGTGGCGGAGCTGAGGAAATCTCTGAGGTGCTAGAT
TCACTTGAAAATTTGATGCGACATTGAGAGGATAACCCAAATCAGATTCGCATATT
GGCACAGACCCTAACAGATGGCGGAGTCATGGATGAGCTAATCAATGAGGAACTT
GAGACATTTAATTCTCGTTGGAGGGAATAACATGAAGAGGCTGTAAGGAGGCAAA
AGTTGCTTGAACAGAGCATCCAGTCTGCCAGGAGACTGAAAAATCCTTACACTT
AATCCAGGAGTCCCTCACATTGACAAGCAGTTGGCAGCTTATATTGCAGAC
AAGGTGGACGCAGCTCAAATGCCTCAGGAAGCCAGAAAATCCAATCTGATTTGA
CAAGTCATGAGATCAGTTTAGAAGAAATGAAGAAACATAATCAGGGGAAGGAGG
CTGCCAAAGAGTCCTGTCTCAGATTGATGTTGCACAGAAAAAATTACAAGATGT
CTCCATGAAGTTTCGATTATTCCAGAAACCAGCCAATTTTGAGCAGCGTCTACAAG
AAAGTAAGATGATTTTAGATGAAGTGAAGATGCACTTGCCCTGCATTGGAAACAAA
GAGTGTGGAACAGGAAGTAGTACAGTCACAGCTAAATCATTGTGTGAAGTTGTAT
AAAAGTCTGAGTGAAGTGAAGTCTGAAGTGGAAATGGTGATAAAGACTGGACGT
CAGATTGTACAGAAAAAGCAGACGGAAAAATCCCAAAGAACTTGATGAAAGAGTA
ACAGCTTTGAAATTGCATTATAATGAGCTGGGAGCAAAGGTAACAGAAAGAAAG
CAACAGTTGGAGAAATGCTTGAAATTGTCCCGTAAGATGCGAAAGGAAATGAATG
TCTTGACAGAAATGGCTGGCAGCTACAGATATGGAATTGACAAAGAGATCAGCAGT
TGAAGGAATGCCTAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAGGCTACTCAA
AAAGAGATTGAGAAACAGAAGGTGCACCTGAAGAGTATCACAGAGGTAGGAGAG
GCCTTGAAAACAGTTTTTGGGCAAGAAGGAGACGTTGGTGGAAGATAAACTCAGTC
TTCTGAATAGTAACTGGATAGCTGTCACCTCCCGAGCAGAAGAGTGGTTAAATCTT

TTGTTGGAATACCAGAAACACATGGAAACTTTTGACCAGAATGTGGACCACATCA
CAAAGTGGATCATTGAGGCTGACACACTTTTGGATGAATCAGAGAAAAAGAAACC
CCAGCAAAAAGAAGACGTGCTTAAGCGTTTAAAGGCAGAACTGAATGACATACG
CCCAAAGGTGGACTCTACACGTGACCAAGCAGCAAACTTGATGGCAAACCGCGGT
GACCACTGCAGGAAATTAGTAGAGCCCCAAATCTCAGAGCTCAACCATCGATTTG
CAGCCATTTACACAGAATTAAGACTGGAAAGGCCTCCATTCCCTTTGAAGGAATT
GGAGCAGTTTAACTCAGATATACAAAAATTGCTTGAACCACTGGAGGCTGAAATT
CAGCAGGGGGTGAATCTGAAAGAGGAAGACTTCAATAAAGATATGAATGAAGAC
AATGAGGGTACTGTAAAAGAATTGTTGCAAAGAGGAGACAACCTTACAACAAAGA
ATCACAGATGAGAGAAAGAGAGAGGAAATAAAGATAAAACAGCAGCTGTTACAG
ACAAAACATAATGCTCTCAAGGATTTGAGGTCTCAAAGAAGAAAAAAGGCTCTAG
AAATTTCTCATCAGTGGTATCAGTACAAGAGGCAGGCTGATGATCTCCTGAAATG
CTTGGATGACATTGAAAAAAATTAGCCAGCCTACCTGAGCCCAGAGATGAAAGG
AAAATAAAGGAAATTGATCGGGAATTGCAGAAGAAGAAAGAGGAGCTGAATGCA
GTGCGTAGGCAAGCTGAGGGCTTGTCTGAGGATGGGGCCGCAATGGCAGTGGAGC
CAACTCAGATCCAGCTCAGCAAGCGCTGGCGGGAAATTGAGAGCAAATTTGCTCA
GTTTCGAAGACTCAACTTTGCACAAATTCACACTGTCCGTGAAGAAACGATGATG
GTGATGACTGAAGACATGCCTTTGGAAATTTCTTATGTGCCTTCTACTTATTTGACT
GAAATCACTCATGTCTCACAAAGCCCTATTAGAAGTGGAAACAACCTTCTCAATGCTCC
TGACCTCTGTGCTAAGGACTTTGAAGATCTCTTTAAGCAAGAGGAGTCTCTGAAG
AATATAAAAGATAGTCTACAACAAAGCTCAGGTCGGATTGACATTATTCATAGCA
AGAAGACAGCAGCATTGCAAAGTGCAACGCCTGTGGAAAGGGTGAAGCTACAGG
AAGCTCTCTCCAGCTTGATTTCCAATGGGAAAAAGTTAACAAAATGTACAAGGA
CCGACAAGGGCGATTTGACAGATCTGTTGAGAAATGGCGGCGTTTTTCATTATGAT
ATAAAGATATTTAATCAGTGGCTAACAGAAGCTGAACAGTTTCTCAGAAAGACAC
AAATTCCTGAGAATTGGGAACATGCTAAATACAAATGGTATCTTAAGGAACTCCA
GGATGGCATTGGGCAGCGGCAAACCTGTTGTCAGAACATTGAATGCAACTGGGGAA
GAAATAATTCAGCAATCCTCAAAAACAGATGCCAGTATTCTACAGGAAAAAATTGG
GAAGCCTGAATCTGCGGTGGCAGGAGGTCTGCAAACAGCTGTCAGACAGAAAAA
AGAGGCTAGAAGAACAAGAATATCTTGTGAGAATTTCAAAGAGATTTAAATGA
ATTTGTTTTATGGTTGGAGGAAGCAGATAACATTGCTAGTATCCCACTTGAACCTG
GAAAAGAGCAGCAACTAAAAGAAAAGCTTGAGCAAGTCAAGTTACTGGTGGAAAG
AGTTGCCCCCTGCGCCAGGGAATTCTCAAACAATTAATGAACTGGAGGACCCGT
GCTTGTAAGTGCTCCCATAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAGCTC
AAGCAGACAAATCTCCAGTGGATAAAGGTTTCCAGAGCTTTACCTGAGAAACAAG
GAGAAATTGAAGCTCAAATAAAAGACCTTGGGCAGCTTGAAAAAAGCTTGAAG
ACCTTGAAGAGCAGTTAAATCATCTGCTGCTGTGGTTATCTCCTATTAGGAATCAG
TTGGAAATTTATAACCAACCAACCAAGAAGGACCATTTGACGTTCAAGGAACTG
AAATAGCAGTTCAAGCTAAACAACCGGATGTGGAAGAGATTTTGTCTAAAGGGCA
GCATTTGTACAAGGAAAAACAGCCACTCAGCCAGTGAAGAGGAAGTTAGAAGA
TCTGAGCTCTGAGTGGAAAGGCGGTAAACCGTTTACTTCAAGAGCTGAGGGCAAAG
CAGCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCTCCTACTCAGACTGT
TACTCTGGTGACACAACCTGTGGTTACTAAGGAACTGCCATCTCCAACTAGAA
ATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCTGGCAGATTTCAACCGGGCTTG
GACAGAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATCACAGAGG
GTGATGGTGGGTGACCTTGAGGATATCAACGAGATGATCATCAAGCAGAAGGCAA

CAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGC
CCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGA
TCGAATTGAAAGAATTCAGAATCAGTGGGATGAAGTACAAGAACACCTTCAGAAC
CGGAGGCAACAGTTGAATGAAATGTTAAAGGATTCAACACAATGGCTGGAAGCTA
AGGAAGAAGCTGAGCAGGTCTTAGGACAGGCCAGAGCCAAGCTTGAGTCATGGA
AGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATCACAGAAACCAAGC
AGTTGGCCAAAGACCTCCGCCAGTGGCAGACAAATGTAGATGTGGCAAATGACTT
GGCCCTGAAACTTCTCCGGGATTATTCTGCAGATGATACCAGAAAAGTCCACATG
ATAACAGAGAATATCAATGCCTCTTGAGAAAGCATTTCATAAAAGGGTGAGTGAGC
GAGAGGCTGCTTTGGAAGAACTCATAGATTACTGCAACAGTTCCCCCTGGACCT
GGAAAAGTTTCTTGCTGGCTTACAGAAGCTGAAACAACCTGCCAATGTCCTACAG
GATGCTACCCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTG
ATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATC
ACAACCTGGATGAAAACAGCCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGA
TGCAGTCCTGTTACAAAGACGTTTGGATAACATGAACTTCAAGTGGAGTGAAGT
CGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTGGA
AGCGTCTGCACCTTTCTCTGCAGGAAGTCTGGTGTGGCTACAGCTGAAAGATGAT
GAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGCAGTTCAGAAGCAGA
ACGATGTACATAGGGCCTTCAAGAGGGAATTGAAAACATAAGAACCTGTAATCAT
GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGACTA
GAGAACTCTACCAGGAGCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCAGAAAT
GTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAAT
TGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCTTGAAAGACT
CCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAG
GTGATCAAGGGATCCTGGCAGCCCCGTGGGCGATCTCCTCATTGACTCTCTCCAAGA
TCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCGCCTCTGAAAGAGAAC
GTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACTTTGGGCATTACAGCTCTC
ACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAG
GTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTC
CAGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATC
TCGCCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAACAACCTTGCTGGG
ACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTCAG
ATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAAGGCCCTTTGCT
TGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGAACAGCACAACTCAAG
CAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTA
TGACCGCCTGGAGCAAGAGCACAAACAATTTGGTCAACGTCCCTCTCTGCGTGGAT
ATGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCC
GTGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGAC
AAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGC
GCAGGCTGGGCCTCCTTCTGCATGATTCTATCCAAATTCCAAGACAGTTGGGTGAA
GTTGCATCCTTTGGGGGCAGTAACATTGAGCCAAGTGTCCGGAGCTGCTTCCAATT
TGCTAATAATAAGCCAGAGATCGAAGCGGCCCTCTTCCTAGACTGGATGAGACTG
GAACCCCAGTCCATGGTGTGGCTGCCCCGTCCTGCACAGAGTGGCTGCTGCAGAAA
CTGCCAAGCATCAGGCCAAATGTAACATCTGCAAAGAGTGTCCAATCATTGGATT
CAGGTACAGGAGTCTAAAGCACTTTAATTATGACATCTGCCAAAGCTGCTTTTTTT
CTGGTCGAGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATATTGCAC

TCCGACTACATCAGGAGAAGATGTTTCGAGACTTTGCCAAGGTAATAAAAAACAAA
TTTCGAACCAAAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTACCTGCCAGTGC
AGACTGTCTTAGAGGGGGACAACATGGAAACTCCCGTTACTCTGATCAACTTCTG
GCCAGTAGATTCTGCGCCTGCCTCGTCCCCTCAGCTTTCACACGATGATACTCATT
CACGCATTGAACATTATGCTAGCAGGCTAGCAGAAATGGAAAACAGCAATGGATC
TTATCTAAATGATAGCATCTCTCCTAATGAGAGCATAGATGATGAACATTTGTAA
TCCAGCATTACTGCCAAAGTTTGAACCAGGACTCCCCCTGAGCCAGCCTCGTAGT
CCTGCCCAGATCTTGATTTCTTAGAGAGTGAGGAAAGAGGGGAGCTAGAGAGAA
TCCTAGCAGATCTTGAGGAAGAAAACAGGAATCTGCAAGCAGAATATGACCGTCT
AAAGCAGCAGCACGAACATAAAGGCCTGTCCCCACTGCCGTCCCCCTCTGAAATG
ATGCCACCTCTCCCCAGAGTCCCCGGGATGCTGAGCTCATTGCTGAGGCCAAGCT
ACTGCGTCAACACAAAGGCCGCCTGGAAGCCAGGATGCAAATCCTGGAAGACCA
CAATAAACAGCTGGAGTCACAGTTACACAGGCTAAGGCAGCTGCTGGAGCAACCC
CAGGCAGAGGCCAAAGTGAATGGCACAACGGTGTCTCTCTTCTACCTCTCTAC
AGAGGTCCGACAGCAGTCAGCCTATGCTGCTCCGAGTGGTTGGCAGTCAAACCTC
GGACTCCATGGGTGAGGAAGATCTTCTCAGTCCTCCCCAGGACACAAGCACAGGG
TTAGAGGAGGTGATGGAGCAACTCAACAACCTCTCCCTAGTTCAAGAGGAAGAA
ATACCCTTGGAAGCCAATGAGAGAGGACACAATGTAGGAAGTCTTTTCCACATG
GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAG
AAGGAGCAGAATAAATGTTTTACAACCTCTGATTCCCGCATGGTTTTTATAATATT
CATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTT
TGTGAAGGGTAGTGGTATTATACTGTAGATTTTCAGTAGTTTCTAAGTCTGTTATTG
TTTTGTAAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAA
GAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAA
CGCATTTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAA
AGTGTGCTTTATAAAAAAAGTTGTTTATAAAAAACCCCTAAAAACAAAACAAACA
CACACACACACATACACACACACACACAAAACCTTTGAGGCAGCGCATTGTTTT
GCATCCTTTTGGCGTGATATCCATATGAAATTCATGGCTTTTTCTTTTTTGCATAT
TAAAGATAAGACTTCCTCTACCACCACACCAATGACTACTACACACTGCTCATT
GAGAACTGTCAGCTGAGTGGGGCAGGCTTGAGTTTTCATTTTCATATATCTATATGT
CTATAAGTATATAAATACTATAGTTATATAGATAAAGAGATACGAATTTCTATAG
ACTGACTTTTTTCCATTTTTTAAATGTTTCATGTCACATCCTAATAGAAAGAAATTACT
TCTAGTCAGTCATCCAGGCTTACCTGCTTGGTCTAGAATGGATTTTTCCCGGAGCC
GGAAGCCAGGAGGAACTACACCACACTAAACATTGTCTACAGCTCCAGATGTT
TCTCATTTTAAACAACTTTCCACTGACAACGAAAGTAAAGTAAAGTATTGGATTTT
TTTAAAGGGAACATGTGAATGAATACACAGGACTTATTATATCAGAGTGAGTAAT
CGGTTGGTTGGTTGATTGATTGATTGATTGATACATTCAGCTTCCTGCTGCTAGCA
ATGCCACGATTTAGATTTAATGATGCTTCAGTGGAATCAATCAGAAGGTATTCTG
ACCTTGTAACATCAGAAGGTATTTTTTAACTCCCAAGCAGTAGCAGGACGATGA
TAGGGCTGGAGGGCTATGGATTCCCAGCCCATCCCTGTGAAGGAGTAGGCCACTC
TTAAGTGAAGGATTGGATGATTGTTTATAATACATAAAGTTCTCTGTAATTACAA
CTAAATTATTATGCCCTCTTCTCACAGTCAAAAGGAACTGGGTGGTTTTGGTTTTTG
TTGCTTTTTTAGATTTATTGTCCCATGTGGGATGAGTTTTTAAATGCCACAAGACAT
AATTTAAATAAATAAATTTGGGAAAAGGTGTAAGACAGTAGCCCCATCACATT
TGTGATACTGACAGGTATCAACCCAGAAGCCCATGAACTGTGTTTTCCATCCTTGC
ATTTCTCTGCGAGTAGTTCCACACAGGTTTGTAAGTAAGTAAGAAAGAAGGCAAA

```

TTGATTCAAATGTTACAAAAAACCCCTTCTTGGTGGATTAGACAGGTAAATATAT
AAACAAACAAACAAAAATTGCTCAAAAAAGAGGAGAAAAAGCTCAAGAGGAAAA
GCTAAGGACTGGTAGGAAAAAGCTTTACTCTTTCATGCCATTTTATTTCTTTTGGAT
TTTTAAATCATTCAATTCAATAGATACCACCGTGTGACCTATAATTTTGCAAATCTG
TTACCTCTGACATCAAGTGTAATTAGCTTTTGGAGAGTGGGCTGACATCAAGTGTA
ATTAGCTTTTGGAGAGTGGGTTTTGTCCATTATTAATAATTAATTAATTAACATCA
AACACGGCTTCTCATGCTATTTCTACCTCACTTTGGTTTTGGGGTGTTCCTGATAAT
TGTGCACACCTGAGTTCACAGCTTCACCACTTGTCCATTGCGTTATTTTCTTTTTCC
TTTATAATTCTTTCTTTTTCTTCATAATTTTCAAAGAAAACCCAAAGCTCTAAGG
TAACAAATTACCAAATTACATGAAGATTTGGTTTTTGTCTTGCAATTTTTTTCCTTTA
TGTGACGCTGGACCTTTTCTTTACCCAAGGATTTTAAACTCAGATTAAAAACAA
GGGGTTACTTTACATCCTACTAAGAAGTTTAAGTAAGTAAGTTTCATTCTAAAATC
AGAGGTAAATAGAGTGCATAAATAATTTGTTTTAATCTTTTGTTTTCTTTTAGA
CACATTAGCTCTGGAGTGAGTCTGTCATAATTTGAACAAAAATTGAGAGCTTTA
TTGCTGCATTTTAAGCATAATTAATTTGGACATTATTTCTGTGTGTGTCTTTATAA
CCACCGAGTATTAACTGTAAATCATAATGTAAGTGAAGCATAAACATCACATGG
CATGTTTTGTGCTATTGTTTTCAGGTACTGAGTTCTTACTTGAGTATCATAATATATTG
TGTTTTAACACCAACACTGTAACATTACGAATTATTTTTTAACTTCAGTTTTAC
TGCATTTTCACAACATATCAGACTTCACCAAATATATGCCTTACTATTGTATTATA
GTACTGCTTTACTGTGTATCTCAATAAAGCACGCAGTTATGTTAC

```

The wild type, full length dystrophin protein (isoform Dp427m) contains 24 spectrin like repeats, at least four hinge regions, actin binding domain (ABD1), Cysteine rich domain (CR), and C terminal Domain (C-term.). The polypeptide sequence of each domain is shown in Table 3, and the nucleotide sequence of each domain is shown in Table 4.

Table 3. Amino Acid Sequences of Dystrophin Domains

Description and Sequence Identifier	Sequence
ABD1 (SEQ ID NO:3)	MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHLENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKIL LSWVRQSTRNYPQVNVINFTTSWSDGLALNALIHSRPLDFDWSNVVC QQSATQRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQV LPQQVSIEAIQEVE
Hinge 1(SEQ ID NO:4)	MLPRPPKVTKEEHFQLHHQMHSYQQITVSLAQGYERTSSPKPRFKSYAY TQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSLMES
Spectrin repeat 1 (SEQ ID NO:5)	EVNLDTRYQTALIEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGY MMDLTAHQGRVGNILQLGSKLIGTGKLSEDEETEVQEQMNLLNSRWEC LRVASMEKQSNLH

Spectrin repeat 2 (SEQ ID NO:6)	RVLMDLQNQKLKELNDWLTKEERTRKMEEEEPLGPDLEDLKRQVQQH KVLQEDLEQEQRVNSLTHMVVVVDESSGDHATAALEEQLKVLGDRW ANICRWTEDRWVLLQDI
Spectrin repeat 3 (SEQ ID NO:7)	LLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAV LKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLDNFARCWD NLVQKLEKSTAQISQA
Hinge 2 (SEQ ID NO:8)	VTQTQPSLTQTVMETVTTVTTREQILVKHAQEELPPPPQKKRQITVD
Spectrin repeat 4 (SEQ ID NO:9)	SEIRKRLDVDITELHSWITRSEAVLQSPEFAIFRKEGNFSDLKEKVNAIER EKA EKFRKLQDASRSAQALVEQMVNEGVNADSIKQASEQLNSRWIEFC QLLSERLNLWLEY
Spectrin repeat 5 (SEQ ID NO:10)	QNNIAFYNNQLQQLEQMTTAEENWLKIQPTTPSEPTAIKSQLKICKDEVN RLSGLQPQIERLKIQSIALKEKGQGMFLDADFVAFTNHFKQVFSDVQA REKELQTIFD
Spectrin repeat 6 (SEQ ID NO:11)	TLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQALQ SSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWKKLSSQL VEHCQKLEEQ
Spectrin repeat 7 (SEQ ID NO:12)	MNKLRLKIQNHQTLKKWMAEVDVFLKEEWPALGDSEILKKQLKQCRL VSDIQTIPSLNSVNEGGQKIKNEAEPEFASRLETTELKELNTQWDHMCQ QVYARKEALKGG
Spectrin repeat 8 (SEQ ID NO:13)	LEKTVSLQKDLSEMHEWMTQAE E EYLERDFEYKTPDELQKAVEEMKR AKEEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELETLTNYQWL CTRLNGKCKTLEEV
Spectrin repeat 9 (SEQ ID NO:14)	WACWHELLSYLEKANKWLNEVEFKLKTENIPGGAE EISEVLDLENL MRHSEDNPNQIRILAQTLTDGGVMDELINEELETFNRSRWRELHEEA VRR QKLLEQS
Spectrin repeat 10 (SEQ ID NO:15)	IQSAQETEKSLHLIQESLTFIDKQLAAYIADKVDA AQMPQEAQKIQSDLT SHEISLEEMKKHNQGKEAAQRVLSQIDVAQKKLQDVSMKFRL
Spectrin repeat 11 (SEQ ID NO:16)	FQKPANFEQRLQESKMILDEVKMHLPALETKSVEQEVVQSQLNHC VNL YKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLHYNELG AKVTERKQQLEKC
Spectrin repeat 12 (SEQ ID NO:17)	LKLSRKMRKEMNVLTEWLAATDMELTKRS AVEGMPSNLDSEVAWGK ATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDKLSLLNSNWIAVT SRAE EWLNLLEY
Spectrin repeat 13 (SEQ ID NO:18)	QKHMETFDQNVDHITKWIIQADTLLDESEKKKPPQKEDVLKRLKAELN DIRPKVDSTRDQAANLMANRGDHCRKLVEPQISELNHRFAAISHRIKTG KASIPLK
Spectrin repeat 14 (SEQ ID NO:19)	ELEQFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQR GDNLQQRITDERKREEIKIKQQLLQTKHNALKDLRSQRRKKALEI

Spectrin repeat (SEQ NO:20)	15 ID	SHQWYQYKRQADDLLKCLDDIEKKLASLPEPRDERKIKEIDRELQKKKE ELNAVRRQAEGLSEDGAAMAVEPTQIQLSKRWREIESKFAQFRRNLFA Q
L3 (20-mer linker) (SEQ ID NO:21)		IHTVREETMMVMTEDMPLEI
Spectrin repeat (SEQ NO:22)	16 ID	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEESLKNIKD SLQQSSGRIDIIHKKTAALQSATPVERVKLQEALSQLDFQWEKVNKMY KDRQGRFDRS
Spectrin repeat (SEQ NO:23)	17 ID	VEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYKWYKELQ DGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLNRWQEVCKQ LSDRKKRLEEQ
Spectrin repeat (SEQ NO:24)	18 ID	KNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQKLEQVKLLVEE LPLRQGILKQLNETGGPVLVSAPISPEEQDKLENKQTNLQWIKVSRA LPEKQGEIEAQIKDLGQL
Spectrin repeat (SEQ NO:25)	19 ID	EKKLEDLEEQLNHLLWLSPIRNQLEIYNQPNQEGPFDVQETEIIVQAK QPDVEEILSKGQHLYKEKPATQPVKRKLEDLSSEWKA VNRLQLRAK QPD
Hinge 3 (SEQ ID NO:26)		APGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLE
Spectrin repeat (SEQ NO:27)	20 ID	VPALADFNRAWTELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKAT MQDLEQRRPQLEELITAAQNLKNKTSNQEARTIITDRIERIQNQWDEVQ EHLQNRQQQLNEM
Spectrin repeat (SEQ NO:28)	21 ID	LKDSTQWLEAKEEAEQVLGQARAKLESWKEGPYTVDAIQKKITETKQL AKDLRQWQTNVDVANDLALKLLRDYSADDTRKVHMITENINASWRSI HKRVSEREAALEET
Spectrin repeat (SEQ NO:29)	22 ID	HRLLQQFPLDLEKFLAWL TEAETTANVLQDATRKERLLED SKGVKELM KQWQDLQGEIEAHTDVYHNL DENSQKILRSLEGSDDAVLLQRRLDNM NFKWSELRKKSLNIRSHLEAS
Spectrin repeat (SEQ NO:30)	23 ID	SDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPAVQKQNDVHRA FKRELKTKEPVMSTLETVRIFL TEQPLEGLEKLYQEPREL PPEERAQNV TRLRLKQAEVNTWEKLNLSADWQRKIDET
Spectrin repeat (SEQ NO:31)	24 ID	LERLQELQEATDELDLKL RQAEVIKGSWQPVGDLLIDSLQDHLEK VKA LRGEIAPLKENVSHVNDLARQL TTLGIQLSPYNLSTLEDLNTRWKLLQV AVEDRVRQLHE

Hinge 4 (SEQ ID NO:32)	AHRDFGPASQHFLLSTSVQGPWERAISPKNVPYYINHETQTTCDWHPKM TELYQSLADLNNVRFSA YRTAMKL
CR (SEQ ID NO:33)	RRLQKALCLDLLSLSAACDALDQHNLKQNDQPM DILQINCLTTIYDRL EQEHNNLVNVPLCVDMLNWLNVYDTGRTGRIRVLSFKTGIISLCKA HLEDKYRYLQVASTGFCDQRRLLGLLHDSIQIPRQLGEVASFGGSNI EPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETA KHQAKCNICKECPIIGFRYRSLKHFN YDICQSCFFSGRVAKGHKMHYPM VEYCTPTTSGEDVRDFAKVLKNKFR TKRYFAKHPRMGYLPVQTVLEGD NMET
C-term (SEQ ID NO:34)	PVTILNFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLND SISPNESIDDEHLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGERIL ADLEENRNQLAEYDRLKQQHEHKGLSPLSPPEMMPTSPQSPRDAELI AEAKLLRQHKGRLEARMQILEDHNKQLESQHLRLRQLLEQPQAEAKVN GTTVSSPSTSLQRS DSSQPMLLRVVGSTSDSMGEEDLLSPPQDTSTGLE EVMEQLNNSFPSSRGRNTPGKPMREDTM

Table 4. Nucleotide Sequences Encoding Dystrophin Domains

Description and Sequence Identifier	Sequence
5' untranslated region (SEQ ID NO:35)	gggatccct cacttcccc ctacaggact cagatctggg aggcaattac ctccggagaa 60 aaacgaatag gaaaaactga agtggtactt ttttaaagc tgctgaagt tgggtgttc120 tcattgttt taagcctact ggagcaataa agtttgaaga acttttacc ggtttttt180 atcgctgcct tgatacac tttcaaa 208
ABD1 (SEQ ID NO:36)	atgctttggt gggaagaagt agaggactgt tatgaagag aagatgttca aaagaaaaca 60 ttcacaaaat gggtaaatgc acaattttc aagtttgga agcagcatat tgagaacctc120 ttcagtgacc tacaggatgg gaggcgcctc ctagacctcc tgaaggcct gacagggcaa180 aaactgcaa aagaaaagg atccacaaga gttcatgccc tgaacaatgt caacaaggca240 ctgcgggtt tgcagaacaa taatgttgat ttagtgaata ttggaagtac tgacatcgta300 gatggaaatc ataaactgac tcttggttg atttggaata taatcctcca ctggcaggtc360 aaaaatgtaa tgaataat catggctgga ttgcaacaaa ccaacagtga aaagatttc420 ctgagctggg tccgacaatc aactcgta taccacagg ttaatgta tcaactcacc480 accagctggt ctgatggcct ggcttgaat gctctcatc atagtcata gccagacctc540 ttgactgga atagtgtggt ttgccagcag tcagccacac aacgactgga acatgcattc600 aacatcgcca gatatcaatt aggcataag aaactactcg atcctgaaga tgttgatacc660 acatatcag ataagaagtc catctaatg tacatcatc cactttcca agttttgcct720 caacaagtga gcattgaagc catccaggaa gtggaa756
Hinge 1 (SEQ ID NO:37)	atgttgcaa ggccaccta agtgactaaa gaagaacatt ttcagtaca tcatcaaatg 60 cactatttc aacagatcac ggtcagtcta gcacaggat atgagagaac ttctccct120 aagcctgat tcaagagcta tgcctacaca caggctgctt atgtaccac ctctgacct180 acacggagcc catttcctc acagcattg gaagctcctg aagacaagtc attggcagt240 tcattgatg agagt 255

Spectrin repeat 1 (SEQ ID NO:38)	gaagtaaacc tggaccgtta tcaaacagct ttagaagaag tattatcgtg gcttcttct 60 gctgaggaca cattgcaagc acaaggagag atttctaattg atgtggaagt ggtgaaagac 120 cagtttcata ctcatgaggg gtacatgatg gatttgacag cccatcaggg ccgggttggt 180 aatattctac aattgggaag taagctgatt ggaacaggaa aattatcaga agatgaagaa 240 actgaagtac aagagcagat gaatctccta aattcaagat gggaaatgcct cagggtagct 300 agcatggaaa aacaagcaa ttacat 327
Spectrin repeat 2 (SEQ ID NO:39)	agagttttaa tggatctcca gaatcagaaa ctgaaagagt tgaatgactg gctaacaaaa 60 acagaagaaa gaacaaggaa aatggaggaa gagcctcttg gacctgatct tgaagacct 120 aaacgccaag tacaacaaca taaggtgctt caagaagatc tagaacaaga acaagtcagg 180 gtcaattctc tcaactcacat ggtgggtgta gttgatgaat ctagtggaga tcacgcaact 240 gctgctttgg aagaacaact taaggtattg ggagatcgat gggcaaacat ctgtagatgg 300 acagaagacc gctgggttct ttacaagac atc 333
Spectrin repeat 3 (SEQ ID NO:40)	cttctcaaat ggcaacgtct tactgaagaa cagtgccttt ttagtgcagt gcttcagaa 60 aaagaagatg cagtgaacaa gattcacaca actggcttta aagatcaaaa tgaatgtta 120 tcaagtcttc aaaaactggc cgttttaaaa gcgcatctag aaaagaaaaa gcaatccatg 180 ggcaaacgtg attcactcaa acaagatctt cttcaacac tgaagaataa gtcagtgaac 240 cagaagacgg aagcatggct ggataacttt gcccggtgtt gggataattt agtcaaaaaa 300 cttgaagaaga gtacagcaca gatttcacag gct 333
Hinge (SEQ ID NO:41)	gtcaccacca ctacgccatc actaacacag acaactgtaa tggaaacagt aactacggtg 60 accacaaggg aacagatcct ggtaaagcat gctcaagagg aacttcacc accacctccc 120 caaaagaaga ggcagattac tgtggt 147
Spectrin repeat 4 (SEQ ID NO:42)	tctgaaatta ggaaaagggt ggatgttgat ataactgaac ttcacagctg gattactcgc 60 tcagaagctg tgttcagag tctgaattt gcaatcttc ggaaggagg caacttctca 120 gacttaaaag aaaaagtcaa tgccatagag cgagaaaaag ctgagaagt cagaaaactg 180 caagatgccg gcagatcagc tcaggccctg gtggaacaga tggatgaatg ggggtgta 240 gcagatagca tcaacaagc ctcaacaaca ctgaacagcc ggtggatcga attctgccag 300 ttgctaagtg agagactta ctggctggag tat 333
Spectrin repeat 5 (SEQ ID NO:43)	cagaacaaca tcacgcctt ctataatcag ctacaacaat tggagcagat gacaactact 60 gctgaaaact ggttgaaaat ccaacccacc acccatcag agccaacagc aattaaaagt 120 cagttaaaaa ttgttaagga tgaagtcaac cggctatcag gtcttcaacc tcaattgaa 180 cgattaaaaa ttcaaagcat agccctgaaa gagaaaggac aaggacccat gtctctggat 240 gcagactttg tggcctttac aaatcattt aagcaagtct ttctgatgt gcaggccaga 300 gagaaagagc tacagacaat tttgac 327
Spectrin repeat 6 (SEQ ID NO:44)	actttgccac caatgcgcta tcaggagacc atgagtcca tcaggacatg ggtccagcag 60 tcagaaacca aactctccat acctcaactt agtgcaccg actatgaaat catggagcag 120 agactcgggg aattgcaggc ttacaaagt tctctgcaag agcaacaag tggcctatac 180 tatctcagca ccactgtgaa agagatgtcg aagaaagcgc cctctgaaat tagccggaaa 240 tatcaatcag aatttgaaga aattgaggga cgctggaaga agctctctc ccagctggtt 300 gagcattgtc aaaagctaga ggagcaa 327

Spectrin repeat 7 (SEQ ID NO:45)	atgaataaac tccgaaaaat tcagaatcac atacaaaccc tgaagaaatg gatggctgaa 60 gttgatgttt ttctgaagga ggaatggcct gcccttgggg attcagaaat tctaaaaaag120 cagctgaaac agtgcagact ttagtcagt gatattcaga caattcagcc cagtctaaac180 agtgtcaatg aaggtgggca gaagataaag aatgaagcag agccagagtt tgctcgaga240 cttgagacag aactcaaaga acttaacact cagtgggac acatgtgcca acaggtctat300 gccagaaagg aggccttgaa gggaggt327
Spectrin repeat 8 (SEQ ID NO:46)	ttggagaaaa ctgtaagcct ccagaaagat ctatcagaga tgcacgaatg gatgacacaa 60 gctgaagaag agtatcttga gagagatttt gaatataaaa ctccagatga attacagaaa120 gcagtgaag agatgaagag agctaagaa gaggcccaac aaaaagaagc gaaagtgaaa180 ctccttactg agtctgtaa tagtgtcata gctcaagtc cacctgtagc acaagaggcc240 ttaaaaaagg aactgaac tctaaccacc aactaccagt ggctctgcac taggctgaat300 gggaaatgca agactttgga agaagtt327
Spectrin repeat 9 (SEQ ID NO:47)	tgggcatgtt ggcatgagtt attgtcatac ttggagaaa caaacaagtg gctaaatgaa 60 gtagaattta aactaaaaac cactgaaaac attcctggcg gagctgagga aatctctgag120 gtgctagatt cactgaaaa ttgatgcga cattcagagg ataaccaaa tcagattcgc180 atattggcac agaccctaac agatggcgga gtcattgatg agctaataca tgaagaaact240 gagacattta attctgttg gaggaacta catgaagagg ctgtaaggag gcaaaagt300 cttgaacaga gc312
Spectrin repeat 10 (SEQ ID NO:48)	atccagtctg cccaggagac tgaaaaatcc ttacacttaa tccaggagtc cctcacattc 60 attgacaagc agttggcagc ttatattgca gacaagggtg acgcagctca aatgcctcag120 gaagcccaga aaatccaatc tgatttgaca agtcattgaga tcagtttaga agaaatgaag180 aaacataatc aggggaagga ggctgccaa agagtctctg ctgagattga tgttcacag240 aaaaaattac aagatgtctc catgaagtt cgatta276
Spectrin repeat 11 (SEQ ID NO:49)	ttccagaaac cagccaattt tgagcagcgt ctacaagaaa gtaagatgat tttagatgaa 60 gtgaagatgc acttgccctgc attggaaca aagagtgtgg aacaggaagt agtacagtca120 cagctaaatc attgtgtgaa ctgtataaa agtctgagtg aagtgaagtc tgaagtggaa180 atggtgataa agactggacg tcagattgta cagaaaaagc agacggaaaa tcccaagaa240 cttgatgaaa gagtaacagc ttgaaattg cattataatg agctgggagc aaaggtaa300 gaaagaaagc aacagttgga gaaatgc327
Spectrin repeat 12 (SEQ ID NO:50)	ttgaaattgt cccgtaagat gcgaaaggaa atgaatgtct tgacagaatg gctggcagct 60 acagatatgg aattgacaaa gagatcagca gttgaaggaa tgcctagtaa ttggattct120 gaagttgcct ggggaaaggc tactcaaaaa gagattgaga aacagaaggt gcacctgaag180 agtatcacag aggtaggaga ggccttgaaa acagttttgg gcaagaagga gacgttggt240 gaagataaac tcagtcttct gaatagtaac tggatagctg tcacctccc agcagaagag300 tggttaaact tttgttga atac 324
Spectrin repeat 13 (SEQ ID NO:51)	cagaaacaca tggaaacttt tgaccagaat gtggaccaca tcacaaagt gatcattcag 60 gctgacacac ttttgatga atcagagaaa aagaaacccc agcaaaaaga agacgtgct120 aagcgttaa aggcagaact gaatgacata cgcccaaagg tggactctac acgtgacaa180 gcagcaaact tgatggcaaa ccgcgtgac cactgcagga aattagtaga gcccaaatc240 tcagagctca accatcgatt tgcagccatt tcacacagaa ttaagactgg aaaggcctcc300

	attcctttga ag312
Spectrin repeat 14 (SEQ ID NO:52)	gaattggagc agtttaactc agatatacaa aaattgcttg aaccactgga ggctgaaatt 60 cagcagggggg tgaatctgaa agaggaagac ttcaataaag atatgaatga agacaatgag120 ggtactgtaa aagaattgtt gcaaagagga gacaacttac aacaaagaat cacagatgag180 agaaagagag aggaataaaa gataaacag cagctgttac agacaaaaca taatgtctc240 aaggatttga ggtctcaaag aagaaaaaag gctctagaaa tt 282
Spectrin repeat 15 (SEQ ID NO:53)	tctcatcagt ggtatcagta caagaggcag gctgatgac tctgaaatg cttggatgac 60 attgaaaaaa aattagccag cctacctgag cccagagatg aaaggaaaat aaaggaaatt120 gatcgggaat tgcagaagaa gaaagaggag ctgaatgcag tgcgtaggca agctgagggc180 ttgtctgagg atggggccgc aatggcagtg gagccaactc agatccagct cagcaagcgc240 tggcgggaaa ttgagagcaa attgtctcag ttctgaagac tcaactttgc acaa294
L3 (20-mer linker)(SEQ ID NO:54)	attcacactg tccgtgaaga aacgatgatg gtgatgactg aagacatgcc ttggaaatt 60
Spectrin repeat 16 (SEQ ID NO:55)	tcttatgtgc cttctactta ttgactgaa atcactcatg tctcacaagc cctattagaa 60 gtggaacaac ttctaatgc tctgacctc tgtgctaagg actttgaaga tctctttaag120 caagaggagt ctctgaagaa tataaaagat agtctacaac aaagctcagg tcggattgac180 attattcata gcaagaagac agcagcattg caaagtgcaa cgctgtgga aagggtgaag240 ctacaggaag ctctctccca gcttgatttc caatgggaaa aagttaaca aatgtacaag300 gaccgacaag ggcgatttga cagatct327
Spectrin repeat 17 (SEQ ID NO:56)	gttgagaaat ggcggcgttt tcattatgat ataaagatat ttaatcagtg gctaacagaa 60 gctgaacagt ttctcagaaa gacacaaatt cctgagaatt gggaacatgc taaatacaaa120 tggtatctta aggaactcca ggtggcatt gggcagcggc aaactgttgt cagaacattg180 aatgcaactg gggaagaaat aattcagcaa tctcaaaaa cagatgccag tattctacag240 gaaaaattgg gaagcctgaa tctgcggtgg caggaggtct gcaaacagct gtcagacaga300 aaaaagaggc tagaagaaca a321
Spectrin repeat 18 (SEQ ID NO:57)	aagaatatct tgtcagaatt tcaagagat taaatgaat ttgtttatg gttggaggaa 60 gcagataaca ttgctagtat ccacttgaa cctggaaaag agcagcaact aaaagaaaag120 cttgagcaag tcaagttact ggtggaagag ttgccctgc gccagggaat tctcaacaa180 ttaaataaaa ctggaggacc cgtgcttga agtgctcca taagcccaga agagcaagat240 aaactgaaa ataagctcaa gcagacaaat ctccagtgga taaaggttc cagagcttta300 cctgagaaac aaggagaaat tgaagctcaa ataaaagacc ttgggcagct t 351
Spectrin repeat 19 (SEQ ID NO:58)	gaaaaaagc ttgaagacct tgaagagcag taaatcatc tgctgctgtg gttatctct 60 attaggaatc agttggaaat ttataacca ccaaccaag aaggaccatt tgacgttcag120 gaaactgaaa tagcagtcca agctaaacaa ccgatgtgg aagagatttt gtctaaaggg180 cagcatttgt acaaggaaaa accagccact cagccagtga agaggaagtt agaagatctg240 agctctgagt ggaaggcggg aaaccgttta ctcaagagc tgagggcaaa gcagcctgac300 cta303

Hinge 3 (SEQ ID NO:59)	gctcctggac tgaccactat tggagcctct cctactcaga ctgttactct ggtgacacaa 60 cctgtgggta ctaaggaaac tgccatctcc aaactagaaa tgccatcttc cttgatgttg120 gag123
Spectrin repeat 20 (SEQ ID NO:60)	gtacctgctc tggcagattt caaccgggct tggacagaac ttaccgactg gctttctctg 60 cttgatcaag ttataaaatc acagaggggtg atgggtgggtg accttgagga tatcaacgag120 atgatcatca agcagaaggc aacaatgcag gatttggaac agaggcgtcc ccagttggaa180 gaactcatta ccgctgcccc aaatttgaaa aacaagacca gcaatcaaga ggctagaaca240 atcattacgg atcgaattga aagaattcag aatcagtggg atgaagtaca agaacacct300 cagaaccgga ggcaacagtt gaatgaaatg 330
Spectrin repeat 21 (SEQ ID NO:61)	ttaaaggatt caacacaatg gctggaagct aaggaagaag ctgagcaggt cttaggacag 60 gccagagcca agcttgagtc atggaaggag ggtccctata cagtagatgc aatccaaaag120 aaaatcacag aaaccaagca gttggccaaa gacctccgcc agtggcagac aaatgtagat180 gtggcaaatg acttggcctt gaaactctc cgggattatt ctgcagatga taccagaaaa240 gtccacatga taacagagaa tatcaatgcc tcttgagaa gcatcataa aagggtgagt300 gagcgagagg ctgcttggga agaaact327
Spectrin repeat 22 (SEQ ID NO:62)	catagattac tgcaacagtt cccctggac ctggaagaat ttctgcctg gcttacagaa 60 gctgaaacaa ctgccaatgt cctacaggat gctaccgta aggaaggct cctagaagac120 tccaaggag taaaagagct gatgaaacaa tggcaagacc tccaagtgta aattgaagct180 cacacagatg ttatcacaa cctggatgaa aacagccaaa aaatcctgag atccctggaa240 ggttccgatg atgcagtcct gttacaaaga cgttggata acatgaact caagtggagt300 gaactcga aaaagtctct caacattagg tccatttgg aagccagt 348
Spectrin repeat 23 (SEQ ID NO:63)	tctgaccagt ggaagcgtct gcaccttct ctgcaggaa ttctggtgtg gctacagctg 60 aaagatgatg aattaagccg gcaggcacct attggaggcg acttccagc agttcagaag120 cagaacgatg tacatagggc cttcaagagg gaattgaaaa ctaaaagaacc tgtaatcatg180 agtactcttg agactgtacg aatattctg acagagcagc ctttggagg actagagaaa240 ctctaccagg agcccagaga gctgcctctt gaggagagag cccagaatgt cactcggtt300 ctacgaaagc aggtgagga ggtcaatact gagtgggaaa aattgaacct gactccgct360 gactggcaga gaaaaataga tgagacc387
Spectrin repeat 24 (SEQ ID NO:64)	cttgaaagac tccaggaact tcaagaggcc acggatgagc tggacctcaa gctgcgcca 60 gctgaggtga tcaagggtat ctggcagccc gtggcgatc tctcattga ctcttccaa120 gatcacctcg agaaagtcaa ggcacttca ggagaaattg cgcctctgaa agagaacgtg180 agccacgtca atgaccttgc tgcagctt accacttgg gcattcagct ctaccgtat240 aacctcagca ctctggaaga cctgaacacc agatggaagc ttctgcaggt ggccgtcgag300 gaccgagtca ggcagctgca tga 324
Hinge 4 (SEQ ID NO:65)	gcccacaggg acttgggtcc agcatctcag cactttctt ccacgtctgt ccagggtccc 60 tgggagagag ccatctcgcc aaacaaagt cctactata tcaaccacga gactcaaca120 acttgcctgg accatcccaa aatgacagag ctctaccagt ctttagctga cctgaataat180 gtcagattct cagcttatag gactgcatg aaactc216

CR (SEQ ID NO:66)	cgaagactgc agaaggccct ttgcttggat ctcttgagcc tgtcagctgc atgtgatgcc 60 ttggaccagc acaacctcaa gcaaaatgac cagcccatgg atatcctgca gattattaat120 tgtttgacca ctattatga ccgcctggag caagagcaca acaatttggc caacgtccct180 ctctgcgtgg atatgtgtct gaactggctg ctgaatgttt atgatacggg acgaacaggg240 aggatccgtg tcctgtcttt taaaactggc atcattccc tgtgtaaagc acatttgaa300 gacaagtaca gatacctttt caagcaagtg gcaagttaa caggattttg tgaccagcgc360 aggctggggc tccttctgca tgattctatc caaattcaa gacagttggg tgaagttgca420 tcctttgggg gcagtaacat tgagccaagt gtccggagct gctccaatt tgctaataat480 aagccagaga tcgaagcggc cctctccta gactggatga gactggaacc ccagtccatg540 gtgtggctgc ccgtcctgca cagagtggct gctgcagaaa ctgccaagca tcaggccaaa600 tgtaacatct gcaaagagtg tccaatcatt ggattcaggt acaggagtct aaagcacttt660 aattatgaca tctgcaaaag ctgcttttt tctggtcgag ttgcaaaagg ccataaaatg720 cactatccca tgggtggaata ttgactccg actacatcag gagaagatgt tcgagacttt780 gccaaggtac taaaaaaca attcgaacc aaaaggtatt ttgcgaagca tccccgaatg840 ggctacctgc cagtgcagac tgtcttagag ggggacaaca tggaaact
C-term (SEQ ID NO:67)	cccgttactc tgatcaactt ctggccagta gattctgcgc ctgctcgtc cctcagctt 60 tcacacgatg atactcattc acgcattgaa cattatgcta gcaggctagc agaaatggaa120 aacagcaatg gatcttatct aaatgatagc atctctccta atgagagcat agatgatgaa180 catttgttaa tccagcatta ctgcaaaagt tgaaccagg actccccct gagccagcct240 cgtagtctcg ccagatctt gatttccta gagagtgagg aaagagggga gctagagaga300 atcctagcag atcttgagga agaaaacagg aatctgcaag cagaatatga ccgtctaaag360 cagcagcacg aacataaagg cctgtcccca ctgccgtccc ctctgaaat gatgccacc420 tctcccaga gtccccggga tctgagctc attgctgagg ccaagctact gcgtcaacac480 aaaggccgcc tggagccag gatgcaaac ctggaagacc acaataaaca gctggagtca540 cagttacaca ggctaaggca gctgctggag caaccccagg cagaggccaa agtgaatggc600 acaacgggtg cctctcttc tacctctcta cagaggtccg acagcagta gcctatgctg660 ctccgagtgg ttggcagta aactcggac tccatgggtg aggaagatct tctcagtcct720 ccccaggaca caagcacagg gttagaggag gtgatggagc aactcaaca ctcttcctct780 agtcaagag gaagaaatac cctggaaag ccaatgagag aggacacaat gtag

The present disclosure is directed to a miniaturized dystrophin polypeptide that is smaller than the full-length dystrophin protein, i.e., isoform Dp427m, and that is not identical to the naturally occurring dystrophin protein isoforms, or a nucleic acid molecule comprising a nucleotide sequence encoding the miniaturized dystrophin polypeptide. When the present disclosure discloses miniaturized dystrophin polypeptides, the present disclosure also discloses nucleic acid molecule comprising a nucleotide sequence encoding the corresponding disclosed miniaturized dystrophin polypeptide, and vice versa. In some embodiments, the nucleic acid molecule encoding the miniaturized dystrophin polypeptide is suitable for gene therapy. Accordingly, the nucleic acid molecule encoding the miniaturized dystrophin polypeptide is constructed not only to fit into a gene therapy vector, e.g., AAV vector, or to be suitable for recombinant expression, but also to reduce

any unwanted immune response (e.g., humoral immune response and/or cellular immune response, e.g., CD4 and/or CD8) against the miniaturized dystrophin polypeptide when administered or expressed *in vivo*.

In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises a junction N-terminal to a unmodified or modified spectrin repeat 16 (R16) domain that varies from the wild-type junction. In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises a modified spectrin repeat 16 (R16) domain, wherein a part of spectrin repeat 16 (R16) domain is replaced by a corresponding part of a different spectrin repeat domain. In some embodiments, the different spectrin repeat domain is spectrin repeat 2 (R2) domain. In some embodiments, the modified R16 domain comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to a sequence selected from the group consisting of SEQ ID NO: 68, 69, 70, 71 and 72. The term junction J4 (or J4 junction), as used herein, refers to the peptide sequence surrounding the junction between spectrin repeat 1 (R1) domain and spectrin repeat 16 (R16) domain. The variants of junction J4 disclosed herein, i.e., J4V4, J4V11, J4V12 and J4V13, are J4 junctions in which, to different degrees, the N-terminal part of spectrin repeat 16 (R16) domain has been replaced by certain N-terminal parts of spectrin repeat 2 (R2).

20

Table 5. Partial amino acid sequence of modified R16 domains/Junction J4 Variants.

SEQ ID NO:	Description	Sequence
68	Modified Spectrin-16 (junction J4V13)	MDLQNQKLTEITHVSQ
69	Modified Spectrin-16 (junction J4V12)	LMDLQNQKTEITHVSQ
70	Modified	LMDLQNQKEITHVSQA

	Spectrin-16 (junction J4V11)	
71	Modified Spectrin-16 (junction J4V4)	LHRVLMDLTYLTEITH
72	Modified Spectrin-16 (junction J4)	MEKQSNLHSYVPSTYL

In some embodiments, the miniaturized dystrophin polypeptide comprises from N terminus to C terminus a hinge 1 (H1) domain, a spectrin repeat 1 (R1) domain, the modified R16 domain, a spectrin repeat 17 (R17) domain, a hinge 3 (H3) domain, a spectrin repeat 23 (R23) domain, a spectrin repeat 24 (R24) domain, and a hinge 4 (H4) domain of dystrophin. In some embodiments, (i) the H1 domain and the R1 domain are fused directly, (ii) the R1 domain and the modified R16 domain are fused directly, (iii) the modified R16 domain and the R17 domain are fused directly, (iv) the R17 domain and the H3 domain are fused directly, (v) the H3 domain and the R23 domain are fused directly, (vi) the R23 domain and the R24 domain are fused directly, or (vii) the R24 domain and the H4 domain are fused directly, or (vii) any combination thereof. In some embodiments, the miniaturized dystrophin polypeptide does not comprise a spectrin repeat 2 (R2) domain, spectrin repeat 3 (R3) domain, spectrin repeat 4 (R4) domain, spectrin repeat 5 (R5) domain, spectrin repeat 6 (R6) domain, spectrin repeat 7 (R7) domain, spectrin repeat 8 (R8) domain, spectrin repeat 9 (R9) domain, spectrin repeat 10 (R10) domain, spectrin repeat 11 (R11) domain, spectrin repeat 12 (R12) domain, spectrin repeat 13 (R13) domain, spectrin repeat 14 (R14) domain, spectrin repeat 15 (R15) domain, spectrin repeat 18 (R18) domain, spectrin repeat 19 (R19) domain, spectrin repeat 20 (R20) domain, spectrin repeat 21 (R21) domain, and/or spectrin repeat 22 (R22) domain. In some embodiments, the miniaturized dystrophin polypeptide further comprises an ABD1 domain and/or a CR domain. In some embodiments, the miniaturized dystrophin polypeptide consists essentially of or consists of, from N terminus to C terminus, the ABD1 domain, the H1 domain, the R1 domain, the modified R16 domain, the R17 domain, the H3 domain, the R23 domain, the R24 domain, the H4 domain, and the CR domain of dystrophin.

Each domain in the miniaturized dystrophin polypeptides can have one or more changes from the corresponding wild-type domain.

For example, the miniaturized dystrophin BXA-212372-J4V13 (BXA-220931) consists of the following protein domains in order:

5 Table 6. Amino acid sequence and domain structure of miniaturized dystrophin polypeptide BXA-212372-J4V13 (BXA-220931).

SEQ ID NO:	Description	Sequence
73	ABD1	MLWWEEVEDCYEREDVQKKTF TKWVNAQFSKFGKQHIEN LFSDLQDGRRLDLLEGLTGQKLPKEKGSTRVHALNNVNK ALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNILHWQV KNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAF NIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQ VSIEAIQEVE
74	Hinge 1	MLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSP KPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSS MES
75	Spectrin-1	EVNLDTRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKD QFHTHEGYMMDLTAHQGRVGNILQLGSKLIGTGKLSEDEE TEVQEQMNLLNSRWECLRVASMEKQSNLH
76	Modified Spectrin-16	RVLMDLQNQKLTEITHVSQALLEVEQLLNAPDLCAKDFED LFKQEEESLKNIKDSLQSSGRIDIIHSSKTAALQSATPVERV KLQEALSQDLDFQWEKVNKMYKDRQGRFDRS
77	Spectrin-17	VEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQE KLGSLNLRWQEVCKQLSDRKKRLEEQ
78	Hinge 3	APGLTTIGASPTQTVTLVTPPVVTKETAISKLEMPSSLMLE
79	Spectrin-23	SDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPAVQK QNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKL YQEPRELPPERAQNVTRLLRKQAEENVTEWEKLNLSAD WQRKIDET
80	Spectrin-24	LERLQELQEATDELCLKLRQAEVIKGSWQPVGDLLIDSLQD HLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNL STLEDLNTRWKLLQVAVEDRVRQLHE
81	Hinge 4	AHRDFGPASQHFLSTSVQGPWERAI SPNKVPYYINHETQTT CWDHPKMTELYQSLADLNNVRFSAYRTAMKL
82	CR	RRLQKALCLDLSLSAACDALDQHNKQNDQPMQDILQIINC LTTIYDRLEQEHNNLVNVPLCVDMLCNWLLNVYDTGRTG RIRVLSFKTGII SLCKAHLEDKYRYLFKQVASSTGFCQRRRL GLLLHDSIQIPRQLGEVASFGGSNI EPSVRSCFQFANNKPEIE AALFLDWMRLEPQSMVWLPVLHRVAAAETA KHQAKCNIC

		KECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMV EYCTPTTSGEDVRDFAKVLKNKFR TKRYFAKHPRMGYLPV QTVLEGDNMET
--	--	---

- In some embodiments, the H1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 74.
- 5 In some embodiments, the R1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 75. In some embodiments, the modified R16 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about
- 10 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 76. In some embodiments, the R17 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 77. In some embodiments, the H3 domain is an amino acid sequence at least about 80%, at
- 15 least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 78. In some embodiments, the R23 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 79.
- 20 In some embodiments, the R24 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 80. In some embodiments, the H4 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about
- 25 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 81. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the N terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 73. In some embodiments, the miniaturized

dystrophin polypeptide further comprises at the C terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 82.

- 5 The various miniaturized dystrophin polypeptides of the present disclosure are shown in Table 7.

Table 7. Amino Acid Sequences of miniaturized dystrophin constructs.

SEQ ID NO and Description	Sequence
SEQ ID NO:83 - BXA-212372-J4V13 (BXA-220931)	MLWWEVEEDCYEREDVQKKTF TKWVNAQFSKFGKQH IENLFS DLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLP RPPKVTKEEHFQLHHQM HYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTS DPTRSPFPSQHLEAPEDKSFGSSLMES E VNLDR YQTAL E EVLSWLLSAEDTLQAQGEISNDVEVVKDQFH THEGYMMDLTA HQGRVGNILQLGSKLIGTGK LSEDEETE VQE QMNLLNSRWECLRVASM EKQSNLHRV LMDLQNQKLTEITHVSQALLEVEQLLNAPDLCAKDFEDLF KQEESLKNIKDSLQSSGRIDIH SKKTAALQSATPVERVKLQEALSQ LDF QWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQLWTEAEQFLRKT QIPENWEHAKYK WYLKELQDGIGQRQT VVRTLNATGEEIIQQSSKTDAS ILQEKLGSNLNRWQEVCKQLSDRKKRLEEQAPGLTTIGASPTQT VTLVT QPVVTKETAISKLEMPSSLMLESDQWKRLHLSLQELLVWLQLKDEL SR QAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEG LEKLYQEPREL PPEERAQNVTRLLRKQAE EVNTEWEKLN LHSADWQRK IDETLERLQELQEATDELDLKLRAEVIKGSWQPVGDLLIDSLQDHLEKV KALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLEDLNTRWKLL QVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISP NKVPYYINH ETQTTCWDHPKMT ELYQSLADLNNVRFSA YRTAMKLRR LQKALCLDLL SLAACDALDQHNLKQNDQPM DILQIINCLTTIYDRLEQEHN NLVNVPL CVDMLCNWLLNVYDTGRTGRIRVLSFKTG IISLCKAHLEDKYRYL FKQV ASSTGFCDQRR LGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNK PEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPII GFRYRSLKHFN YDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVR DFAKVLKNKFR TKRYFAKHPRMGYLPVQTVLEGDNMET
SEQ ID NO:84 - BXA-	MLWWEVEEDCYEREDVQKKTF TKWVNAQFSKFGKQH IENLFS DLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLP RPPKVTKEEHFQLHHQM HYSQQITVSLAQGYERTSSPK

212372- J4V12	PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSLMSEVNLD YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLNSRWECLRVASM EKQSNLHRVLMDLQONQKTEITHVSQALLEVEQLLNAPDLCAKDFEDLFK QEESLKNIKDSLQSSGRIDIHHSKKTAAALQSATPVERVKLQEALSQLDFO WEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI PENWEHAKYKWYLKELQDGIGQRQTVVRTL NATGEEIIQQSSKTDASIL QEKLGSLNLRWQEVCKQLSDRKKRLEEQA PGLTTIGASPTQTVTLVTQP VVTKETAI SKLEMPSSLMLES DQWKRLHLSLQELLVWLQLKDDELSRQ APIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGL EKLYQEPREL PPEERAQNVTRLLRKQAE EVNTEWEKLNLSADWQRKI DETLERLQELQEATDELDLKL RQAEVIKGSWQPVGDLLIDSLQDHLEKV KALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLEDLNTRWKLL QVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISP NKVPYYINH ETQTTCDWHPKMTELYQSLADLNNVRFSAYRTAMKLRLQKALCLDLL SLSAACDALDQHNLKQNDQPM DILQIINCLTTIYDRLEQEHNNLVNPL CVDMLCNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQV ASSTGFCDQRR LGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNK PEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPII GFRYRSLKHFN YDICQSCFFSGR VAKGHKMHYPMVEYCTPTTSGEDVR DFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
SEQ ID NO:85 BXA- 212372- J4V11	MLWWEVEEDCYEREDVQKKTF TKWVNAQFSKFGKQH IENLFSDLQDG RRLDLLEGLTGQKL PKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVD TTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLP RPPKVTKEEHFQLHHQM HYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSLMSEVNLD YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLNSRWECLRVASM EKQSNLHRVLMDLQONQKEITHVSQALLEVEQLLNAPDLCAKDFEDLFK QEESLKNIKDSLQSSGRIDIHHSKKTAAALQSATPVERVKLQEALSQLDFO WEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI PENWEHAKYKWYLKELQDGIGQRQTVVRTL NATGEEIIQQSSKTDASIL QEKLGSLNLRWQEVCKQLSDRKKRLEEQA PGLTTIGASPTQTVTLVTQP VVTKETAI SKLEMPSSLMLES DQWKRLHLSLQELLVWLQLKDDELSRQ APIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGL EKLYQEPREL PPEERAQNVTRLLRKQAE EVNTEWEKLNLSADWQRKI DETLERLQELQEATDELDLKL RQAEVIKGSWQPVGDLLIDSLQDHLEKV KALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLEDLNTRWKLL QVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISP NKVPYYINH ETQTTCDWHPKMTELYQSLADLNNVRFSAYRTAMKLRLQKALCLDLL SLSAACDALDQHNLKQNDQPM DILQIINCLTTIYDRLEQEHNNLVNPL CVDMLCNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQV ASSTGFCDQRR LGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNK PEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPII

		GFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVR DFAKVLKNKFRTRKRYFAKHPRMGYLPVQTVLEGDNMET
SEQ ID NO:86 BXA- 212372- J4V4	-	MLWWEVEEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYDPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFSPQHLEAPEDKSFGSSLMSEVNLDL YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVQEQMNNLSNRWECLRVASM EKQSNLHRVLMDLTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQ EESLKNIKDSLQQSSGRIDIIHSKKTAAALQSATPVERVKLQEALSQDLFQ WEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQI PENWEHAKYKWYKELQDGIGQRQTIVRTL NATGEEIIQQSSKTDASIL QEKLGSLNLRWQEVCKQLSDRKKRLEEQA PGLTTIGASPTQTVTLVTQP VVTKET AISKLEMPSSLMLESDQWKRLHLSLQELLVWLQLKDDELSRQ APIGGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGL EKLYQEPREL PPEERAQNVTRLLRKQAEEVNTEWEKLNLSADWQRKI DETLERLQELQEATDELDLKLRAEVIKGSWQPVGDLLIDSLQDHLEKV KALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLEDLNTRWKLL QVAVEDRVRQLHEAHRDFGPASQHFLLSTSVQGPWERAI SPNKVPYYINH ETQTTCDWHPKMTLEYQSLADLNNVRFSA YRTAMKLRLRLQKALCLDLL SLSAACDALDQHNKQNDQPM DILQIINCLTTIYDRLEQEHNNLVNVPL CVDMLCNWLLNVYDTGRTGRIRVLSFKTGHS LCKAHLEDKYRYLFKQV ASSTGFC DQRRGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNK PEIEAALFLDWMRLPQSMVWLPVLHRVAAAETA KHQAKCNICKECPII GFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVR DFAKVLKNKFRTRKRYFAKHPRMGYLPVQTVLEGDNMET
SEQ ID NO:87 BXA- 212372-J4	-	MLWWEVEEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYDPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPRPPKVTKEEHFQLHHQMHYSSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFSPQHLEAPEDKSFGSSLMSEVNLDL YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVQEQMNNLSNRWECLRVASM EKQSNLHSYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEE SLKNIKDSLQQSSGRIDIIHSKKTAAALQSATPVERVKLQEALSQDLFQWE KVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPE NWEHAKYKWYKELQDGIGQRQTIVRTL NATGEEIIQQSSKTDASILQE KLGS LNLRWQEVCKQLSDRKKRLEEQA PGLTTIGASPTQTVTLVTQPVV TKETAISKLEMPSSLMLESDQWKRLHLSLQELLVWLQLKDDELSRQAPI GGDFPAVQKQNDVHRAFKRELKTKEPVIMSTLETVRIFLTEQPLEGLEKL YQEPREL PPEERAQNVTRLLRKQAEEVNTEWEKLNLSADWQRKIDET

		LERLQELQEATDELDLKLRLQAEVIKGSWQPVGDLLIDSLQDHLEKVKAL RGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLEDLNTRWKLLQVA VEDRVRQLHEAHRDFGPASQHFLLSTSVQGPWERAISPKNVPYYINHETQ TTCWDHPKMTLEYQSLADLNNVRFSAYRTAMKLRRLQKALCLDLSLS AACDALDQHNKQNDQPMQDILQIINCLTTIYDRLEQEHNNLVNPLCVD MCLNWLLNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLQVASS TGFCQRRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEI EALFLDWMRLEPQSMVWLPVLRVAAAETAKHQAKCNICKECPIIGF RYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDF AKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMET
SEQ ID NO:88 BXA- 212372	-	MLWWEVEEDCYEREDVQKKTFKWNVAQFSKFGKQHINLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHSYQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSPHLEAPEDKSFSSLMSESEVNLDR YQTALVEEVLWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLNSRWECLRVASM EKQSNLHIHTVREETMMVMTEMDMPLEISYVPSTYLTEITHVSQALLEVEQ LLNAPDLCAKDFEDLFKQEEESLKNIKDSLQSSGRIDIIHSSKKAALQSAT PVERVKLQEALSQDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKI FNQWLTEAEQFLRKTQIPENWEHAKYKWYKELQDGIGQRQTVVRTLNL ATGEEIIQQSSKTDASILQEKLGSLNLRWQEVCKQLSDRKKRLEEQAPGL TTIGASPTQTVTLVTQPVTKETAIKLEMPSSLMLESQWKRHLHSLQE LLVWLQLKDDELSRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVIMST LETVRIFLTEQPLEGLEKLYQEPRELPEERAQNVTRLLRKQAEENVTEW EKLNLHSADWQRKIDETLERLQELQEATDELDLKLRLQAEVIKGSWQPV GDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYN LSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLLSTSVQGPW ERAISPKNVPYYINHETQTTCWDHPKMTLEYQSLADLNNVRFSAYRTA MKLRRLQKALCLDLSLSAACDALDQHNKQNDQPMQDILQIINCLTTIY DRLEQEHNNLVNPLCVDMMCLNWLLNVYDTGRTGRIRVLSFKTGIISL KAHLEDKYRYLQVASSTGFCQRRRLGLLLHDSIQIPRQLGEVASFGGS NIEPSVRSCFQFANNKPEIEALFLDWMRLEPQSMVWLPVLRVAAAET AKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYP MVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEG DNMET

In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 83. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence identical to SEQ ID NO: 83. In

some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 84. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 85. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 86. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 87.

In some embodiments, the amino acid sequence of the miniaturized dystrophin disclosed herein when expressed has at least one dystrophin activity.

In some embodiments, a nucleic acid sequence encoding each domain can be the following:

Table 8: Nucleotide sequence (and domain structure) encoding miniaturized dystrophin polypeptide BXA-220931.

SEQ ID NO:	Description	Nucleotide Sequence
89	5' UTR	CCGCCTTCGGCACCATTCTCACGACACCCAAATATGGCGAC GGGTGAGGAATGGTGGGGAGTTATTTTATAGAGCGGTGAGGAA GGTGGGCAGGCAGCAGGTGTTGGCGCTCTAAAAATAACTCCC GGGAGTTATTTTATAGAGCGGAGGAATGGTGGACACCCAAATA TGGCGACGGTTCCTACCCGTCGCCATATTTGGGTGTCCGCCC TCGGCCGGGGCCGCATTCTGGGGGCCGGGCGGTGCTCCCGC CCGCCTCGATAAAAGGCTCCGGGGCCGGCGGCGGCCACGA GCTACCCGGAGGAGCGGGAGGCACGCGTCTCTAAGGTAAAT ATAAAATTTTAAAGTGTATAATGTGTAAACTACTGATTCTAA TTGTTTCTCTCTTTTAGATTCCAACCTTTGGAAGTATCTAGA CCACC
90	ABD1	ATGCTTTGGTGGGAAGAAGTCGAGGACTGCTACGAGCGCGAG GACGTGCAGAAGAAAACCTTCACCAAATGGGTCAACGCCCA

		G TTCAGCAAGTTCGGCAAGCAGCACATCGAGAACCTGTTTCAG CGACCTGCAGGATGGCAGAAGGCTGCTGGATCTGCTGGAAGG CCTGACAGGCCAGAAGCTGCCTAAAGAGAAGGGCAGCACAA GAGTGCACGCCCTGAACAACGTGAACAAGGCCCTGAGAGTG CTGCAGAACAAACAGTGGACCTGGTCAACATCGGCAGCACC GACATCGTGGACGGCAATCACAACCTGACCCTGGGCCTGATC TGGAACATCATCCTGCACTGGCAAGTGAAGAACGTGATGAAG AACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGAT TCTGCTGAGCTGGGTCCGACAGAGCACCCGGAACCTACCCTCA AGTGAACGTGATCAACTTCACCACCTCTTGAGCGACGGACT GGCCCTGAATGCCCTGATTCACAGCCACAGACCTGACCTGTT CGACTGGAATAGCGTCGTGTGTCAGCAGAGCGCCACACAGAG ACTGGAACACGCCTTCAATATCGCCAGATACCAGCTGGGCAT CGAGAACTGCTGGACCCCGAGGATGTGGACACCACCTATCC TGACAAGAAATCCATCCTCATGTACATCACCAGCCTGTTCCA GGTGCTGCCCCAGCAAGTGTCTATCGAGGCCATTCAAGAGGT CGAG
91	Hinge 1	ATGCTGCCCAGACCTCCTAAAGTGACCAAAGAGGAACACTTC CAGCTGCACCACCAGATGCACTACTCTCAGCAGATCACCGTG TCTCTGGCCCAGGGCTACGAGAGAACAAGCAGCCCCAAGCCT CGGTTCAAGAGCTACGCCTATACACAGGCCGCCTACGTGACC ACCAGCGATCCCACAAGAAGCCCATTTCCAAGCCAGCATCTG GAAGCCCCTGAGGACAAGAGCTTTGGCAGCAGCCTGATGGA AAGC
92	Spectrin-1	GAAGTGAACCTGGATAGATACCAGACAGCCCTGGAAGAGGT GCTGTCTTGGCTGCTGTCTGCCGAAGATACACTGCAGGCTCA GGGCGAGATCAGCAACGACGTGGAAGTGGTCAAGGACCAGT TTCACACCCACGAGGGCTACATGATGGACCTGACAGCCCATC AGGGCAGAGTGGGCAATATCCTGCAGCTGGGCTCTAAGCTGA TCGGCACAGGCAAGCTGAGCGAGGACGAAGAGACAGAGGTG CAAGAGCAGATGAACCTGCTGAACAGCAGATGGGAGTGTCT GAGAGTGGCCAGCATGGAAAAGCAGAGCAACCTGCAC
93	Modified Spectrin-16	CGGGTCCTGATGGATCTGCAGAATCAGAAGCTGACCGAGATC ACCCACGTGTCACAGGCCCTGCTTGAAGTGGAAACAGCTGCTG AACGCCCCTGATCTGTGCGCCAAGGACTTCGAGGATCTGTTC AAGCAAGAGGAAAGCCTGAAGAATATCAAGGACTCTCTGCA GCAGTCCAGCGGCCGGATCGACATCATCCACAGCAAGAAAA CAGCTGCCCTGCAGTCCGCCACACCTGTGGAAAGAGTGAAAC TGCAAGAGGGCCCTGTCTCAGCTGGACTTCCAGTGGGAGAAAG TGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCGC TCT
94	Spectrin-17	GTGGAAAAATGGCGGAGATTCCACTACGACATCAAGATCTTC AACCAGTGGCTGACAGAGGCCGAGCAGTTCTTGAGAAAGAC ACAGATCCCCGAGAAGTGGGAGCACGCCAAGTACAAGTGGT ATCTGAAAGAACTGCAGGACGGCATCGGCCAGAGGCAGACA GTCGTTAGAACACTGAATGCCACCGGCGAGGAAATCATCCAG

		CAGAGCAGCAAGACCGACGCCAGCATCCTGCAAGAGAAGCT GGGCAGCCTGAACCTGAGATGGCAAGAAGTGTGCAAGCAGC TGTCCGACCGGAAGAAGAGGCTGGAAGAACAG
95	Hinge 3	GCCCCCTGGCCTGACAACAATCGGAGCCTCTCTACACAGACC GTGACACTGGTCACACAGCCCGTGGTCACCAAAGAGACAGCC ATCAGCAAGCTGGAAATGCCCTCTAGCCTGATGCTCGAG
96	Spectrin-23	AGCGACCAGTGGAAAGAGACTGCACCTGTCTCTGCAAGAGCTG CTCGTGTGGCTGCAGCTGAAGGACGATGAACTGAGCAGACAG GCCCCAATCGGAGGCGATTTTCCTGCCGTGCAGAAACAGAAC GACGTGCACAGAGCCTTCAAGCGGGAAGTGAACAAAAGA ACCCGTGATCATGAGCACCCCTGGAAACCGTGC GGATCTTCCT GACAGAGCAGCCTCTCGAAGGCCTGGAAGAGCTGTACCAAG AGCCTAGAGAGCTGCCTCCTGAGGAACGGGCCAGAAATGTGA CCAGACTGCTGAGAAAGCAGGCCGAAGAGGTCAACACCGAA TGGGAGAAGCTGAACCTGCACAGCGCCGACTGGCAGAGAAA GATCGACGAGACA
97	Spectrin-24	CTGGAACGGCTGCAAGAACTCCAAGAAGCCACCGACGAGCT GGACCTGAAACTGAGGCAGGCTGAAGTGATCAAAGGCAGCT GGCAGCCAGTGGGCGACCTGCTGATTGATAGTCTGCAGGACC ACCTGGAAAAAGTGAAGGCCCTGCGGGGAGAGATCGCCCCA CTGAAAGAAAACGTGTCCACGTGAACGACCTGGCCAGACA GCTGACAACCCTGGGAATCCAGCTGTCCCCTTACAACCTGTC CACACTGGAAGATCTGAACACCCGGTGGAAACTGCTCCAGGT GGCCGTGGAAGATAGAGTGCGACAGCTGCACGAG
98	Hinge 4	GCCCACAGAGATTTTGGACCAGCCAGCCAGCACTTCCTGTCT ACATCTGTGCAAGGCCCTTGGGAGAGAGCTATCAGCCCTAAC AAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGT TGGGATCACCCCAAGATGACCGAGCTGTATCAGAGCCTGGCC GACCTGAACAATGTGCGCTTTAGCGCCTACCGGACCGCCATG AAGCTG
99	CR	CGGAGACTGCAGAAAGCCCTGTGTCTGGACCTGCTGTCTCTG TCTGCAGCCTGTGATGCCCTGGACCAGCACAACTGAAGCAG AACGACCAGCCTATGGACATCCTCCAGATCATCAACTGCCTG ACCACCATCTACGACCGGCTGGAACAAGAGCACAAACACCTC GTGAATGTGCCCTGTGCGTGGACATGTGTCTGAACTGGCTG CTGAATGTGTACGACACCGGCAGAACCGGCAGGATCAGAGT GCTGAGCTTCAAGACCGGCATCATCTCCCTGTGCAAAGCCCA CCTCGAGGACAAGTACAGATACCTGTTCAAACAGGTGGCCAG CTCCACCGGCTTTTGCATCAAAGAAGGCTGGGCCTGCTGCT GCACGACAGCATCCAGATTCTAGACAGCTGGGCGAAGTGGC CTCCTTCGGCGGATCTAATATTGAGCCTAGCGTGCGGAGCTG CTTCCAGTTCGCCAACAACAAGCCTGAGATCGAGGCCGCTCT GTTCCCTGGATTGGATGCGCCTGGAACCTCAGAGCATGGTTTG GCTGCCTGTGCTGCATAGAGTGGCCGCTGCCGAAACAGCCAA GCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCAT CGGCTTCCGGTACAGATCCCTGAAGCACTTCAACTACGATAT

		CTGCCAGAGCTGTTTCTTCTCTGGCCGCGTGGCCAAGGGCCA CAAAATGCACTACCCCATGGTGGGAATACTGCACCCCTACCAC ATCTGGCGAAGATGTGCGGGATTTCGCCAAGGTGCTGAAAAA CAAGTTCGGACCAAGCGGTACTTCGCTAAGCACCCCAGAAT GGGCTATCTGCCCCGTGCAGACAGTGCTCGAGGGCGATAACAT GGAAACCTGA
--	--	---

In some embodiments, a nucleic acid sequence encoding the H1 domain in the miniaturized dystrophin polypeptide is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 91. In some embodiments, a nucleic acid sequence encoding the R1 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 92. In some embodiments, a nucleic acid sequence encoding the modified R16 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 93. In some embodiments, a nucleic acid sequence encoding the R17 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 94. In some embodiments, a nucleic acid sequence encoding the H3 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 95. In some embodiments, a nucleic acid sequence encoding the R23 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 96. In some embodiments, a nucleic acid sequence encoding the R24 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 97.

In some embodiments, a nucleic acid sequence encoding the H4 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 98. In some embodiments, a nucleic acid sequence encoding the ABD1 domain in the miniaturized dystrophin polypeptide is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 90. In some embodiments, a nucleic acid sequence encoding the CR/C-term. polypeptide is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 99.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule exhibits one or more properties selected from the group consisting of (i) having a lower CD4 proliferation compared to BXA-212372 (SEQ ID NO: 88), (ii) having a lower CD8 proliferation compared to BXA-212372 (SEQ ID NO: 88), and (iv) any combination thereof.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule has formula (I):

H1-R1-mR16-R17-H3-R23-R24-H4

(I)

wherein: H1 is a hinge 1 domain of dystrophin; R1 is a spectrin repeat 1 domain of dystrophin; mR16 is a modified spectrin repeat 16 of dystrophin; R17 is a spectrin repeat 17 of dystrophin; H3 is a hinge 3 domain of dystrophin; R23 is a spectrin repeat 23 of dystrophin; R24 is a spectrin repeat 24 of dystrophin; H4 is a hinge 4 domain of dystrophin; and (-) is a peptide bond.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 83.

In some embodiments, the miniaturized dystrophin polypeptide exhibits a higher expression of the miniaturized dystrophin polypeptide than BXA-212372 (SEQ ID NO: 88). In some other embodiments, the miniaturized dystrophin polypeptide expression is at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about 1.9 fold, at least about 2 fold, at least about 2.1 fold, at least about 2.2 fold, at least about 2.3 fold, at least about 2.4 fold, at least about 2.5 fold, at least about 2.6 fold, at least about 2.7 fold, at least about 2.8 fold, at least about 2.9 fold or at least about 3 fold higher than the BXA-212372 (SEQ ID NO: 88) polypeptide expression.

In some embodiments, the miniaturized dystrophin polypeptides can be encoded by nucleotide sequences. Some examples of the nucleotide sequences are shown in Table 9.

Table 9. Nucleotide Sequences of Dystrophin constructs.

SEQ ID NO and Description	Sequence
SEQ ID NO:100 - BXA- 220931	ATGCTTTGGTGGGAAGAAGTCGAGGACTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTCAACGCCAGTTCAGCAAG TTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGATGG CAGAAGGCTGCTGGATCTGCTGGAAGGCCTGACAGGCCAGAAGCTG CCTAAAGAGAAGGGCAGCACAAAGAGTGCACGCCCTGAACAACGTGA ACAAGGCCCTGAGAGTGCTGCAGAACAACAACGTGGACCTGGTCAA CATCGGCAGCACCGACATCGTGGACGGCAATCACAACTGACCCTGG GCCTGATCTGGAACATCATCCTGCACTGGCAAGTGAAGAACGTGATG AAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATTC TGCTGAGCTGGGTCCGACAGAGCACCCGGAACCTACCCTCAAGTGAAC GTGATCAACTTCACCACCTCTTGGAGCGACGGACTGGCCCTGAATGC CCTGATTCACAGCCACAGACCTGACCTGTTCGACTGGAATAGCGTCG TGTGTCAGCAGAGCGCCACACAGAGACTGGAACACGCCTTCAATATC GCCAGATACCAGCTGGGCATCGAGAAACTGCTGGACCCCGAGGATGT GGACACCACCTATCCTGACAAGAAATCCATCCTCATGTACATCACCA GCCTGTTCCAGGTGCTGCCCCAGCAAGTGTCTATCGAGGCCATTCAA GAGGTCGAGATGCTGCCCAGACCTCCTAAAGTGACCAAAGAGGAAC ACTTCCAGCTGCACCACCAGATGCACTACTCTCAGCAGATCACCGTG TCTCTGGCCCAGGGCTACGAGAGAACAAGCAGCCCCAAGCCTCGGTT CAAGAGCTACGCCTATACACAGGCCGCCTACGTGACCACCAGCGATC CCACAAGAAGCCCATTTCAGCCAGCATCTGGAAGCCCCTGAGGAC AAGAGCTTTGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGATA GATACCAGACAGCCCTGGAAGAGGTGCTGTCTTGGCTGCTGTCTGCC GAAGATACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGAAG TGGTCAAGGACCAGTTTCACACCCACGAGGGCTACATGATGGACCTG ACAGCCCATCAGGGCAGAGTGGGCAATATCCTGCAGCTGGGCTCTAA

SEQ ID NO and Description	Sequence
	GCTGATCGGCACAGGCAAGCTGAGCGAGGACGAAGAGACAGAGGTG CAAGAGCAGATGAACCTGCTGAACAGCAGATGGGAGTGTCTGAGAG TGGCCAGCATGGAAAAGCAGAGCAACCTGCACCGGGTCCTGATGGA TCTGCAGAATCAGAAGCTGACCGAGATCACCCACGTGTCACAGGCC TGCTTGAAGTGGAACAGCTGCTGAACGCCCCTGATCTGTGCGCCAAG GACTTCGAGGATCTGTTCAAGCAAGAGGAAAGCCTGAAGAATATCA AGGACTCTCTGCAGCAGTCCAGCGGCCGATCGACATCATCCACAGC AAGAAAACAGCTGCCCTGCAGTCCGCCACACCTGTGGAAAGAGTGA AACTGCAAGAGGGCCCTGTCTCAGCTGGACTTCCAGTGGGAGAAAGTG AACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCGCTCTGTGG AAAAATGGCGGAGATTCCACTACGACATCAAGATCTTCAACCAAGTGG CTGACAGAGGCCGAGCAGTTCCCTGAGAAAGACACAGATCCCCGAGA ACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAAGTGCAGGA CGGCATCGGCCAGAGGCAGACAGTCGTTAGAACACTGAATGCCACC GGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCC TGCAAGAGAAGCTGGGCAGCCTGAACCTGAGATGGCAAGAAGTGTG CAAGCAGCTGTCCGACCGGAAGAAGAGGCTGGAAGAACAGGCCCT GGCCTGACAACAATCGGAGCCTCTCCTACACAGACCGTGACACTGGT CACACAGCCCGTGGTCAACAAAGAGACAGCCATCAGCAAGCTGGAA ATGCCCTCTAGCCTGATGCTCGAGAGCGACCAAGTGGAAAGAGACTGCA CCTGTCTCTGCAAGAGCTGCTCGTGTGGCTGCAGCTGAAGGACGATG AACTGAGCAGACAGGCCCCAATCGGAGGCGATTTTCTGCCGTGCAG AAACAGAACGACGTGCACAGAGCCTTCAAGCGGGAAGTGAACAA AAGAACCCGTGATCATGAGCACCTTGGAACCGTGCGGATCTTCTG ACAGAGCAGCCTCTCGAAGGCCTGGAAGAGCTGTACCAAGAGCCTA GAGAGCTGCCTCCTGAGGAACGGGCCCAGAATGTGACCAGACTGCTG AGAAAGCAGGCCGAAGAGGTCAACACCGAATGGGAGAAGCTGAACC TGCACAGCGCCGACTGGCAGAGAAAGATCGACGAGACACTGGAACG GCTGCAAGAACTCCAAGAAGCCACCGACGAGCTGGACCTGAAACTG AGGCAGGCTGAAGTGATCAAAGGCAGCTGGCAGCCAGTGGGCGACC TGCTGATTGATAGTCTGCAGGACCACCTGGAAAAAGTGAAGGCCCTG CGGGGAGAGATCGCCCCACTGAAAGAAAACGTGTCCACGTGAACG ACCTGGCCAGACAGCTGACAACCCTGGGAATCCAGCTGTCCCCTTAC AACCTGTCCACACTGGAAGATCTGAACACCCGGTGGAAACTGCTCCA GGTGGCCGTGGAAGATAGAGTGCGACAGCTGCACGAGGCCACAGA GATTTTGGACCAGCCAGCCAGCACTTCTGTCTACATCTGTGCAAGG CCCTTGGGAGAGAGCTATCAGCCCTAACAAGGTGCCCTACTACATCA ACCACGAGACACAGACCACCTGTTGGGATCACCCCAAGATGACCGA GCTGTATCAGAGCCTGGCCGACCTGAACAATGTGCGCTTTAGCGCCT ACCGGACCGCCATGAAGCTGCGGAGACTGCAGAAAGCCCTGTGTCTG GACCTGCTGTCTGTCTGCAGCCTGTGATGCCCTGGACCAGCACAA CCTGAAGCAGAACGACCAGCCTATGGACATCCTCCAGATCATCAACT GCCTGACCACCATCTACGACCGGCTGGAACAAGAGCACAAACCTC

SEQ ID NO and Description	Sequence
	GTGAATGTGCCCCTGTGCGTGGACATGTGTCTGAACTGGCTGCTGAA TGTGTACGACACCGGCAGAACCGGCAGGATCAGAGTGCTGAGCTTCA AGACCGGCATCATCTCCCTGTGCAAAGCCCACCTCGAGGACAAGTAC AGATACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTGCGATCA AAGAAGGCTGGGCCTGCTGCTGCACGACAGCATCCAGATTCCTAGAC AGCTGGGCGAAGTGGCCTCCTTCGGCGGATCTAATATTGAGCCTAGC GTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCTGAGATCGAGGC CGCTCTGTTCCCTGGATTGGATGCGCCTGGAACCTCAGAGCATGGTTTG GCTGCCTGTGCTGCATAGAGTGGCCGCTGCCGAAACAGCCAAGCACC AGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGG TACAGATCCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGTTTC TTCTCTGGCCGCGTGGCCAAGGGCCACAAAATGCACTACCCCATGGT GGAATACTGCACCCCTACCACATCTGGCGAAGATGTGCGGGATTTTCG CCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTCGCTAAG CACCCAGAATGGGCTATCTGCCCCTGCAGACAGTGCTCGAGGGCGA TAACATGGAAACCTGA
SEQ ID NO:101 - BXA- 212372- J4V13	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACAGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCCTG GGCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCAGCTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCTTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG

SEQ ID NO and Description	Sequence
	GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACCGGGTCTCTGAT GGATCTGCAGAATCAGAAGCTGACCGAGATCACCCACGTGTCCAGG CTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCC AAGGACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATA TCAAGGACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCATCCAC AGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAAAGAG TGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAA AGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGC GTGGAAAAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCAACCA GTGGCTGACCGAGGCGGAGCAGTTCTTGAGAAAGACCCAGATCCCCG AGAACTGGGAGCAGCCAAAGTACAAGTGGTATCTGAAAGAGCTGCA GGACGGCATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGAATGCC ACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCA TCCTGCAGGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGT GTGCAAGCAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAACAGGCC CCTGGCCTGACCACAATCGGCGCCAGCCCTACCCAGACCGTGACCCT GGTGACACAGCCCGTGGTGACAAAAGAGACAGCCATCAGCAAGCTG GAAATGCCCAGCAGCCTGATGCTGGAAAGCGACCAGTGGAAGCGGC TGCACCTGAGCCTGCAGGAACTGCTGGTCTGGCTGCAGCTGAAGGAC GACGAGCTGAGCAGACAGGCCCCCATCGGCGGCGATTTCCTCCGCCGT GCAGAAACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGCTGAAA ACAAAGAACCCTGATCATGAGCACCTTGGAACCGTGCGGATCTT CCTGACCGAGCAGCCCCTGGAAGGCCTGGAAGAGCTGTACCAGGAA CCCAGAGAGCTGCCCCCGAGGAACGGGCCCAGAACGTGACCAGAC TGCTGCGGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAGAAAGCT GAACCTGCACAGCGCCGACTGGCAGCGGAAGATCGACGAGACACTG GAACGGCTGCAGGAACTGCAGGAGGCCACCGACGAGCTGGACCTGA AGCTGAGACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCCGTGGG CGACCTGCTGATCGACTCCCTGCAGGACCACCTGGAAAAAGTGAAGG CCCTGCGGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTG AACGACCTGGCCCGGCAGCTGACCACCCTGGGCATCCAGCTGAGCCC CTACAACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGAAGCTGC TGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCCA CAGAGACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCTCTGTGC AGGGACCCTGGGAGAGAGCCATCAGCCCCAACAAAGGTGCCCTACTA CATCAACCACGAGACACAGACCACCTGTTGGGACCACCCCAAGATGA CCGAGCTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGGTTTCAGC GCCTACCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTG CCTGGATCTGCTGAGCCTGAGCGCCGCCTGCGACGCCCTGGACCAGC ACAACCTGAAGCAGAACGACCAGCCCATGGATATCCTGCAGATCATC AACTGCCTGACCACAATCTACGACAGGCTGGAACAGGAACACAACA ATCTGGTCAACGTGCCCTGTGCGTGGACATGTGCCTGAATTGGCTG

SEQ ID NO and Description	Sequence
	CTGAATGTGTACGACACCGGCCGGACCGGCAGAATCCGGGTGCTGAG CTTCAAGACCGGCATCATCAGCCTGTGCAAGGCCACCTGGAAGATA AGTACCGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTGC GACCAGCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCAGATCCC CAGACAGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGC CCAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATC GAGGCCGCCCTGTTCTGGACTGGATGAGACTGGAACCCAGAGCAT GGTGTGGCTGCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCA AGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGC TTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAG CTGCTTCTTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGCACTACC CCATGGTGGAACTACTGCACCCCCACCACCAGCGGCGAGGATGTGCGG GACTTCGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTT TGCCAAGCACCCCCGGATGGGCTACCTGCCCCGTGCAGACAGTGCTGG AAGGCGACAACATGGAAACCTGA
SEQ ID NO:102 - BXA- 212372- J4V12	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACAGTGGAACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCAGCTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCTTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCTGC

SEQ ID NO and Description	Sequence
	GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACAGAGTTTAAAT GGATCTCCAGAATCAGAAAACCGAGATCACCCACGTGTCCCAGGCTC TGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAG GACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCA AGGACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCATCCACAGC AAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAAAGAGTGA AGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGT GAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGCGTG GAAAAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAGTG GCTGACCGAGGCCGAGCAGTTTCTGAGAAAGACCCAGATCCCCGAG AACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAGCTGCAGG ACGGCATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGAATGCCACC GGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCC TGCAGGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTG CAAGCAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAACAGGCCCT GGCCTGACCACAATCGGCGCCAGCCCTACCCAGACCGTGACCCTGGT GACACAGCCCGTGGTGACAAAAGAGACAGCCATCAGCAAGCTGGAA ATGCCCAGCAGCCTGATGCTGGAAAGCGACCAGTGGAAGCGGCTGC ACCTGAGCCTGCAGGAAGTGTGCTGGTCTGGCTGCAGCTGAAGGACGAC GAGCTGAGCAGACAGGCCCCCATCGGCGGCGATTTCGCCCGCCGTGCA GAAACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGCTGAAAACA AAAGAACCCGTGATCATGAGCACCCCTGGAAACCGTGCGGATCTTCCT GACCGAGCAGCCCCCTGGAAGGCCTGGAAGAGCTGTACCAGGAACCC AGAGAGCTGCCCCCGAGGAACGGGCCCAGAACGTGACCAGACTGC TGCGGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAGAAGCTGAA CCTGCACAGCGCCGACTGGCAGCGGAAGATCGACGAGACACTGGAA CGGCTGCAGGAAGTGCAGGAGGCCACCGACGAGCTGGACCTGAAGC TGAGACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCCGTGGGCGA CCTGCTGATCGACTCCCTGCAGGACCACCTGGAAAAAGTGAAGGCC TGCGGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTGAAC GACCTGGCCCGGCAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTA CAACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGAAAGCTGCTGC AGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCACAG AGACTTTGGCCCTGCCAGCCAGCACTTCTGAGCACCTCTGTGCAGG GACCCTGGGAGAGAGCCATCAGCCCCAACAAGGTGCCCTACTACATC AACCACGAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCG AGCTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCC TACCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTGCCT GGATCTGCTGAGCCTGAGCGCCGCCTGCGACGCCCTGGACCAGCACA ACCTGAAGCAGAACGACCAGCCCATGGATATCCTGCAGATCATCAAC TGCCTGACCACAATCTACGACAGGCTGGAACAGGAACACAACAATCT GGTCAACGTGCCCCGTGTGCGTGGACATGTGCCTGAATTGGCTGCTGA ATGTGTACGACACCGGCCGGACCGGCAGAAATCCGGGTGCTGAGCTTC

SEQ ID NO and Description	Sequence
	AAGACCGGCATCATCAGCCTGTGCAAGGCCACCTGGAAGATAAGTA CCGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTGCGACC AGCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGA CAGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGCCCAG CGTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGG CCGCCCTGTTCTGGAAGTGGATGAGACTGGAACCCAGAGCATGGTG TGGCTGCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCAAGCA CCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCC GGTACAGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGC TTCTTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGCACTACCCCAT GGTGGAATACTGCACCCCCACCACCGCGGCGAGGATGTGCGGGACT TCGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCC AAGCACCCCCGGATGGGCTACCTGCCCCGTGCAGACAGTGCTGGAAGG CGACAACATGGAAACCTGA
SEQ ID NO:103 - BXA- 212372- J4V11	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGTGTCAGAAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACAGAGTTTAAAT

SEQ ID NO and Description	Sequence
	GGATCTCCAGAATCAGAAAGAGATCACCCACGTGTCCCAGGCTCTGC TGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGAC TTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGG ACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCATCCACAGCAAG AAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAAAGAGTGAAGCT GCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGTGAAC AAGATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAA AGTGGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAAGTGGCTG ACCGAGGCCGAGCAGTTCCTGAGAAAGACCCAGATCCCCGAGAACT GGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGG CATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGAATGCCACCGGCG AGGAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCCTGCA GGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAG CAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAACAGGCCCTGGCC TGACCACAATCGGCGCCAGCCCTACCCAGACCGTGACCCTGGTGACA CAGCCCGTGGTGACAAAAGAGACAGCCATCAGCAAGCTGGAAATGC CCAGCAGCCTGATGCTGGAAGCGACCAAGTGGAAAGCGGCTGCACCT GAGCCTGCAGGAAGTGTGGTCTGGCTGCAGCTGAAGGACGACGAG CTGAGCAGACAGGCCCCCATCGGCGGCGATTTCGCCCGCGTGACAGAA ACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGCTGAAAACAAAA GAACCCGTGATCATGAGCACCTGGAACCGTGCGGATCTTCCTGAC CGAGCAGCCCCTGGAAGGCCTGGAAGAGCTGTACCAGGAACCCAGA GAGCTGCCCCCGAGGAACGGGCCCAGAACGTGACCAGACTGCTGC GGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAGAAGCTGAACCT GCACAGCGCCGACTGGCAGCGGAAGATCGACGAGACACTGGAACGG CTGCAGGAAGTGCAGGAGGCCACCGACGAGCTGGACCTGAAGCTGA GACAGGCCGAAGTATCAAGGGCAGCTGGCAGCCCCTGGGCGACCT GCTGATCGACTCCCTGCAGGACCACCTGGAAGAGTGAAGGCCCTGC GGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTGAACGA CCTGGCCCCGCGAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTACA ACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAG GTGGCCGTGGAAGATAGAGTGCAGGAGCTGCACGAGGCCACAGAG ACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCTCTGTGCAGGGA CCCTGGGAGAGAGCCATCAGCCCCAACAAGGTGCCCTACTACATCAA CCACGAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAG CTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCCTA CCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTGCCTGG ATCTGCTGAGCCTGAGCGCCGCCTGCGACGCCCTGGACCAGCACAAC CTGAAGCAGAACGACCAGCCATGGATATCCTGCAGATCATCAACTG CCTGACCACAATCTACGACAGGCTGGAACAGGAACACAACAATCTG GTCAACGTGCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAA TGTGTACGACACCGGCCGACCGGCAGAAATCCGGGTGCTGAGCTTCA AGACCGGCATCATCAGCCTGTGCAAGGCCACCTGGAAGATAAGTAC

SEQ ID NO and Description	Sequence
	CGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTGCGACCA GCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGAC AGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGCCCAGC GTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGC CGCCCTGTTCTTGACTGGATGAGACTGGAACCCAGAGCATGGTGT GGCTGCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCAAGCAC CAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCG GTACAGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGCT TCTTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGCACTACCCCATG GTGGAATACTGCACCCCCACCACCAGCGGCGAGGATGTGCGGGACTT CGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCCA AGCACCCCCGGATGGGCTACCTGCCCCGTGCAGACAGTGCTGGAAGGC GACAACATGGAAACCTGA
SEQ ID NO:104 - BXA- 212372- J4V4	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA AACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCTTGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACAGAGTTTAAAT GGATCTCACCTACCTGACCGAGATCACCCACGTGTCCAGGCTCTGC

SEQ ID NO and Description	Sequence
	TGGAAGTGGAAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGAC TTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGG ACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCATCCACAGCAAG AAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAAAGAGTGAAGCT GCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGTGAAC AAGATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAA AGTGGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAAGTGGCTG ACCGAGGCCGAGCAGTTCCTGAGAAAGACCCAGATCCCCGAGAACT GGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGG CATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGAATGCCACCGGCG AGGAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCCTGCA GGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAG CAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAACAGGCCCCCTGGCC TGACCACAATCGGCGCCAGCCCTACCCAGACCGTGACCCTGGTGACA CAGCCCGTGGTGACAAAAGAGACAGCCATCAGCAAGCTGGAAATGC CCAGCAGCCTGATGCTGGAAAGCGACCAAGTGGAAAGCGGCTGCACCT GAGCCTGCAGGAAGTGTGGTCTGGCTGCAGCTGAAGGACGACGAG CTGAGCAGACAGGCCCCCATCGGCGGCGATTTCGCCCGCGTGCAAG ACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGCTGAAAACAAAA GAACCCGTGATCATGAGCACCCCTGGAAACCGTGCGGATCTTCCTGAC CGAGCAGCCCCTGGAAGGCCTGGAAAAGCTGTACCAGGAACCCAGA GAGCTGCCCCCGAGGAACGGGCCCAGAACGTGACCAGACTGCTGC GGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAGAAGCTGAACCT GCACAGCGCCGACTGGCAGCGGAAGATCGACGAGACACTGGAACGG CTGCAGGAAGTGCAGGAGGCCACCGACGAGCTGGACCTGAAGCTGA GACAGGCCGAAGTATCAAGGGCAGCTGGCAGCCCGTGGGCGACCT GCTGATCGACTCCCTGCAGGACCACCTGGAAAAAGTGAAGGCCCTGC GGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTGAACGA CCTGGCCCCGCGAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTACA ACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAG GTGGCCGTGGAAGATAGAGTGCAGGAGCTGCACGAGGCCACAGAG ACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCTCTGTGCAGGGA CCCTGGGAGAGAGCCATCAGCCCCAACAAAGGTGCCCTACTACATCAA CCACGAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAG CTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCCTA CCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTGCCTGG ATCTGCTGAGCCTGAGCGCCGCTGCGACGCCCTGGACCAGCACAAAC CTGAAGCAGAACGACCAGCCATGGATATCCTGCAGATCATCAACTG CCTGACCACAATCTACGACAGGCTGGAACAGGAACACAACAATCTG GTCAACGTGCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAA TGTGTACGACACCGGCCGACCGGCAGAAATCCGGGTGCTGAGCTTCA AGACCGGCATCATCAGCCTGTGCAAGGCCACCTGGAAGATAAGTAC CGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTGCGACCA

SEQ ID NO and Description	Sequence
	GCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGAC AGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGCCCAGC GTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGC CGCCCTGTTCTTGACTGGATGAGACTGGAACCCAGAGCATGGTGT GGCTGCCCCTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCAAGCAC CAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCG GTACAGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGCT TCTTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGCACTACCCCATG GTGGAATACTGCACCCCCACCACCAGCGGCGAGGATGTGCGGGACTT CGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCCA AGCACCCCCGGATGGGCTACCTGCCCCTGCAGACAGTGCTGGAAGGC GACAACATGGAAACCTGA
SEQ ID NO:105 - BXA- 212372-J4	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACAGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCAGCTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCTTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACAGCTACGTGCC CAGCACCTACCTGACCGAGATCACCCACGTGTCCCAGGCTCTGCTGG AAGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGACTTC

SEQ ID NO and Description	Sequence
	GAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGACT CCCTGCAGCAGTCCAGCGGCCGGATCGACATCATCCACAGCAAGAAA ACAGCCGCCCTGCAGTCCGCCACCCCGTGGAAAGAGTGAAGCTGCA GGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAG ATGTACAAGGACCGGCAGGGCAGATTCGACCGCAGCGTGGAAAAGT GGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAAGTGGCTGACC GAGGCCGAGCAGTTCTTGAGAAAGACCCAGATCCCCGAGAAGTGGG AGCACGCCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCAT CGGCCAGCGGCAGACAGTGGTCCGCACCCTGAATGCCACCGGCGAG GAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCCTGCAGG AAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGAAGCA GCTGAGCGACCGGAAGAAGCGGCTGGAAGAACAGGCCCCTGGCCTG ACCACAATCGGCGCCAGCCCTACCCAGACCGTGACCCTGGTGACACA GCCCCGTGGTGACAAAAGAGACAGCCATCAGCAAGCTGGAAATGCC AGCAGCCTGATGCTGGAAAGCGACCAGTGAAGCGGCTGCACCTGA GCCTGCAGGAAGTCTGGTCTGGCTGCAGCTGAAGGACGACGAGCTG AGCAGACAGGCCCCCATCGGCGGCGATTTCGCCCGCTGCAGAAACA GAACGACGTGCACCGGGCCTTCAAGCGCGAGCTGAAAACAAAAGAA CCCGTGATCATGAGCACCTGGAAACCGTGCGGATCTTCCTGACCGA GCAGCCCCTGGAAGGCCTGGAAAAGCTGTACCAGGAACCCAGAGAG CTGCCCCCGAGGAACGGGCCCAGAACGTGACCAGACTGCTGCGGA AGCAGGCCGAAGAGGTCAACACCGAGTGGGAGAAGCTGAACCTGCA CAGCGCCGACTGGCAGCGGAAGATCGACGAGACACTGGAACGGCTG CAGGAACTGCAGGAGGCCACCGACGAGCTGGACCTGAAGCTGAGAC AGGCCGAAGTGATCAAGGGCAGCTGGCAGCCCGTGGGCGACCTGCT GATCGACTCCCTGCAGGACCACCTGGAAAAAGTGAAGGCCCTGCGG GGCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTGAACGACCT GGCCCGGCAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTACAACC TGTCCACCCTGGAAGATCTGAACACCCGGTGAAGCTGCTGCAGGTG GCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCACAGAGACT TTGGCCCTGCCAGCCAGCACTTCCTGAGCACCTCTGTGCAGGGACCC TGGGAGAGAGCCATCAGCCCCAACAAGGTGCCCTACTACATCAACCA CGAGACACAGACCACCTGTTGGGACCACCCAAGATGACCGAGCTGT ACCAGAGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCCTACCGG ACCGCCATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTGCCTGGATCT GCTGAGCCTGAGCGCCGCTGCGACGCCCTGGACCAGCACAACTGA AGCAGAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTG ACCACAATCTACGACAGGCTGGAACAGGAACACAACAATCTGGTCA ACGTGCCCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTG TACGACACCGGCCGGACCGGCAGAAATCCGGGTGCTGAGCTTCAAGAC CGGCATCATCAGCCTGTGCAAGGCCACCTGGAAGATAAGTACCGCT ACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTTCGACCAGCGG AGACTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGACAGCT

SEQ ID NO and Description	Sequence
	GGGCGAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGCCCAGCGTGC GGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCC CTGTTCCCTGGACTGGATGAGACTGGAACCCCGAGAGCATGGTGTGGCT GCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCAAGCACCAGG CCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTAC AGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGCTTCTT CAGCGGCAGAGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTG GAATACTGCACCCCCACCACCAGCGGCGAGGATGTGCGGGACTTCGC CAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCCAAGC ACCCCCGGATGGGCTACCTGCCCCGTGCAGACAGTGCTGGAAGGCGAC AACATGGAAACCTGA
SEQ ID NO:106 - BXA- 212372	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACAGTGGAACCTGGTGA ACATCGGCAGCACCGACATCGTGACGGCAACCACAAGCTGACCCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCTCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACATCCACACCGT GCGGGAAGAGACAATGATGGTGATGACCGAGGACATGCCCCCTGGAA ATCAGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCACGTGTC CCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGT

SEQ ID NO and Description	Sequence
	GCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAA GAATATCAAGGACTCCCTGCAGCAGTCCAGCGGCCGATCGACATCA TCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAA AGAGTGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGG AGAAAGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCG CAGCGTGGAAAAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCA ACCAGTGGCTGACCGAGGCCGAGCAGTTCTGAGAAAGACCCAGAT CCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAG CTGCAGGACGGCATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGA ATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGC CAGCATCCTGCAGGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAG GAAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAAC AGGCCCCTGGCCTGACCACAATCGGCGCCAGCCCTACCCAGACCGTG ACCCTGGTGACACAGCCCCTGGTGACAAAAGAGACAGCCATCAGCA AGCTGGAAATGCCCAGCAGCCTGATGCTGGAAAGCGACCAAGTGAA GCGGCTGCACCTGAGCCTGCAGGAAGTGTGGTCTGGCTGCAGCTGA AGGACGACGAGCTGAGCAGACAGGCCCCCATCGGCGGCGATTTCCTCC GCCGTGCAGAAACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGC TGAAAACAAAAGAACCCGTGATCATGAGCACCCCTGGAAACCGTGCG GATCTTCCTGACCGAGCAGCCCCCTGGAAGGCCTGGAAAAGCTGTACC AGGAACCCAGAGAGCTGCCCCCGAGGAACGGGCCAGAACGTGAC CAGACTGCTGCGGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAG AAGCTGAACCTGCACAGCGCCGACTGGCAGCGGAAGATCGACGAGA CACTGGAACGGCTGCAGGAAGTGCAGGAGGCCACCGACGAGCTGGA CCTGAAGCTGAGACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCC GTGGGCGACCTGCTGATCGACTCCCTGCAGGACCACCTGGAAAAAGT GAAGGCCCTGCGGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCC CACGTGAACGACCTGGCCCCGGCAGCTGACCACCCTGGGCATCCAGCT GAGCCCCTACAACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGAA AGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGA GGCCACAGAGACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCT CTGTGCAGGGACCCCTGGGAGAGAGCCATCAGCCCCAACAAAGGTGCC CTACTACATCAACCACGAGACACAGACCACCTGTTGGGACCACCCCA AGATGACCGAGCTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGG TTCAGCGCCTACCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGC TCTGTGCCTGGATCTGCTGAGCCTGAGCGCCGCTGCGACGCCCTGG ACCAGCACAACTGAAGCAGAACGACCAGCCCATGGATATCCTGCA GATCATCAACTGCCTGACCACAATCTACGACAGGCTGGAACAGGAAC ACAACAATCTGGTCAACGTGCCCCCTGTGCGTGGACATGTGCCTGAAT TGGCTGCTGAATGTGTACGACACCGGCCGGACCGGCAGAAATCCGGGT GCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCAAGGCCACCTGG AAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGC TTTTGCGACCAGCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCA

SEQ ID NO and Description	Sequence
	GATCCCCAGACAGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACA TTGAGCCCAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCC GAGATCGAGGCCGCCCTGTTCTTGGAAGTGGATGAGACTGGAACCCCA GAGCATGGTGTGGCTGCCCCTGCTGCATCGGGTGGCCGCTGCCGAGA CAGCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATC ATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATCTG CCAGAGCTGCTTCTTCAGCGGCAGAGTGCCAAGGGCCACAAAATGC ACTACCCCATGGTGGAATACTGCACCCCCACCACCAGCGGCGAGGAT GTGCGGGACTTCGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGC GGTACTTTGCCAAGCACCCCCGGATGGGCTACCTGCCCCTGCAGACA GTGCTGGAAGGCGACAACATGGAAACCTGA

SEQ ID NO:100 and SEQ ID NO: 101 encode the same miniaturized dystrophin, except that the SEQ ID NO: 100 is codon optimized vis-à-vis SEQ ID NO: 101.

In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 100, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 101, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 102, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least

about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 103, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In some
5 embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 104, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide
10 sequence has at least one dystrophin activity. In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 105, wherein the miniaturized
15 dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity.

Non-Coding Polynucleotides

In some aspects, provided herein are nucleic acid molecules, *e.g.*, DNA or RNA, comprising a nucleotide sequence encoding a miniaturized dystrophin polypeptide.

20 In some embodiments, the nucleic acid molecules disclosed herein comprise non-coding components. In some embodiments, the nucleic acid molecules disclosed herein comprise promoters. Certain exemplary regulatory sequences for mammalian host cell expression include viral elements that direct high levels of protein expression in mammalian cells, such as promoters and/or enhancers derived from cytomegalovirus
25 (CMV), Simian Virus 40 (SV40), adenovirus, (*e.g.*, the adenovirus major late promoter (AdMLP) and polyoma. Alternatively, nonviral regulatory sequences can be used, such as the ubiquitin promoter or β -globin promoter. Still further, regulatory elements composed of sequences from different sources may be used, such as the SRa promoter system, which contains sequences from the SV40 early promoter and the long terminal repeat of human T
30 cell leukemia virus type 1 (Takebe, Y. *et al.*, *Mol. Cell. Biol.* 8:466-472 (1988)). In certain

embodiments, the regulatory sequence comprises a tissue specific promoter. In some embodiments, the tissue specific promoter drives expression of the gene of interest in a tissue selected from the group consisting of heart, liver, lungs, eyes, nervous system, lymphatic system, central nervous system, neuronal cells, muscle and stem cells.

5 In some embodiments, the promoters disclosed herein are tissue-specific promoters. In some embodiments, the promoter drives expression of the therapeutic protein in hepatocytes, muscle cells, endothelial cells, sinusoidal cells, or neuronal cells, or any combination thereof. In some embodiments, the promoter is selected from the group consisting of C5-12(T) promoter, MLC2v-cTNT455 promoter, a synapsin 1 gene promoter,
10 a mouse thyretin promoter (mTTR), an endogenous human factor VIII promoter (F8), a human alpha-1-antitrypsin promoter (hAAT), a human albumin minimal promoter, a mouse albumin promoter, a tristetraprolin (TTP) promoter, a CASI promoter, a CAG promoter, a cytomegalovirus (CMV) promoter, an α 1-antitrypsin (AAT) promoter, a muscle creatine kinase (MCK) promoter, a myosin heavy chain alpha (α MHC) promoter, a myoglobin
15 (MB) promoter, a desmin (DES) promoter, a SPc5-12 promoter, a 2R5Sc5-12 promoter, a dMCK promoter, a tMCK promoter, an α -synuclein promoter and a phosphoglycerate kinase (PGK) promoter. In some embodiments, the promoter is the C5-12(T) promoter.

In some embodiments, the nucleic acid molecules disclosed herein comprise an intronic sequence. In some embodiments, the intronic sequence is positioned 5' to the
20 nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the intronic sequence is positioned 3' to the promoter. In some embodiments, the intronic sequence comprises a synthetic intronic sequence. In some embodiments, the intronic sequence is an SV40 intronic sequence.

In some embodiments, the nucleic acid molecules disclosed herein comprise a post-
25 transcriptional regulatory element. In some embodiments, the post-transcriptional regulatory element is positioned 3' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the post-transcriptional regulatory element comprises a mutated woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), a microRNA binding site, or a DNA nuclear targeting sequence, or any
30 combination thereof.

In some embodiments, the nucleic acid molecules disclosed herein comprise a 3'UTR poly(A) tail sequence. In some embodiments, the 3'UTR poly(A) tail sequence is selected from the group consisting of bGH poly(A), actin poly(A), hemoglobin poly(A), dystrophin poly(A), and any combination thereof. In some embodiments, the 3'UTR poly(A) tail sequence comprises nucleotides from the N-terminal portion of the endogenous dystrophin 3'UTR. In some embodiments, the 3'UTR poly(A) tail sequence comprises the 25 nucleotides from the N-terminal portion of the endogenous dystrophin 3'UTR.

In some embodiments, the nucleic acid molecules disclosed herein comprise an enhancer sequence. In some embodiments, the nucleic acid molecules disclosed herein comprise a first inverted terminal repeat (ITR) and/or a second ITR. In some embodiments, the first ITR and the second ITR are identical. In some embodiments, the first ITR and/or the second ITR are derived from adeno-associated virus. In some embodiments, the first ITR is derived from adeno-associated virus, and the second ITR is derived from adeno-associated virus.

It is further recognized that the nucleic acid molecule can comprise additional elements that aid in the translation of the polypeptide. Such sequences include, for example, Kozak sequences attached to the 5' end of the polynucleotide encoding polypeptide. The Kozak consensus sequence is a sequence which occurs on eukaryotic mRNA that plays a role in the initiation of the translation process and has the consensus (gee)gccRccAUGG (SEQ ID NO:107); wherein (1) a lower case letter denotes the most common base at a position where the base can nevertheless vary; (2) upper case letters indicate highly-conserved bases, i.e. the 'AUGG' sequence is constant or rarely, if ever, changes, with the exception being the IUPAC ambiguity code 'R' which indicates that a purine (adenine or guanine) is normally observed at this position; and (3) the sequence in brackets ((gee)) is of uncertain significance.

In one non-limiting embodiment, the nucleic acid molecule comprises a functional variant or fragment thereof of a Kozak sequence. A functional variant or fragment thereof of a Kozak sequence will retain the ability to increase translation of the protein when compared to the level of translation from a sequence lacking the leader. Such a functional fragment can comprise at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40 continuous nucleotides of a Kozak sequence or the sequence set forth in SEQ ID NO:107

or SEQ ID NO:108 (gccaccATGG). Alternatively, a functional variant can comprise at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the Kozak sequence or the sequence set forth in SEQ ID NO:107 or SEQ ID NO:108.

In some embodiments, a nucleotide sequence of the present invention driving expression of a miniaturized dystrophin polypeptide comprises the sequence shown in Table 10.

Table 10: Nucleotide sequence (and domain structure thereof) driving expression of and encoding miniaturized dystrophin polypeptide BXA-220931.

SEQ ID NO:	Description	Nucleotide Sequence
109	C5-12(T) Promoter	CCGCCTTCGGCACCATTCTCAGACACCCAAATATGGCGAC GGGTGAGGAATGGTGGGGAGTTATTTTAGAGCGGTGAGGAA GGTGGGCAGGCAGCAGGTGTTGGCGCTCTAAAAATAACTCCC GGGAGTTATTTTAGAGCGGAGGAATGGTGGACACCCAAATA TGGCGACGGTTCCTACCCGTCGCCATATTTGGGTGTCCGCC TCGGCCGGGGCCGCATTCTGGGGGCCGGGCGGTGCTCCCGC CCGCCTCGATAAAAGGCTCCGGGGCCGGCGGCGGCCACGA GCTACCCGGAGGAGCGGGAGGCACGCGT
110	SV40 Intron	CTCTAAGGTAAATATAAAATTTTAAGTGTATAATGTGTTAA CTACTGATTCTAATTGTTTCTCTCTTTTAGATTCCAACCTTTG AACTGATCTAGACCACC
111	Coding Sequence for miniaturized Dystrophin BXA-220931	ATGCTTTGGTGGGAAGAAGTCGAGGACTGCTACGAGCGCGAG GACGTGCAGAAGAAAACCTTCACCAAATGGGTCAACGCCCA GTTTCAGCAAGTTCGGCAAGCAGCACATCGAGAACCTGTTTCAG CGACCTGCAGGATGGCAGAAGGCTGCTGGATCTGCTGGAAGG CCTGACAGGCCAGAAGCTGCCTAAAGAGAAGGGCAGCACAA GAGTGCACGCCCTGAACAACGTGAACAAGGCCCTGAGAGTG CTGCAGAACAAACGTGGACCTGGTCAACATCGGCAGCACC GACATCGTGGACGGCAATCACAACTGACCCTGGGCCTGATC TGGAACATCATCCTGCACTGGCAAGTGAAGAACGTGATGAAG AACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGAT TCTGCTGAGCTGGGTCCGACAGAGCACCCGGAACCTACCCTCA AGTGAACGTGATCAACTTCACCACCTCTTGGAGCGACGGACT GGCCCTGAATGCCCTGATTACAGCCACAGACCTGACCTGTT CGACTGGAATAGCGTCGTGTGTCAGCAGAGCGCCACACAGAG ACTGGAACACGCCTTCAATATCGCCAGATACCAGCTGGGCAT CGAGAACTGCTGGACCCCGAGGATGTGGACACCACCTATCC TGACAAGAAATCCATCCTCATGTACATCACCAGCCTGTTCCA GGTGCTGCCCCAGCAAGTGTCTATCGAGGCCATTCAAGAGGT CGAGATGCTGCCCAGACCTCCTAAAGTGACCAAAGAGGAAC

		ACTTCCAGCTGCACCACCAGATGCACTACTCTCAGCAGATCA CCGTGTCTCTGGCCAGGGCTACGAGAGAACAAGCAGCCCCA AGCCTCGGTTCAAGAGCTACGCCTATACACAGGCCGCCTACG TGACCACCAGCGATCCCACAAGAAGCCCATTTCGAAGCCAGC ATCTGGAAGCCCCTGAGGACAAGAGCTTTGGCAGCAGCCTGA TGGAAGCGAAGTGAACCTGGATAGATACCAGACAGCCCTG GAAGAGGTGCTGTCTTGGCTGCTGTCTGCCGAAGATACACTG CAGGCTCAGGGCGAGATCAGCAACGACGTGGAAGTGGTCAA GGACCAGTTTCACACCCACGAGGGCTACATGATGGACCTGAC AGCCCATCAGGGCAGAGTGGGCAATATCCTGCAGCTGGGCTC TAAGCTGATCGGCACAGGCAAGCTGAGCGAGGACGAAGAGA CAGAGGTGCAAGAGCAGATGAACCTGCTGAACAGCAGATGG GAGTGTCTGAGAGTGGCCAGCATGGAAAAGCAGAGCAACCT GCACCGGGTCCTGATGGATCTGCAGAATCAGAAGCTGACCGA GATCACCCACGTGTACAGGCCCTGCTTGAAGTGGAACAGCT GCTGAACGCCCCTGATCTGTGCGCCAAGGACTTCGAGGATCT GTTCAAGCAAGAGGAAAGCCTGAAGAATATCAAGGACTCTCT GCAGCAGTCCAGCGGCCGGATCGACATCATCCACAGCAAGA AAACAGCTGCCCTGCAGTCCGCCACACCTGTGGAAGAGTGA AACTGCAAGAGGCCCTGTCTCAGCTGGACTTCCAGTGGGAGA AAGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGAC CGCTCTGTGGA AAAAATGGCGGAGATTCCACTACGACATCAAG ATCTTCAACCAAGTGGCTGACAGAGGCCGAGCAGTTCTTGAGA AAGACACAGATCCCCGAGAACTGGGAGCACGCCAAGTACAA GTGGTATCTGAAAGAACTGCAGGACGGCATCGGCCAGAGGC AGACAGTCGTTAGAACA CTGAATGCCACCGGCGAGGAAATC ATCCAGCAGAGCAGCAAGACCGACGCCAGCATCCTGCAAGA GAAGCTGGGCAGCCTGAACCTGAGATGGCAAGAAGTGTGCA AGCAGCTGTCCGACCGGAAGAAGAGGCTGGAAGAACAGGCC CCTGGCCTGACAACAATCGGAGCCTCTCTACACAGACCGTG ACACTGGTCACACAGCCCGTGGTCACCAAAGAGACAGCCATC AGCAAGCTGGA AATGCCCTCTAGCCTGATGCTCGAGAGCGAC CAGTGGAAGAGACTGCACCTGTCTCTGCAAGAGCTGCTCGTG TGGCTGCAGCTGAAGGACGATGAACTGAGCAGACAGGCCCC AATCGGAGGCGATTTTCCTGCCGTGCAGAAACAGAACGACGT GCACAGAGCCTTCAAGCGGGA ACTGAAAACAAAAGAACCCG TGATCATGAGCACCCCTGGAAACCGTGCGGATCTTCCTGACAG AGCAGCCTCTCGAAGGCCTGGAAAAGCTGTACCAAGAGCCTA GAGAGCTGCCTCCTGAGGAACGGGCCCAGAATGTGACCAGA CTGCTGAGAAAGCAGGCCGAAGAGGTCAACACCGAATGGGA GAAGCTGAACCTGCACAGCGCCGACTGGCAGAGAAAGATCG ACGAGACACTGGAACGGCTGCAAGAACTCCAAGAAGCCACC GACGAGCTGGACCTGAAACTGAGGCAGGCTGAAGTGATCAA AGGCAGCTGGCAGCCAGTGGGCGACCTGCTGATTGATAGTCT GCAGGACCACCTGGAAAAAGTGAAGGCCCTGCGGGGAGAGA TCGCCCCACTGAAAGAAAACGTGTCCACGTGAACGACCTGG
--	--	---

		CCAGACAGCTGACAACCCTGGGAATCCAGCTGTCCCCTTACA ACCTGTCCACACTGGAAGATCTGAACACCCGGTGGAAACTGC TCCAGGTGGCCGTGGAAGATAGAGTGCACAGCTGCACGAG GCCCACAGAGATTTTGGACCAGCCAGCCAGCACTTCCTGTCT ACATCTGTGCAAGGCCCTTGGGAGAGAGCTATCAGCCCTAAC AAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGT TGGGATCACCCCAAGATGACCGAGCTGTATCAGAGCCTGGCC GACCTGAACAATGTGCGCTTTAGCGCCTACCGGACCGCCATG AAGCTGCGGAGACTGCAGAAAGCCCTGTGTCTGGACCTGCTG TCTCTGTCTGCAGCCTGTGATGCCCTGGACCAGCACAACCTG AAGCAGAACGACCAGCCTATGGACATCCTCCAGATCATCAAC TGCCTGACCACCATCTACGACCGGCTGGAACAAGAGCACAAC AACCTCGTGAATGTGCCCTGTGCGTGGACATGTGTCTGAAC TGGCTGCTGAATGTGTACGACACCGGCAGAACCGGCAGGATC AGAGTGCTGAGCTTCAAGACCGGCATCATCTCCCTGTGCAA GCCCACCTCGAGGACAAGTACAGATACCTGTTCAAACAGGTG GCCAGCTCCACCGGCTTTTGCATCAAAGAAGGCTGGGCCTG CTGCTGCACGACAGCATCCAGATTCCTAGACAGCTGGGCGAA GTGGCCTCCTTCGGCGGATCTAATATTGAGCCTAGCGTGC GG AGCTGCTTCCAGTTCGCCAACAAAGCCTGAGATCGAGGCC GCTCTGTTCTTGATTGGATGCGCCTGGAACCTCAGAGCATG GTTTGGCTGCCTGTGCTGCATAGAGTGGCCGCTGCCGAAACA GCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCC ATCATCGGCTTCCGGTACAGATCCCTGAAGCACTTCAACTAC GATATCTGCCAGAGCTGTTTCTTCTCTGGCCGCGTGGCCAAGG GCCACAAAATGCACTACCCCATGGTGGAATACTGCACCCCTA CCACATCTGGCGAAGATGTGCGGGATTTCGCCAAGGTGCTGA AAAACAAGTTCCGGACCAAGCGGTACTTCGCTAAGCACCCCA GAATGGGCTATCTGCCCCGTGCAGACAGTGCTCGAGGGCGATA ACATGGAAACCTGA
112	3' UTR	GAAGTCTTTTCCACATGGCAGATGA
113	PolyA	AATAAAAGATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTT TGTGTG

In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, 15 at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to the combined sequence of SEQ ID NO: 109 to 113.

Heterologous Moieties

In some embodiments, the polypeptides of the present disclosure can further comprise an additional element, *e.g.*, heterologous moiety. Such elements can aid in the expression of the polypeptide, aid in the secretion of the polypeptide, improve the stability
5 of the polypeptide, allow for more efficient purification of the polypeptide, and/or modulate the activity of the polypeptide. In some embodiments, the heterologous moiety is a polypeptide moiety. In other embodiments, the heterologous moiety is a non-polypeptide moiety.

In some embodiments, the polypeptide comprises a heterologous moiety fused to
10 the polypeptide.

In some embodiments, the polypeptide disclosed herein comprises one or more additional heterologous moieties. In some embodiments, the heterologous moieties are half-life extending moieties. In some embodiments, the heterologous moiety comprises albumin or a fragment thereof, an immunoglobulin Fc region, the C-terminal peptide (CTP)
15 of the β subunit of human chorionic gonadotropin, a PAS sequence, a HAP sequence, a transferrin or a fragment thereof, or an albumin-binding moiety or a derivative thereof, or any combination thereof.

In some embodiments, the polypeptides disclosed herein comprise one or more additional heterologous moieties. In some embodiments, the heterologous moieties are
20 half-life extending moieties. In some embodiments, the heterologous moiety comprises albumin, an immunoglobulin constant region or a portion thereof, an immunoglobulin-binding polypeptide, an immunoglobulin G (IgG), albumin-binding polypeptide (ABP), a PASylation moiety, a HESylation moiety, XTEN, a PEGylation moiety, or an Fc region, or any combination thereof.

Cells

In certain aspects, provided herein are cells (*e.g.*, host cells) expressing (*e.g.*, recombinantly) proteins described herein and expression vectors comprising nucleotides that encode proteins described herein.

In some embodiments, the host cell comprises the nucleic acid molecules described
30 herein. In some embodiments, the host cell comprises the vectors described herein.

In some embodiments, the host cell is a eukaryotic cell. In some embodiments, the host cell is selected from the group consisting of a mammalian cell, an insect cell, a yeast cell, a transgenic mammalian cell, and a plant cell. In some embodiments, the host cell is a prokaryotic cell. In some embodiments, the prokaryotic cell is a bacterial cell.

5 In some embodiments, the host cell is a mammalian cell. Such mammalian host cells include but are not limited to CHO, VERO, BHK, Hela, MDCK, HEK 293, NIH 3T3, W138, BT483, Hs578T, HTB2, BT2O and T47D, NS0 (a murine myeloma cell line that does not endogenously produce any immunoglobulin chains), CRL7030, COS (*e.g.*, COS1 or COS), PER.C6, VERO, HsS78Bst, HEK-293T, HepG2, SP210, R1.1, B-W, L-M, BSC1,
10 BSC40, YB/20, BMT10, HBK, NSO, HT1080 and HsS78Bst cells.

Vectors

Adeno-Associate Virus (AAV)

Overview

Provided herein are vectors (*e.g.*, expression vectors) comprising nucleic acid
15 molecules comprising nucleotide sequences encoding a miniaturized dystrophin protein for recombinant expression in host cells and cells targeted for therapeutic intervention. The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked; or an entity comprising such a nucleic acid molecule capable of transporting another nucleic acid. One type of vector is
20 a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors, or polynucleotides that are part of vectors, are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication, and episomal
25 mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").
30 In general, expression vectors of utility in recombinant DNA techniques are often in the

form of plasmids. In the present specification, "plasmid" and "vector" can sometimes be used interchangeably, depending on the context, as the plasmid is the most commonly used form of vector. However, also disclosed herein are other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

In some embodiments, the polynucleotides disclosed herein are expressed using an adeno-associated virus (AAV). AAV is a nonenveloped, single-stranded DNA virus of the *Parvoviridae* family. In contrast to most other members of the *Parvoviridae* family, AAV is replication defective and is only able to replicate efficiently in the presence of a helper virus such as adenovirus or herpes virus.

AAV was first discovered in the mid 1960's as a contaminant of viral preparations of adenovirus. *See* Atchison R.W., Casto B.C., Hammon W.M., *Science*. 149(3685), 754-756 (1965). Since then, progressively safer and more effective methods to use AAV as a recombinant DNA vector have been developed. *See, e.g.*, Hermonat P.L. and Muzyczka N., *Proc Natl Acad Sci USA*. 81(20):6466-6470 (1984); Laughlin C.A. *et al.*, *Gene*, 23(1): 65-73 (1983); Matsushita T. *et al.*, *Gene Ther.* 5(7):938-945 (1998); and Xiao X. *et al.*, *Journal of Virology* 72(3):2224-2232 (1998). Low numbers of AAV genomes have been shown to integrate into the host chromosome. *See* Cheung A.K., Hoggan M.D., Hauswirth W.W. *et al.*, Integration of the adeno-associated virus genome into cellular DNA in latently infected human detroit 6 cells, *J. Virol.* 33:739-748 (1980). AAV is immunologically distinct from any known adenovirus antigen. The AAV capsid contains a single-stranded DNA (ssDNA) genome. *See* Rose J.A., Berns K.I., Hoggan M.D. *et al.*, *Proc. Natl. Acad. Sci. USA* 64:863-869 (1969).

AAV has a single stranded, 4.7 kb DNA genome encoding a replication (rep) gene and a capsid (cap) genes flanked by two inverted terminal repeats (ITRs). It is predominantly non-integrating, and forms stable episomes in non-dividing tissue. In spite of its high seroprevalence in the adult human population, AAV has not been associated with any human disease. *See* Gonçalves M., *Virol. J.* 2, 43 (2005). AAV's stable expression in tissues, its lack of pathogenicity, and its ease of high titer production have made it a very attractive vector and popular gene transfer platform.

A recombinant AAV (rAAV) is a genetically manipulated AAV in which typically part or all of the rep and cap genes have been replaced with heterologous transgene sequences. Recombinant AAVs too can trigger long-term transgene expression in postmitotic cells, most likely because the recombinant AAV genome persist as largely circular episomes within the nucleus. rAAVs' only DNA *cis*-element required for the production of rAAVs is the AAV inverted terminal repeats (ITRs), whereas *rep*, *cap*, and adenoviral helper genes can be provided in *trans*. Thus, in some embodiments disclosed herein, rAAVs contain only heterologous transgene DNA flanked by the ITRs, and this genome is encapsidated within a serotype-specific AAV capsid.

AAV possesses unique features that make it attractive as a vector system for delivering foreign DNA into cells. AAV infection of cells in culture has generally been noncytopathic, and natural infection of humans and other animals is silent and asymptomatic. Moreover, AAV infects many different types of mammalian cells allowing the possibility of targeting many different tissues *in vivo*. AAV also possesses additional advantages that make it a particularly attractive viral system for gene delivery, including the promotion of an immune response that is relatively mild compared to other forms of gene delivery, and persistent expression in both dividing and quiescent cells based on non-integrating, episomal vector DNA. Also, AAV withstands the conditions used to inactivate adenovirus (56° to 65° C for several hours), making cold preservation of rAAV-based vaccines less critical.

Replication of the viral DNA is not required for integration into the host-cell genome, and thus helper virus is not required for this process. The AAV proviral genome is infectious as cloned DNA in plasmids which makes construction of recombinant genomes feasible. Furthermore, because the signals directing AAV replication, genome encapsidation and integration are contained within the ITRs of the AAV genome, the internal approximately 4.7 kb of the genome (encoding the replication and structural capsid proteins, rep-cap) can thus be replaced with foreign DNA such as a gene cassette containing a promoter, a DNA of interest and a polyadenylation signal.

AAV vectors can include additional elements that function in *cis* or in *trans*. In particular embodiments, an AAV vector that includes a vector genome also has one or more inverted terminal repeat (ITR) sequences that flank the 5' or 3' terminus of the donor

sequence; an expression control element that drives transcription (*e.g.*, a promoter or enhancer) of the donor sequence, such as a constitutive or regulatable control element, or tissue-specific expression control element; an intron sequence, a stuffer or filler polynucleotide sequence; and/or a poly-Adenine sequence located 3' of the donor sequence.

5 In some embodiments, AAV replicates using a helper virus. A variety of such helper viruses for AAV are known in the art, including adenoviruses, herpesviruses and poxviruses such as vaccinia. The adenoviruses encompass a number of different subgroups, although Adenovirus type 5 of subgroup C is most commonly used. Numerous adenoviruses of human, non-human mammalian and avian origin are known and available
10 from depositories such as the ATCC. Viruses of the herpes family include, for example, herpes simplex viruses (HSV) and Epstein-Barr viruses (EBV), as well as cytomegaloviruses (CMV) and pseudorabies viruses (PRV); which are also available from depositories such as ATCC.

Exemplary AAV vectors include capsid sequences of any of AAV1, AAV2, AAV3,
15 AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8, or a capsid variant of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8. Recombinant AAV vectors of the invention also include AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8, and variants thereof. Particular capsid
20 variants include capsid variants of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8, such as a capsid sequence with an amino acid substitution, deletion or insertion/addition. In one embodiment, the AAV vector is AAV9. In one embodiment, the AAV vector is AAV5. In one embodiment, the AAV vector is AAV8.

25 In some aspects the disclosure relates to AAVs having distinct tissue targeting capabilities (*e.g.*, tissue tropisms). In some embodiments, the variant AAV capsid polypeptides further exhibit increased transduction or tropism in one or more human stem cell types as compared to non-variant parent capsid polypeptides. In some embodiments, the human stem cell types include but are not limited to embryonic stem cells, adult tissue
30 stem cells (*i.e.*, somatic stem cells), bone marrow stem cells, progenitor cells, induced pluripotent stem cells, and reprogrammed stem cells. In some embodiments, adult stem

cells can include organoid stem cells (i.e., stem cells derived from any organ or organ system of interest within the body). In some embodiments, the target tissue of an AAV is gonad, diaphragm, heart, stomach, liver, spleen, pancreas, muscle or kidney. In some embodiments, the AAV targets organs of the body that include, but are not limited to, skin, hair, nails, sense receptors, sweat gland, oil glands, bones, muscles, brain, spinal cord, nerve, pituitary gland, pineal gland, hypothalamus, thyroid gland, parathyroid, thymus, adrenals, pancreas (islet tissue), heart, blood vessels, lymph nodes, lymph vessels, thymus, spleen, tonsils, nose, pharynx, larynx, trachea, bronchi, lungs, mouth, pharynx, esophagus, stomach, small intestine, large intestine, rectum, anal canal, teeth, salivary glands, tongue, liver, gallbladder, pancreas, appendix, kidneys, ureters, urinary bladder, urethra, testes, ductus (vas) deferens, urethra, prostate, penis, scrotum, ovaries, uterus, uterine (fallopian) tubes, vagina, vulva, and mammary glands (breasts). Organ systems of the body include but are not limited to the integumentary system, skeletal system, muscular system, nervous system, endocrine system, cardiovascular system, lymphatic system, respiratory system, digestive system, urinary system, and reproductive system. In some embodiments, transduction and/or tropism of an AAV with variant capsid polypeptides is increased by about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, 65%, about 70%%, about 75%, about 80%, about 85%, about 90%, about 95%, about 99%, or about 100%, by comparison to an AAV having non-variant capsid polypeptides. In some embodiments, transduction and/or tropism is increased by about 5% to about 80%, about 10% to about 70%, about 20% to about 60% or about 30% to about 60%.

Replication, Capsid, and Assembly AAV Genes

The single-stranded genome of AAV comprises three genes, *rep* (Replication), *cap* (Capsid), and *aap* (Assembly). These three genes give rise to at least nine gene products through the use of three promoters, alternative translation start sites, and differential splicing.

The *rep* gene encodes four proteins (Rep78, Rep68, Rep52, and Rep40), which are required for viral genome replication and packaging.

Cap gene expression gives rise to the viral capsid proteins (VP1; VP2; VP3), which form the outer capsid shell that protects the viral genome, as well as being actively involved in cell binding and internalization. It is estimated that the viral coat is comprised of 60 proteins arranged into an icosahedral structure.

5 The *aap* gene encodes the assembly-activating protein (AAP) in an alternate reading frame overlapping the *cap* gene. This nuclear protein is thought to provide a scaffolding function for capsid assembly and plays a role in nucleolar localization of VP proteins in some AAV serotypes.

10 In some embodiments, one or more of the *rep*, *cap*, or *aap* genes are naturally occurring, *e.g.* the *rep*, *cap*, or *aap* genes comprise all or a portion of parvovirus *rep*, *cap*, or *aap* genes. In some embodiments, the one or more of the *rep*, *cap*, or *aap* genes comprise a synthetic sequence.

15 In one embodiment, the *rep* gene comprises a synthetic sequence. In one embodiment, the *cap* gene comprises a synthetic sequence. In one embodiment, the *aap* gene comprises a synthetic sequence. In one embodiment, the *rep* and *cap* genes comprise a synthetic sequence. In one embodiment, the *rep* and *aap* genes comprise a synthetic sequence. In one embodiment, the *cap* and *aap* genes comprise a synthetic sequence. In one embodiment, the *rep*, *cap*, and *aap* genes comprise a synthetic sequence.

20 In some embodiments, *rep* is from an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, *rep* is from the AAV1 genome. In a particular embodiment, *rep* is from the AAV2 genome. In a particular embodiment, *rep* is from the AAV3 genome. In a particular embodiment, *rep* is from the AAV4 genome. In a particular embodiment, *rep* is from the AAV5 genome. In a particular embodiment, *rep* is from the AAV6 genome. In a particular embodiment, *rep* is from the AAV7 genome. In a particular embodiment, *rep* is from the AAV8 genome. In a particular embodiment, *rep* is from the AAV9 genome. In a particular embodiment, *rep* is from the AAV10 genome. In a particular embodiment, *rep* is from the AAV11 genome.

25 In some embodiments, *cap* is from an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, *cap* is from the AAV1 genome. In a particular embodiment, *cap* is from the AAV2 genome. In a particular embodiment, *cap* is from the AAV3 genome. In a particular embodiment, *cap* is from the AAV4 genome. In a particular embodiment, *cap* is from the AAV5 genome. In a particular embodiment, *cap* is from the AAV6 genome. In a particular embodiment, *cap* is from the AAV7 genome. In a particular embodiment, *cap* is from the AAV8 genome. In a particular embodiment, *cap* is from the AAV9 genome. In a particular embodiment, *cap* is from the AAV10 genome. In a particular embodiment, *cap* is from the AAV11 genome.

particular embodiment, *cap* is from the AAV2 genome. In a particular embodiment, *cap* is from the AAV3 genome. In a particular embodiment, *cap* is from the AAV4 genome. In a particular embodiment, *cap* is from the AAV5 genome. In a particular embodiment, *cap* is from the AAV6 genome. In a particular embodiment, *cap* is from the AAV7 genome. In a particular embodiment, *cap* is from the AAV8 genome. In a particular embodiment, *cap* is from the AAV9 genome. In a particular embodiment, *cap* is from the AAV10 genome. In a particular embodiment, *cap* is from the AAV11 genome.

In some embodiments, *aap* is from an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, *aap* is from the AAV1 genome. In a particular embodiment, *aap* is from the AAV2 genome. In a particular embodiment, *aap* is from the AAV3 genome. In a particular embodiment, *aap* is from the AAV4 genome. In a particular embodiment, *aap* is from the AAV5 genome. In a particular embodiment, *aap* is from the AAV6 genome. In a particular embodiment, *aap* is from the AAV7 genome. In a particular embodiment, *aap* is from the AAV8 genome. In a particular embodiment, *aap* is from the AAV9 genome. In a particular embodiment, *aap* is from the AAV10 genome. In a particular embodiment, *aap* is from the AAV11 genome.

It is to be understood that a particular AAV genome described herein could have genes derived from different AAV genomes (*e.g.*, genomes from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and AAV11). Thus, disclosed herein are AAVs that comprise any possible permutation/combination of *rep*, *cap*, or *aap*.

In some embodiments disclosed herein, the AAV is recombinant AAV (rAAV). In some embodiments, the rAAV lacks one or more of the *rep* gene, the *cap* gene, and the *aap* gene. In some embodiments, the rAAV lacks a *rep* gene. In some embodiments, the rAAV lacks a *cap* gene. In some embodiments, the rAAV lacks an *aap* gene. In some embodiments, the rAAV lacks a *rep* gene and lacks a *cap* gene. In some embodiments, the rAAV lacks a *rep* gene and lacks an *aap* gene. In some embodiments, the rAAV lacks a *cap* gene and lacks an *aap* gene. In some embodiments, the rAAV lacks a *rep* gene, a *cap* gene, and an *aap* gene.

In some embodiments disclosed herein, the rAAV is modified so that one or more of the *rep* gene, the *cap* gene, and the *aap* gene is mutated so that expression of one or more

of the AAV genes is modified. In some embodiments, the *rep* gene is mutated. In some embodiments, the *cap* gene is mutated. In some embodiments, the *aap* gene is mutated. In some embodiments, the *rep* gene and the *cap* gene are mutated. In some embodiments, the *rep* gene and the *aap* gene are mutated. In some embodiments, the *cap* gene and the *aap* gene are mutated. In some embodiments, the *cap* gene, the *rep* gene, and the *aap* gene are mutated.

Inverted Terminal Repeats

In certain embodiments, the AAV comprises a first ITR, *e.g.*, a 5' ITR, and second ITR, *e.g.*, a 3' ITR. Typically, ITRs are involved in parvovirus (*e.g.*, AAV) DNA replication and rescue, or excision, from prokaryotic plasmids (Samulski R.J. *et al.*, *Cell* 33(1):135-143 (1983), *Journal of Virology* 61:3096-3101 (1987); Senapathy P. *et al.*, *Journal of Molecular Biology* 179(1):1-20 (1984); Gottlieb J. and Muzyczka N., *Molecular and Cellular Biology* 6(8): 2513-2522 (1988)). In addition, ITRs have been reported to be the minimum sequences required for AAV proviral integration and for packaging of AAV DNA into virions (McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). These elements are essential for efficient multiplication of a parvovirus genome.

In some embodiments, the ITR comprises a naturally occurring ITR, *e.g.*, the ITR comprises all or a portion of a parvovirus ITR. In some embodiments, the ITR comprises a synthetic sequence. In one embodiment, the first ITR or the second ITR comprises a synthetic sequence. In another embodiment, each of the first ITR and the second ITR comprises a synthetic sequence. In some embodiments, the first ITR or the second ITR comprises a naturally occurring sequence. In another embodiment, each of the first ITR and the second ITR comprises a naturally occurring sequence.

In some embodiments, the ITR comprises an ITR from an AAV genome. In some embodiments, the ITR is an ITR of an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, the ITR is an ITR of the AAV2 genome. In another embodiment, the ITR is a synthetic sequence genetically engineered to include at its 5' and 3' ends ITRs derived from one or more of AAV genomes. In some embodiments, the ITRs are derived from the same genome, *e.g.*, from the genome of the same virus, or from

different genomes, *e.g.*, from the genomes of two or more different AAV genomes. In certain embodiments, the ITRs are derived from the same AAV genome. In a specific embodiment, the two ITRs present in the nucleic acid molecule of the invention are the same, and can in particular be AAV2 ITRs. In one particular embodiment, the first ITR and the second ITR are identical.

In some embodiments, the ITRs form hairpin loop structures. In one embodiment, the first ITR forms a hairpin structure. In another embodiment, the second ITR forms a hairpin structure. Still in another embodiment, both the first ITR and the second ITR form hairpin structures.

In some embodiments, an ITR in a nucleic acid molecule described herein is a transcriptionally activated ITR. A transcriptionally-activated ITR can comprise all or a portion of a wild-type ITR that has been transcriptionally activated by inclusion of at least one transcriptionally active element. Various types of transcriptionally active elements are suitable for use in this context. In some embodiments, the transcriptionally active element is a constitutive transcriptionally active element. Constitutive transcriptionally active elements provide an ongoing level of gene transcription, and can be used when it is desired that the transgene be expressed on an ongoing basis. In other embodiments, the transcriptionally active element is an inducible transcriptionally active element. Inducible transcriptionally active elements generally exhibit low activity in the absence of an inducer (or inducing condition), and are up-regulated in the presence of the inducer (or switch to an inducing condition). Inducible transcriptionally active elements can be used when expression is desired only at certain times or at certain locations, or when it is desirable to titrate the level of expression using an inducing agent. Transcriptionally active elements can also be tissue-specific; that is, they exhibit activity only in certain tissues or cell types. Transcriptionally active elements, can be incorporated into an ITR in a variety of ways. In some embodiments, a transcriptionally active element is incorporated 5' to any portion of an ITR or 3' to any portion of an ITR. In other embodiments, a transcriptionally active element of a transcriptionally-activated ITR lies between two ITR sequences. If the transcriptionally active element comprises two or more elements which must be spaced apart, those elements can alternate with portions of the ITR. In some embodiments, a hairpin structure of an ITR is deleted and replaced with inverted repeats of a transcriptional

element. This latter arrangement would create a hairpin mimicking the deleted portion in structure. Multiple tandem transcriptionally active elements can also be present in a transcriptionally-activated ITR, and these can be adjacent or spaced apart. In addition, protein binding sites (*e.g.*, Rep binding sites) can be introduced into transcriptionally active elements of the transcriptionally-activated ITRs. A transcriptionally active element can comprise any sequence enabling the controlled transcription of DNA by RNA polymerase to form RNA, and can comprise, for example, a transcriptionally active element, as defined below.

Transcriptionally-activated ITRs provide both transcriptional activation and ITR functions to the nucleic acid molecule in a relatively limited nucleotide sequence length which effectively maximizes the length of a transgene which can be carried and expressed from the nucleic acid molecule. Incorporation of a transcriptionally active element into an ITR can be accomplished in a variety of ways. A comparison of the ITR sequence and the sequence requirements of the transcriptionally active element can provide insight into ways to encode the element within an ITR. For example, transcriptional activity can be added to an ITR through the introduction of specific changes in the ITR sequence that replicates the functional elements of the transcriptionally active element. A number of techniques exist in the art to efficiently add, delete, and/or change particular nucleotide sequences at specific sites (see, for example, Deng W.P and Nickoloff J.A., *Anal. Biochem.* 200:81-88 (1992)). Another way to create transcriptionally-activated ITRs involves the introduction of a restriction site at a desired location in the ITR. In addition, multiple transcriptionally activate elements can be incorporated into a transcriptionally-activated ITR, using methods known in the art.

By way of illustration, transcriptionally-activated ITRs can be generated by inclusion of one or more transcriptionally active elements such as: TATA box, GC box, CCAAT box, Sp1 site, Inr region, CRE (cAMP regulatory element) site, ATF-1/CRE site, APB β box, APB α box, CArG box, CCAC box, or any other element involved in transcription as known in the art.

Gene of Interest and Other Sequences

Certain aspects of the present disclosure are directed to methods of administering to a subject an AAV therapy. In some embodiments, the AAV comprises a gene of interest (GOI). In some embodiments, the GOI is a nucleic acid molecule comprising a nucleotide sequence as disclosed herein, which encodes a miniaturized dystrophin polypeptide as disclosed herein.

The GOI being expressed can be either a DNA segment encoding a protein, with any necessary control elements (e.g., promoters, operators) desired by the user, or a non-coding DNA segment, the transcription of which produces all or part of some RNA-containing molecule, such as a ribozyme or an anti-sense molecule.

In some embodiments, the AAV comprises more than one GOI. In AAVs with more than one GOI, some embodiments include elements such as IRES or 2A, to co-express them from one promoter. In some embodiments, the AAV comprises two genes of interest separated by an IRES element. In some embodiments, the AAV comprises two genes of interest separated by a 2A element. In some embodiments, the AAV comprises three genes of interest separated by an IRES element between the genes of interest (e.g., GOI-IRES-GOI-IRES-GOI). In some embodiments, the AAV comprises three genes of interest separated by 2A elements between the genes of interest.

In some embodiments, the AAV comprises a regulatory sequence. In some embodiments, the AAV comprises non-coding regulatory DNA. In some embodiments, the AAV genome comprises regulatory sequences that control the expression of the antibody chain genes in a host cell. The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals) that control the transcription or translation of the antibody chain genes. Such regulatory sequences are described, for example, in Goeddel (Gene Expression Technology. Methods in Enzymology 185, Academic Press, San Diego, CA (1990)). It will be appreciated by those skilled in the art that the design of the AAV, including the selection of regulatory sequences, can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. In some embodiments, the AAV genome comprises mRNA splice donor/splice acceptor sites. Certain regulatory sequences for mammalian host cell expression include viral elements that direct high levels of protein expression in mammalian cells, such as promoters and/or enhancers derived from

cytomegalovirus (CMV), Simian Virus 40 (SV40), adenovirus, (*e.g.*, the adenovirus major late promoter (AdMLP) and polyoma. Alternatively, nonviral regulatory sequences can be used, such as the ubiquitin promoter or β -globin promoter. Still further, regulatory elements composed of sequences from different sources, such as the SRa promoter system, which
5 contains sequences from the SV40 early promoter and the long terminal repeat of human T cell leukemia virus type 1 (Takebe, Y. *et al.*, *Mol. Cell. Biol.* 8:466-472 (1988)). In certain embodiments, the regulatory sequence comprises a tissue specific promoter. In some embodiments, the tissue specific promoter drives expression of the gene of interest in a tissue selected from the group consisting of heart, liver, lungs, eyes, nervous system,
10 lymphatic system, muscle and stem cells.

AAV Formulations

In some embodiments, the AAV vector is formulated with a delivery agent. In some embodiments, the delivery agent comprises a lipid nanoparticle. In some embodiments, the delivery agent is selected from the group consisting of liposomes, non-lipid polymeric
15 molecules, endosomes, and any combination thereof.

Non-AAV Vectors

A vector which comprises the above-described polynucleotides operably linked to a promoter is also provided herein. A nucleotide sequence is "operably linked" to an expression control sequence (*e.g.*, a promoter) when the expression control sequence
20 controls and regulates the transcription and translation of that sequence. The term "operably linked" when referring to a nucleotide sequence includes having an appropriate start signal (*e.g.*, ATG) in front of the nucleotide sequence to be expressed and maintaining the correct reading frame to permit expression of the sequence under the control of the expression control sequence and production of the desired product encoded by the sequence. If a gene
25 that one desires to insert into a recombinant nucleic acid molecule does not contain an appropriate start signal, such a start signal can be inserted in front of the gene. A "vector" is a replicon, such as plasmid, phage or cosmid, to which another nucleic acid segment can be attached so as to bring about the replication of the attached segment. The promoter can be, or is identical to, a bacterial, yeast, insect or mammalian promoter.

In some embodiments, the vector can be a plasmid, cosmid, yeast artificial chromosome (YAC), bacteriophage or eukaryotic viral DNA. Other numerous vector backbones known in the art as useful for expressing protein can be employed. Such vectors include, but are not limited to:

5 adenoviral vector, a retroviral vector, poxvirus vector, a baculovirus vector, a herpes viral vector, simian virus 40 (SV40), cytomegalovirus (CMV), mouse mammary tumor virus (MMTV), and Moloney murine leukemia virus. Further, one class of vectors comprises DNA elements derived from viruses such as bovine papilloma virus, polyoma virus, baculovirus, retroviruses, or Semliki Forest virus. Such vectors can be obtained
10 commercially or assembled from the sequences described by methods well-known in the art.

In some embodiments, the vector described herein is formulated with a delivery agent. In some embodiments, the delivery agent comprises a lipid nanoparticle. In some embodiments, the delivery agent is selected from the group consisting of liposomes, non-
15 lipid polymeric molecules, endosomes, and any combination thereof.

Pharmaceutical Compositions

The various polypeptides and polynucleotides disclosed herein (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the polypeptide, or
20 polynucleotides, and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active compounds is well known in the art. Except insofar
25 as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a polypeptide as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a composition comprising a polypeptide as described herein and (b) a pharmaceutically acceptable excipient.

5 In some embodiments, disclosed is a pharmaceutical composition comprising (a) a polynucleotide as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a vector (e.g., rAAV) as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a host cell as described herein and (b) a pharmaceutically acceptable excipient.

10 A pharmaceutical composition of the disclosure is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral, transdermal (topical), and transmucosal, and any combination thereof. Another route of administration includes pulmonary administration. In addition, it can be desirable to administer a therapeutically
15 effective amount of the pharmaceutical composition locally to an area in need of treatment. This can be achieved by, for example, local or regional infusion or perfusion during surgery, topical application, injection, catheter, suppository, or implant (for example, implants formed from porous, non-porous, or gelatinous materials, including membranes, such as sialastic membranes or fibers), and the like. In another embodiment, the
20 therapeutically effective amount of the pharmaceutical composition is delivered in a vesicle, such as liposomes (see, e.g., Langer, Science 249:1527-33, 1990 and Treat *et al.*, in Liposomes in the Therapy of Infectious Disease and Cancer, Lopez Berestein and Fidler (eds.), Liss, N.Y., pp. 353-65, 1989).

In yet another embodiment, the therapeutically effective amount of the
25 pharmaceutical composition can be delivered in a controlled release system. In one example, a pump can be used (see, e.g., Langer, Science 249:1527-33, 1990; Sefton, Crit. Rev. Biomed. Eng. 14:201-40, 1987; Buchwald *et al.*, Surgery 88:507-16, 1980; Saudek *et al.*, N Engl. J Med. 321:574-79, 1989). In another example, polymeric materials can be used (see, e.g., Levy *et al.*, Science 228:190-92, 1985; During *et al.*, Ann. Neural. 25:351-
30 56, 1989; Howard *et al.*, J Neurosurg. 71:105-12, 1989). Other controlled release systems, such as those discussed by Langer (Science 249:1527-33, 1990), can also be used.

Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride, benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); 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, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (*e.g.*, Zn-protein complexes); and/or non-ionic surfactants such as TWEEN™, PLURONICS™ or polyethylene glycol (PEG).

Pharmaceutically acceptable carriers used in parenteral preparations include aqueous vehicles, nonaqueous vehicles, antimicrobial agents, isotonic agents, buffers, antioxidants, local anesthetics, suspending and dispersing agents, emulsifying agents, sequestering or chelating agents and other pharmaceutically acceptable substances. Examples of aqueous vehicles include Sodium Chloride Injection, Ringers Injection, Isotonic Dextrose Injection, Sterile Water Injection, Dextrose and Lactated Ringers Injection. Nonaqueous parenteral vehicles include fixed oils of vegetable origin, cottonseed oil, corn oil, sesame oil and peanut oil. Antimicrobial agents in bacteriostatic or fungistatic concentrations can be added to parenteral preparations packaged in multiple-dose containers which include phenols or cresols, mercurials, benzyl alcohol, chlorobutanol, methyl and propyl p-hydroxybenzoic acid esters, thimerosal, benzalkonium chloride and benzethonium chloride. Isotonic agents include sodium chloride and dextrose. Buffers include phosphate and citrate. Antioxidants include sodium bisulfate. Local anesthetics include procaine hydrochloride. Suspending and dispersing agents include sodium carboxymethylcellulose, hydroxypropyl methylcellulose and polyvinylpyrrolidone. Emulsifying agents include Polysorbate 80 (TWEEN® 80). A sequestering or chelating agent of metal ions includes EDTA. Pharmaceutical carriers also include ethyl alcohol,

polyethylene glycol and propylene glycol for water miscible vehicles; and sodium hydroxide, hydrochloric acid, citric acid or lactic acid for pH adjustment.

Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for
5 injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with
10 acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes, or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous
15 preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELs (BASF; Parsippany, NJ), or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating
20 action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion, and by
25 the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride, in the composition. Prolonged absorption of the injectable compositions can be
30 brought about by including in the composition an agent that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the methods of preparation can be vacuum drying and freeze-drying, which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser that contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer. Systemic administration can also be by transmucosal or transdermal means.

For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions can also be used as pharmaceutically acceptable carriers. These can be

prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used
5 herein refers to physically discrete units suited as unitary dosages for the subject to be treated with each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the disclosure are dictated by and
10 directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such a functional compound for the treatment of individuals. The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Uses and Methods

15 Methods of Producing Miniaturized Dystrophins

Also disclosed herein are methods of producing a miniaturized dystrophin polypeptide, comprising: culturing a host cell described herein under suitable conditions and recovering the miniaturized dystrophin polypeptide.

As used herein, an "isolated" polynucleotide or nucleic acid molecule is one which
20 is separated from other nucleic acid molecules which are present in the natural source (*e.g.*, in a mouse or a human) of the nucleic acid molecule. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. For example, the
25 language "substantially free" includes preparations of polynucleotide or nucleic acid molecule having less than about 15%, 10%, 5%, 2%, 1%, 0.5%, or 0.1% (in particular less than about 10%) of other material, *e.g.*, cellular material, culture medium, other nucleic acid molecules, chemical precursors and/or other chemicals. In a specific embodiment, a nucleic acid molecule(s) encoding a polypeptide described herein is isolated or purified.

The polynucleotides can be obtained, and the nucleotide sequence of the polynucleotides determined, by any method known in the art. Nucleotide sequences encoding polypeptides described herein, *e.g.*, the polypeptides described in Tables 3 and 4, and modified versions of these polypeptides can be determined using methods well known
5 in the art, *i.e.*, nucleotide codons known to encode particular amino acids are assembled in such a way to generate a nucleic acid that encodes the polypeptides. Such a polynucleotide encoding the polypeptide can be assembled from chemically synthesized oligonucleotides (*e.g.*, as described in Kutmeier G. *et al.*, (1994), BioTechniques 17: 242-6), which, briefly, involves the synthesis of overlapping oligonucleotides containing portions of the sequence
10 encoding the polypeptide, annealing and ligating of those oligonucleotides, and then amplification of the ligated oligonucleotides by PCR.

Alternatively, a polynucleotide encoding a polypeptide described herein can be generated from nucleic acid from a suitable source (*e.g.*, a hybridoma) using methods well known in the art (*e.g.*, PCR and other molecular cloning methods). For example, PCR
15 amplification using synthetic primers hybridizable to the 3' and 5' ends of a known sequence can be performed using genomic DNA obtained from hybridoma cells producing the polypeptide of interest. Such PCR amplification methods can be used to obtain nucleic acids comprising the sequence encoding *e.g.*, IL2, a linker sequence, or IL2-R α . The amplified nucleic acids can be cloned into vectors for expression in host cells and for further
20 cloning, for example, to generate polypeptides.

If a clone containing a nucleic acid encoding a particular polypeptide is not available, but the sequence of the polypeptide molecule is known, a nucleic acid encoding the polypeptide can be chemically synthesized or obtained from a suitable source (*e.g.*, a cDNA library or a cDNA library generated from, or nucleic acid, preferably poly A⁺ RNA, isolated from, any tissue or cells expressing the proteins of interest, such as hybridoma cells
25 selected to express a polypeptide described herein) by PCR amplification using synthetic primers hybridizable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe specific for the particular gene sequence to identify, *e.g.*, a cDNA clone from a cDNA library that encodes the polypeptides. Amplified nucleic acids
30 generated by PCR can then be cloned into replicable cloning vectors using any method well known in the art.

DNA encoding polypeptides described herein can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the polypeptides disclosed herein). Hybridoma cells can serve as a source of such DNA. Once isolated, the DNA can be placed into expression
5 vectors, which are then transfected into host cells such as *E. coli* cells, simian COS cells, Chinese hamster ovary (CHO) cells (*e.g.*, CHO cells from the CHO GS SYSTEM™ (Lonza)), or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of polypeptides in the recombinant host cells.

Therapeutic Uses and Methods

10 The miniaturized dystrophin polypeptides, polynucleotides encoding miniaturized dystrophin polypeptides, vectors (*e.g.*, rAAV) harboring polynucleotides encoding miniaturized dystrophin polypeptides and methods described herein have numerous *in vitro* and *in vivo* utilities. For example, the nucleotide sequence encoding a miniaturized dystrophin polypeptide, *e.g.*, a vector, *e.g.*, an AAV vector, or the polypeptides described
15 herein can be administered to cells in culture, *in vitro* or *ex vivo*, or to human subjects, *e.g.*, *in vivo*, to treat diseases.

Accordingly, disclosed herein are therapeutic methods using any of the miniaturized dystrophin nucleic acid molecules as disclosed herein, polypeptides as disclosed herein, host cells as disclosed herein, vectors as disclosed herein, or
20 pharmaceutical compositions as disclosed herein, or any combination thereof.

In some embodiments, disclosed herein is a method of expressing a miniaturized dystrophin polypeptide in a subject in need thereof, comprising administering to the subject a nucleic acid as disclosed herein, a vector as disclosed herein, a host cell as disclosed herein, or a pharmaceutical composition as disclosed herein.

25 In some embodiments, disclosed herein is a method of treating a subject having a disease or condition comprising administering to the subject a nucleic acid as disclosed herein, a vector as disclosed herein, a polypeptide as disclosed herein, a host cell as disclosed herein, or a pharmaceutical composition as disclosed herein. In some
30 embodiments, the disease or condition is caused by dystrophin deficiency. In some embodiments, the disease is Duchene muscular dystrophy (DMD), Becker muscular

dystrophy (BMD), X-linked dilated cardiomyopathy (XLDC), facioscapulohumeral muscular dystrophy, myotonic muscular dystrophy, limb-girdle muscular dystrophy, oculopharyngeal muscular dystrophy, Emery-Dreifuss muscular dystrophy, distal muscular dystrophy, and/or congenital muscular dystrophy. In other embodiments, the disease to be
5 treated is Sarcopenia, heart disease, cachexia.

In some embodiments, a nucleic acid molecule as disclosed herein, a polypeptide as disclosed herein, a vector (e.g., rAAV) as disclosed herein, a host cell as disclosed herein, or a pharmaceutical composition as disclosed herein is administered intravenously, transdermally, intradermally, subcutaneously, orally, or pulmonarily, or any combination
10 thereof. In some embodiments, the nucleic acid molecule as disclosed herein, the polypeptide as disclosed herein, the vector as disclosed herein, the host cell as disclosed herein, or the pharmaceutical composition as disclosed herein is administered via a topical, epidermal mucosal, intranasal, oral, vaginal, rectal, sublingual, topical, intravenous, intraperitoneal, intramuscular, intraarterial, intrathecal, intralymphatic, intralesional,
15 intracapsular, intraorbital, intracardiac, intradermal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural or intrasternal route. In some embodiments, the nucleic acid molecule, the vector (e.g., rAAV), the host cell as disclosed herein, or the polypeptide is administered intravenously.

In some embodiments, the method of treatment further comprises administering to
20 the subject a second agent.

As used herein, the term "subject" includes any human or non-human animal. For example, the methods and compositions described herein can be used to treat a subject having cancer. The term "non-human animal" includes all vertebrates, e.g., mammals and non-mammals, such as non-human primates, sheep, dog, cow, chickens, amphibians,
25 reptiles, etc. In some embodiments, the subject is a human.

In some embodiments, the administration of the nucleic acid molecule, the vector (e.g., rAAV), the polypeptide, the host cell, or the pharmaceutical composition to the subject results in an increased dystrophin protein expression, relative to dystrophin protein expression in the subject prior to the administration, wherein the dystrophin protein
30 expression is increased by at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 6-fold, at least about 7-fold, at least about 8-fold, at least

about 9-fold, at least about 10-fold, at least about 11-fold, at least about 12-fold, at least about 13-fold, at least about 14-fold, at least about 15-fold, at least about 20-fold, at least about 25-fold, at least about 30-fold, at least about 35-fold, at least about 40-fold, at least about 50-fold, at least about 60-fold, at least about 70-fold, at least about 80-fold, at least about 90-fold, or at least about 100-fold.

In certain aspects of the disclosure, the method comprises, or further comprises, administering an AAV therapy to the subject. In some embodiments, the AAV therapy comprises administering a recombinant AAV. Any recombinant AAV known in the art and/or disclosed herein can be used in the methods of the present disclosure. In some embodiments, the AAV therapy comprises administering an AAV selected from the group consisting of AAV type 1, AAV type 2, AAV type 3 (including types 3A and 3B), AAV type 4, AAV type 5, AAV type 6, AAV type 7, AAV type 8, AAV type 9, AAV type 10, AAV type 11, AAV type 12, AAV type 13, snake AAV, avian AAV, bovine AAV, canine AAV, equine AAV, ovine AAV, goat AAV, shrimp AAV, and any combination thereof.

In certain embodiments, the AAV therapy comprises administering an AAV type 1. In certain embodiments, the AAV therapy comprises administering an AAV type 2. In certain embodiments, the AAV therapy comprises administering an AAV type 3. In certain embodiments, the AAV therapy comprises administering an AAV type 4. In certain embodiments, the AAV therapy comprises administering an AAV type 5. In certain embodiments, the AAV therapy comprises administering an AAV type 6. In certain embodiments, the AAV therapy comprises administering an AAV type 7. In certain embodiments, the AAV therapy comprises administering an AAV type 8. In certain embodiments, the AAV therapy comprises administering an AAV type 9. In certain embodiments, the AAV therapy comprises administering an AAV type 10. In certain embodiments, the AAV therapy comprises administering an AAV type 11. In certain embodiments, the AAV therapy comprises administering an AAV type 12. In certain embodiments, the AAV therapy comprises administering an AAV type 13.

In some embodiments, treatment of a subject with the miniaturized dystrophin nucleic acid molecules as disclosed herein, polypeptides as disclosed herein, host cells as disclosed herein, vectors as disclosed herein, or pharmaceutical compositions as disclosed herein, or any combination thereof, does not cause significant inflammatory reactions, *e.g.*,

immune-mediated pneumonitis, immune-mediated colitis, immune mediated hepatitis, immune-mediated nephritis or renal dysfunction, immune-mediated hypophysitis, immune-mediated hypothyroidism and hyperthyroidism, or other immune-mediated adverse reactions. In some embodiments, treatment of a subject with the miniaturized dystrophin nucleic acid molecules as disclosed herein, polypeptides as disclosed herein, host cells as disclosed herein, vectors as disclosed herein, pharmaceutical compositions as disclosed herein, or any combination thereof does not cause significant cardiac disorders, *e.g.*, ventricular arrhythmia; eye disorders, *e.g.*, iridocyclitis; infusion-related reactions; increased amylase, increased lipase; nervous system disorders, *e.g.*, dizziness, peripheral and sensory neuropathy; skin and subcutaneous tissue disorders, *e.g.*, rash, pruritus, exfoliative dermatitis, erythema multiforme, vitiligo or psoriasis; respiratory, thoracic and mediastinal disorders, *e.g.*, cough; fatigue; nausea; decreased appetite; constipation; arthralgia; or diarrhea.

Kits

Also disclosed herein are kits comprising one or more nucleic acid molecules disclosed herein, one or more vectors (*e.g.*, rAAV) as disclosed herein, one or more polypeptides as disclosed herein, or one or more host cells as disclosed herein, or any combination thereof. In some embodiments, the kit also comprises instructions for administering any of the aforesaid, or a combination thereof, to a subject in need thereof.

The terms "kit" and "system," as used herein are intended to refer to at least one or more nucleic acid molecules disclosed herein, one or more vectors (*e.g.*, rAAV) as disclosed herein, one or more polypeptides as disclosed herein, or one or more host cells as disclosed herein, or any combination thereof, which, in specific embodiments, are in combination with one or more other types of elements or components (*e.g.*, other types of biochemical reagents, containers, packages, such as packaging intended for commercial sale, instructions of use, and the like).

In some embodiments, disclosed is a kit comprising (a) one or more of a miniaturized dystrophin polypeptide as described herein, a composition comprising a miniaturized dystrophin polypeptide as described herein, a nucleic acid encoding for a miniaturized dystrophin polypeptide as described herein, a vector (*e.g.*, rAAV), and/or a

host cell; and (b) and instructions for administering any of the aforesaid to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a miniaturized dystrophin polypeptide as described herein and (b) and instructions for administering the miniaturized dystrophin polypeptide to a subject in need thereof. In some embodiments, disclosed is a
5 kit comprising (a) a composition comprising a miniaturized dystrophin polypeptide as described herein and (b) and instructions for administering the composition to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a nucleic acid encoding for a miniaturized dystrophin polypeptide as described herein and (b) and instructions for administering the nucleic to a subject in need thereof. In some
10 embodiments, disclosed is a kit comprising (a) a vector as described herein and (b) and instructions for administering the vector to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) an AAV vector as described herein and (b) and instructions for administering the vector to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a host cell as described herein and (b) and instructions for
15 administering the host cell to a subject in need thereof.

In a specific embodiment, provided herein is a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions described herein, such as one or more miniaturized dystrophin peptides provided herein. In some embodiments, the kits contain a
20 pharmaceutical composition described herein and any prophylactic or therapeutic agent, such as those described herein. In certain embodiments, the kits can contain a T cell mitogen, such as, *e.g.*, phytohaemagglutinin (PHA) and/or phorbol myristate acetate (PMA), or a TCR complex stimulating antibody, such as an anti-CD3 antibody and anti-CD28 antibody. Optionally associated with such container(s) can be a notice in the form
25 prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

Also provided herein are kits that can be used in the above methods. In one embodiment, a kit comprises a miniaturized dystrophin polypeptide described herein,
30 preferably a purified miniaturized dystrophin polypeptide, in one or more containers. In a specific embodiment, kits described herein contain a substantially isolated miniaturized

dystrophin polypeptide as a control. In another specific embodiment, the kits described herein further comprise a control protein which does not react with a miniaturized dystrophin polypeptide antigen. In another specific embodiment, kits described herein contain one or more elements for detecting the binding of the miniaturized dystrophin polypeptide to a dystrophin antigen (*e.g.*, the miniaturized dystrophin polypeptide can be conjugated to a detectable substrate such as a fluorescent compound, an enzymatic substrate, a radioactive compound or a luminescent compound, or a second antibody which recognizes the first antibody can be conjugated to a detectable substrate). In specific embodiments, a kit provided herein can include a recombinantly produced or chemically synthesized miniaturized dystrophin polypeptide. The antigen to a miniaturized dystrophin polypeptide disclosed herein as provided in the kit can also be attached to a solid support. In a more specific embodiment, the detecting means of the above described kit includes a solid support to which an antigen of the miniaturized dystrophin polypeptide is attached. Such a kit can also include a non-attached reporter-labeled anti-human antibody or anti-mouse/rat antibody. In this embodiment, binding of the miniaturized dystrophin polypeptide to an antigen can be detected by binding of the said reporter-labeled antibody.

The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook *et al.*, ed. (1989) *Molecular Cloning A Laboratory Manual* (2nd ed.; Cold Spring Harbor Laboratory Press); Sambrook *et al.*, ed. (1992) *Molecular Cloning: A Laboratory Manual*, (Cold Springs Harbor Laboratory, NY); D. N. Glover ed., (1985) *DNA Cloning*, Volumes I and II; Gait, ed. (1984) *Oligonucleotide Synthesis*; Mullis *et al.* U.S. Pat. No. 4,683,195; Hames and Higgins, eds. (1984) *Nucleic Acid Hybridization*; Hames and Higgins, eds. (1984) *Transcription And Translation*; Freshney (1987) *Culture Of Animal Cells* (Alan R. Liss, Inc.); Immobilized Cells And Enzymes (IRL Press) (1986); Perbal (1984) *A Practical Guide To Molecular Cloning*; the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); Miller and Calos eds. (1987) *Gene Transfer Vectors For Mammalian Cells*, (Cold Spring Harbor Laboratory); Wu *et al.*, eds., *Methods In Enzymology*, Vols. 154 and 155; Mayer and Walker, eds. (1987) *Immunochemical Methods In Cell And Molecular Biology*

(Academic Press, London); Weir and Blackwell, eds., (1986) Handbook Of Experimental Immunology, Volumes I-IV; Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1986);); Crooks, Antisense drug Technology: Principles, strategies and applications, 2nd Ed. CRC Press (2007) and in
5 Ausubel *et al.* (1989) Current Protocols in Molecular Biology (John Wiley and Sons, Baltimore, Md.).

All of the references cited above, as well as all references cited herein and the amino acid or nucleotide sequences (*e.g.*, GenBank numbers and/or Uniprot numbers), are incorporated herein by reference in their entireties.

10 The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1: Design of novel miniaturized dystrophins

15 Mutations in the dystrophin gene often result in an impairment of the stability of the corresponding dystrophin protein, which in turn leads to proteosomal degradation of the unstable dystrophin protein, and dystrophic pathophysiology. Similarly, miniaturizing dystrophin-encoding DNA to accommodate the limited packaging capacity of AAV can impair the stability of the corresponding miniaturized dystrophin protein. Novel
20 miniaturized dystrophins with novel junctions as depicted in FIG. 2 were designed for further testing.

Example 2: Assessment of stability of miniaturized dystrophin proteins expressed in tissue culture cells

25 The stability of the various miniaturized dystrophin proteins depicted in FIG. 2 was examined by comparing the presence of miniaturized dystrophin protein in cells transfected with the corresponding miniaturized dystrophin expression vectors. Male human isogenic induced-pluripotent stem cell (iPSC)-derived induced cardiomyocytes (iCMs) that carry an E2035X premature stop codon in the dystrophin gene that prevented endogenous
30 dystrophin expression were used for these protein stability studies (Fujifilm Cellular Dynamics, Inc., Madison, WI). These cells were transfected with various plasmids

expressing miniaturized dystrophin proteins and the presence of miniaturized dystrophin protein was examined after the transfected cells had been cultured *in vitro* for 24 days by a Meso Scale Discovery (MSD) ELISA assay (Meso Scale Diagnostics, Rockville, MD). The miniaturized dystrophins tested and the test results of the aforesaid assay are shown in FIG. 2 and FIG. 3, respectively. The data indicated that miniaturized dystrophin peptide BXA-212372-J4V13 (SEQ ID NO:83) provides the best protein expression among the miniaturized dystrophin expression vectors and peptides tested.

10 Example 3: Assessment of immunogenicity of novel junctions in miniaturized dystrophin proteins

The immunogenicity of each of the peptides listed in Table 5 (SEQ ID NOs: 68 to 72), representing the novel J4 junctions created within the miniaturized dystrophin designs tested for protein stability (see FIG. 3 and Example 2), were tested using an *in silico* immunogenicity prediction tool. The novel junctions of the BXA-212372-J4V11 and particularly BXA-212372-J4V12 and BXA-212372-J4V13 designs (SEQ ID NO: 70, SEQ ID NO: 69, and SEQ ID NO: 68, respectively) were determined to have minimal immunogenic risk, based on the aforesaid *in silico* approach (see FIGS. 4A and 4B).

The immunogenic potential of the aforesaid junction peptides (see Table 5) were then tested using an *in vitro* T cell proliferation assay as described below. Briefly, samples of peripheral blood mononuclear cells (PBMC) were isolated from healthy volunteer human subjects by Ficoll (GE Healthcare Chicago, IL) gradient centrifugation and characterized, regarding human lymphocyte antigen (HLA) Class I and II expression, using a combination of polymerase chain reaction (PCR) amplification and hybridization with oligonucleotide probes (ProImmune, Sarasota, FL).

25 A panel of PBMC samples from 40 donors, having an HLA expression profile closely matching world population frequencies, was used for further analysis. PBMC samples were labeled with CFSE (Invitrogen, Carlsbad, CA) to monitor proliferation and plated in 96 well plates in six replicates at 200,000 cells per well in RPMI (Lonza, Basel, Switzerland) containing 10% human AB serum (Bioreclamation, Westbury, NY), non-essential amino acids and pen-strep (both Gibco/Fisher Scientific). The BXA-212372 junction peptides listed in Table 5 and control peptides were each cultured with the panel

of 40 PBMC samples at 1 μ M for 7 days, after which the media was washed away and cells were labeled with an anti-CD4 and an anti CD8 APC monoclonal antibody (BD Biosciences, Franklin Lake, NJ). After removal of the unbound antibodies through washing, cells were fixed with 3.7% formalin (Sigma, St. Louis, MO) in PBS and analyzed by flow cytometry to determine the percentage of proliferating CD4⁺ cells or CD8⁺ cells. The percentage of samples (among the 40 donor samples) that showed a positive response after seven days in culture with the different BXA-212372 junction peptides - defined as a significant increase in the number of CD4⁺ or CD8⁺ T proliferating cells compared to PBMC incubated in media without junction peptides or control peptides - is shown in FIG. 5A (CD4⁺) and FIG. 5B (CD8⁺). Control peptides used were: (1) Avastin Framework Peptide; (2) IL-21R Peptide; and (3) CEFX Peptide Pool. It was found that the junction peptide of design BXA-212372-J4V13 (see FIG. 2 and Table 5) was among the best in terms of immunogenic risk (see FIG. 4 and FIG. 5).

Example 4: Codon optimization

The nucleotide sequence encoding miniaturized dystrophin design BXA-212372-J4V13 (SEQ ID NO: 101) was then codon optimized to optimize protein expression, resulting in construct BXA-220931 (SEQ ID NO: 100) (see Table 9 and FIG. 6).

Example 5: Promoter/intron screening for expression of miniaturized dystrophin

A series of promoters and introns were evaluated for their suitability for driving the expression of miniaturized dystrophin. To that end, promoters and introns were cloned in a GFP reporter plasmid the expression of which was then evaluated by transfection into human iCMs (see Example 2). The results indicated that the C5-12T promoter (see FIG. 7A) (US 2004/0175727) and SV40 intron (see FIG. 7B) were superior to other tested designs in driving the expression of GFP protein. Both elements were therefore included in the miniaturized dystrophin expression constructs used, as described below.

Example 6: *In vivo* expression of miniaturized dystrophin constructs and nNOS restoration

Mdx mice (dystrophin deficient *mdx^{scsn}*) were treated systemically with 2e14vg/kg AAV9 virus containing miniaturized dystrophin expression constructs (BXA-212372-J4V4, BXA-212372-J4V11, BXA-212372-J4V12, and BXA-212372-J4V13) via retroorbital injection at 2 weeks of age and examined at either 4 weeks of age or 12 weeks of age. Miniaturized dystrophin expression was driven by the C5-12(T) promoter. The heart and skeletal muscles of treated animals were dissected, frozen in OCT 2-methylbutane in liquid N₂. 10µm frozen sections were immunostained for human dystrophin with monoclonal antibody Manex1011B directly conjugated to Alex488 (DSHB, Iowa City, IA) (see FIG. 8) or with a polyclonal antibody against nNOS (ThermoFisher, Waltham, MA) detected with a secondary fluorescent antibody (see FIG. 9). WGA-conjugated with Alexa-694 was used as control to label muscle cells. Sections were coverslipped in medium containing DAPI to label nuclei. Slides were imaged using a Leica SP8 confocal microscope (Leica Microsystems; see FIGS. 8 and 9). It was found that all AAV9 constructs tested expressed well in the examined muscle tissue, *i.e.*, the heart, diaphragm (Dia), tibialis anterior (TA), and the gastrocnemius muscle (Gast) (FIG. 8A, BXA-212372-J4V4; FIG. 8B, BXA-212372-J4V11; FIG. 8C, BXA-212372-J4V12; and FIG. 8D, BXA-212372-J4V13 (BXA-220931)). No dystrophin protein aggregates were detectable (FIGS. 8A-D). FIG. 9 shows nNOS restoration at the tibialis anterior (TA) muscle sarcolemma of *mdx* mice treated with the indicated AAV9 constructs. Untreated wild-type mice served as positive controls, and *mdx* mice treated with an unrelated miniaturized dystrophin construct served as the negative control. Note that all J4 variants, except J4V12, restored nNOS to the sarcolemma.

Example 7: Effect of miniaturized dystrophin on *in vitro* physiology of human iPSC-derived iCMs

Human iPSC-derived iCMs have been reported to have electrophysiological properties close to primary adult cardiomyocytes and to respond similarly to a range of cardiac ion channel inhibitors as well as adrenergic and muscarinic receptor agonists and antagonists. By comparison to isogenic wild-type iCMs, DMD iCMs carrying the E2035X mutation (see above) have a lower Na⁺ channel amplitude, a prolonged cFPD (Q-T

interval), and a greater beat rate variability, as determined using multi-electrode arrays. It was examined whether the expression of miniaturized dystrophin (e.g., BXA-220931) in DMD (E2035X) iCMs can mitigate the cells' DMD phenotype and improve their physiological traits. Multi-electrode arrays, impedance contraction assay, and Ca^{2+} transients can be used to measure the effect of miniaturized dystrophin expression. Human iPSC-derived DMD (E2035X) iCMs were purchased from Fujifilm Cellular Dynamics, Inc. (Madison, WI). Our own work has shown that co-culturing hiPSC-derived iCMs with fibroblasts provides a more stable preparation for electrophysiological studies on multi-electrode arrays (MEAs). Human ventricular fibroblasts were purchased from Lonza (Walkersville, MD).

Microelectrode array (MEA) technology enables high content spatiotemporal analysis of excitable cells or tissues from an array of embedded substrate-integrated extracellular electrodes onto which cells can be cultured or tissues placed. Extracellular field potentials (FPs) are recorded by each electrode and correspond to cellular action potentials. Assessment of FP morphology, duration and conduction velocity provides a picture of ion channel activities of a treatment as well as effects on repolarization and conduction.

Human iPSC-derived DMD (E2035X) iCMs were cultured with 7% CO_2 on 0.1 % gelatin treated 6-well culture plates for 7 days, then trypsinized and diluted with human adult cardiac fibroblasts at an approximately 5:1 ratio (iCMs vs. fibroblasts). DMD (E2035X) iCMs and fibroblasts were then co-cultured on laminin-coated 9-well multi-electrode array (MEA) plates (256-9wellMEA300/30iR-ITO-mq; Multichannel Systems GmbH, Reutlingen, Germany). After 5 days of culture, the cells formed a spontaneously beating monolayer over recording electrodes imbedded in each well. Spontaneous extracellular field potentials (FPs) were recorded from 28 electrodes/well (30 μm diameter, 300 μm center to center spacing) at a sampling frequency of 10 kHz using an USB-MEA256-System and MC Rack acquisition software (Multichannel Systems GmbH, Reutlingen, Germany). Following a 20-minute equilibration period in a humidified environment at 37°C with constant 5% CO_2 and 95% O_2 supply, wells were either infected with AAV8-BXA-220931 (AAV8 virus including as cargo a transgene including the coding sequence for BXA-220931 and the C5-12(T) Promoter, SV40 Intron, 3' UTR and

polyA as set forth in Table 10 herein) at MOI of 1×10^6 in 300 μ l maintenance medium for 48 hrs, or were left untreated as negative control.

The DMD (E2035X) iCMs were then evaluated for effects of the expression of the miniaturized dystrophin BXA-220931 on electrophysiological parameters 5 days, 7 days, and 9 days after infection. Electrophysiological parameters measured were field potential (FP) duration, a surrogate for repolarization, field potential conduction velocity, and inter-pulse intervals (IPIs). Field potential duration was corrected for beat rate changes (FPDc). Conduction velocity was quantified by measuring field potential activation times for each electrode imbedded in an MEA well during a synchronized single propagated beat. The digitized recordings of field potentials from each electrode were smoothed using a 21-point least squares smoothing polynomial (Savitsky & Golay, Analytical Chemistry, 1964) with a window of 2.1ms. The activation time was the value for the peak in the negative derivative of each field potential waveform. The time between two of the earliest and latest activation times was the conduction time for field potential propagation across a monolayer of DMD (E2035X) iCMs and the distance between these two electrodes was the conduction distance. The conduction time divided by the conduction distance of each propagation was the conduction velocity of each beat of the monolayer DMD (E2035X) iCMs in an MEA well. Data were analyzed with custom software written in MatLab (Mathworks, Natick, MA). Beat rate (beats/minute), a surrogate for heart rate, was calculated by using $BR = 60000/IPI$, where the IPI is the averaged IPIs (msec) of 100 second recording at steady-state under each condition. All treatments had at least 7 replicates and the study was repeated twice.

The data showed that miniaturized dystrophin BXA-220931 significantly improved conduction velocity by ~49% compared to untreated DMD (E2035X) iCMs (two-way ANOVA *** $P < 0.001$ with Sidak's post-test $n = 6$) (see FIG. 10B). Expression of miniaturized dystrophin in the DMD (E2035X) iCMs was confirmed by ELISA (see FIG. 10C).

Example 8: *In vivo* studies – analysis of expression, biodistribution and the ability to prevent the dystrophic phenotype in *mdx^{scsn}* mice of miniaturized dystrophins

Two miniaturized dystrophin viral constructs were used in these studies. One construct included the coding sequence for BXA-220931 and the C5-12(T) Promoter, SV40

Intron, 3' UTR and polyA as set forth in Table 10 herein. The other included the same non-coding elements but expressed miniaturized dystrophin BXA-212374, which has been described in Banks *et. al.*, (PLOS Genetics, volume 6(5), 2010) and has the following domain structure: ABD1/H1/R1/R2/R3/H3/R24/H4/CR). Dystrophin deficient *mdx^{scsn}* mice were treated by retro-orbital injection with about 2e14vg/kg AAV9-BXA-220931 virus or AAV9-BXA-212374 virus at 2 weeks of age. Treated and untreated mice were terminated two-weeks after virus administration (n = 3) to examine expression levels and biodistribution of human miniaturized dystrophin (see FIGS. 11 and 12). Additional mice (n = 10-12) were terminated at 12 weeks of age and examined for expression levels and biodistribution of human miniaturized dystrophin and prevention of dystrophy (see FIGS. 13 to 20 and this and subsequent Examples 9-13). Untreated wild type mice and endogenous mouse dystrophin expression served as controls.

Muscle tissue of treated and untreated *mdx^{scsn}* mice was analyzed for the amount of virus genomes present as well as dystrophin mRNA and protein expression, as described in more detail below. The data showed that sufficient virus was administered to dystrophin deficient *mdx^{scsn}* mice to achieve expression levels (mRNA and protein) of miniaturized dystrophin in striated muscle in these animals at 4 weeks of age and at 12 weeks of age that were higher than corresponding expression levels of endogenous dystrophin in wild-type animals (see FIG. 11 and FIG. 14, respectively).

The skeletal muscles in dystrophin deficient *mdx^{scsn}* mice typically undergo necrosis and regeneration from ~3-4 weeks of age. The regenerated muscle fibers are typically more variable in size and contain centrally located nuclei in frozen transverse sections. Also, fibrosis becomes more prevalent in regenerated muscles. The muscle fiber size, proportion of centrally located nuclei, and fibrosis in untreated *mdx* muscles and *mdx* muscles treated with miniaturized dystrophin BXA-220931 or BXA-212374 (partly) were measured by histology and immune-fluorescence analysis of tissue sections, as described in more detail below. The proportion of muscle fibers expressing the miniaturized dystrophins was also quantified in a similar fashion, as described in more detail below. The data showed that miniaturized dystrophins BXA-220931 and BXA-212374 were expressed in nearly all analyzed myofibers/myocytes of virus-treated *mdx^{scsn}* mice, including the heart, and prevented the central nucleation to a degree similar to wild-type muscles at 4 weeks of age

and 12 weeks of age (see FIG. 12 and FIGS. 15-16, respectively). Importantly, the expression and biodistribution of miniaturized dystrophin was maintained more than two months post AAV treatment.

Miniaturized dystrophin BXA-220931 also prevented the dystrophic pathology seen in untreated *mdx^{scsn}* mice, as shown by histological and immuno-fluorescence analysis of muscle tissue sections (FIG. 13 and FIG. 15).

Vector genome quantitation /genomic DNA isolation and qPCR - For genomic DNA isolation, striated muscle tissue was homogenized using Qiagen TissueLyser (Qiagen, Venlo, Netherlands) and genomic DNA was isolated from homogenized tissue using a Qiagen DNeasy 96 Blood & Tissue Kit (Qiagen, Venlo, Netherlands, #69581). Tissue (~10mg) was placed in 96 well plates (Costar® 96-Well Assay Block 1 ml, #3958) containing 200µl of proteinase K-buffer ATL and one 5mm steel bead, homogenized using the Qiagen TissueLyser at 30hz for 2 min, which was repeated until the tissue was homogenized. Genomic DNA isolation was performed in accordance with the manufacturer's instructions. For genomic qPCR, each DNA sample was run in duplicates with primer/probe sets (wild-type dystrophin F-5' AAGGCCTGACAGGGCAAAA3', R-5' CAGGGCATGAACTCTTGTGGAT3', probe 6FAM-CTGCCAAAAGAAAAA-MGBNFQ; BXA-220931 F-5' CGCGAGGACGTGCAGAA3', R-5' TTGCTGAACTGGGCGTTGA3', Probe 6FAM-AAACCTTCACCAAATGG-MGBNFQ; BXA-212374 F-5' TGGAAGATTGCTACGAGCGC3', R-5' CAGGTCGCTGAACAGGTTCT3', Probe 6FAM-GCAAGTTCGGCAAGCAGCACA-MGBNFQ) in 384 well clear reaction plates (Applied Biosystems, Waltham, MA, #4483285). To each qPCR reaction, 2µl of genomic DNA (80ng) and 8µl of master mix (5µl of Applied Biosystems™ TaqMan™ Fast Advanced Master Mix (ThermoFisher), 0.5µl 20x FAM primer probe mix and 2.5µl water) was added and plates were centrifuged for 1 min at 1000rpm. Samples were incubated at 95°C for 2min followed by 40 cycles at 95°C for 15sec and 60°C for 1 min using the ViiA™ 7 Real-Time PCR System and QuantStudio software for data analysis and vector genome

quantitation (Applied Biosystems, Waltham, MA). Total genomic DNA was quantitated by absorption spectroscopy.

mRNA isolation - For isolation of total RNA, tissue is was homogenized using Qiagen TissueLyzer (Qiagen, Venlo, Netherlands) and RNA was isolated from homogenized tissue using a Qiagen RNeasy 96 Universal Tissue Kit (Qiagen, Venlo, Netherlands, #74881). Tissue (~15mg) was placed in RNeasy kit collection microtubes containing 750µl of QIAzol Lysis Reagent (Qiagen, Venlo, Netherlands) and one 5mm steel bead, homogenized using TissueLyzer at 30hz for 2min, which was repeated until the tissue was homogenized. This step was followed by a centrifugation at 6000xg for 1min at 4°C. To each tube 150 ml of chloroform were added and samples were vortexed vigorously for 15sec. Following a 3 min incubation step at room temperature, samples were spun at 6000xg for 15min at 4°C. The aqueous phase was removed (~360µl) and transferred to a new tube containing 1 volume of RNase free 70% EtOH. All samples were transferred to a 96 well RNeasy 96 plates, which were then sealed with AirPore tape (Qiagen, Venlo, Netherlands) and centrifuge at 5600xg for 4 min at room temperature. 400µl of RW1 buffer was added per well and plates resealed and spun for 4 min at 5600xg. During this step, a DNaseI stock solution was prepared by adding 550µl of RNase free water per DNase vial (Qiagen, Venlo, Netherlands). 670µl of the DNase I stock solution was diluted into 7.3mls RDD buffer, mixed and stored at 4°C. When centrifugation was completed, the flow-through was discarded and 80µl of DNase I mix was added directly to the center of each well and the plate was incubated at room temperature for 15min. Following incubation, 400µl of RW1 was added to each well and the plate was sealed and centrifuged for 4min at 5600xg. Flow-through was discarded and 800µl of RPE buffer were added per well and the plate was re-sealed and spun for 4 min at 5600xg. This process was repeated and the plate was centrifuged for 10min at 5600xg. Each sample was then eluted into a fresh tube by adding 60µl of RNase free water to the center of each well and centrifuging the tubes for 4min at 5600xg. To improve recovery, the eluted 60µl were re-applied back onto the plate and centrifuged for an additional 4min at 5600xg. RNA yield was quantitated using a NanoDrop™ 8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

ddPCR Quantification of mRNA - For cDNA synthesis and subsequent quantitative PCR, 1µg of RNA was added to one well of a 96 well plate in 10µl H₂O (Axygen™ 96-

Well PCR Microplates, EMSCO Scientific Enterprises, Inc., Philadelphia, PA). To each well 10µl of master mix (High Capacity cDNA Reverse Transcription Kit, Applied Biosystems, Waltham, MA) was added and the plate was centrifuged at 1000rpm. cDNA synthesis was carried out at 25°C for 10min, 37°C for 120 min, and 85°C for 5min, which was then followed by a hold at 4°C. For ddPCR, each sample was then run in duplicate with the following primer/probe sets: wild-type dystrophin F-5' AAGGCCTGACAGGGCAAAA3', R-5'CAGGGCATGAACTCTTGTGGAT3', probe 6FAM-CTGCCAAAAGAAAAA-MGBNFQ; BXA-220931 F-5'CGCGAGGACGTGCAGAA3', R-5' TTGCTGAACTGGGCGTTGA3', Probe 6FAM-AAACCTTCACCAAATGG-MGBNFQ; BXA-212374 F-5'TGGAAGATTGCTACGAGCGC3', R-5'CAGGTCGCTGAACAGGTTCT3', Probe 6FAM-GCAAGTTCGGCAAGCAGCACA-MGBNFQ. To each reaction, 7.5µl of cDNA and 17.5µl of master mix (12.5µl ddPCR Supermix (BIO-RAD Laboratories, Hercules, CA), 0.5µl 20x FAM primer probe mix and 4.5µl water) were added to Eppendorf twin.tec® semi-skirted 96 well plates (Eppendorf, Germany, #951022055), which were then sealed and centrifuged for 1min at 1000rpm and processed for droplet generation in DG32 Automated Droplet Generator Cartridges (Bio-Rad Laboratories, Hercules, CA, #1864108). Samples were PCR-amplified in the Bio-Rad C1000 Touch Thermocycler (95°C for 10 min followed by 40 cycles at 94°C for 30sec/ 60°C 1min; 98°C 10min) and immediately analyzed for fluorescence reading in a BioRad Droplet Reader and mRNA levels in target striated muscle tissue were determined. Dystrophin mRNA was quantitated in copy numbers relative to total RNA (µg, quantitated by absorption spectroscopy).

Protein expression determination by MSD-ELISA - Miniaturized dystrophin protein expression in target striated muscle tissue was determined by ELISA assay (Meso Scale Delivery-Enzyme Linked Immunosorbent Assay, Model 1201 MESOTM Sector S 600, Meso Scale Diagnostics, Rockville, MD). Multi-assay 384-well plates (Meso Scale Diagnostics, Rockville, MD) were pre-coated with monoclonal mouse anti-human dystrophin antibody Manex 1011b (DSBH, University of Iowa, Developmental Studies Hybridoma Bank) at a concentration of 2µg/ml in bicarbonate buffer (pH 9.4) overnight. Plates were then washed 3x with PBS and then blocked with blocking buffer (5% BSA in PBS) for 4hrs with shaking at room temperature. Tissues (~20mg) were homogenized in

RIPA buffer at a concentration of 1mg tissue/10µl lysis buffer (Millipore Sigma, Germany, #R0278) with protease inhibitor cocktail tablets (Roche, #04693159 001) using Qiagen TissueLyzer at 30hz for 5 min, which was repeated until the tissue was homogenized. The tissue/RIPA lysates were diluted 1:3 in binding buffer (1% BSA, 0.05% Tween-20, 20 mM Tris pH 7.5 in PBS). Tissue lysates and sulfo-conjugated mouse anti-human dystrophin antibody Mandys 106 (DSBH, 0.2 µg/ml) were added to the pre-coated 384 well plates and incubated at 4°C with shaking overnight. Plates were washed with PBS with 0.05% Tween-20 and additional 40µl MSD Read Buffer T with surfactant (Meso Scale Diagnostics, Rockville, MD, #R92TC-1). Plates were then read on an MSD Sector[®] 6000 Imager (Meso Scale Diagnostics, Rockville, MD). Endogenous dystrophin was assayed using the same protocol but employing antibodies binding murine dystrophin.

Protein expression determination by liquid chromatography–mass spectrometry (LC–MS)- Striated (heart, skeletal) muscle tissues were collected and immediately frozen. Prior to analyses, the tissues were homogenized with RIPA buffer in a 1:20 ratio. The homogenates were digested with trypsin and after fractionation for peptide enrichment, the samples were analyzed by LC-MS/MS by monitoring a previously identified unique peptides common to both mouse and human dystrophin (LLDLLEGLTGQK). Stable isotope labeled analogs (SIL) for human and mouse peptides were spiked into the homogenate and were used to estimate the measured levels. Total protein was also obtained and used for normalization purposes.

Immuno-fluorescence slide preparation - *mdx^{scsn}* mice were treated with AAV9-BXA-220931 or AAV9-BXA-212374 at 2 weeks of age. Heart and skeletal muscle tissue was collected from these mice at 4 weeks of age or at 12 weeks of age. Muscle tissue was frozen in OCT in liquid N₂ and sectioned at 5µm. Sections were mounted on standard microscope slides and stored at -80°C. Frozen sections were brought to room temperature and blocked with 200µl of blocking buffer (Dulbeccos Phosphate Buffered Saline (DPBS) (ThermoFisher, Waltham, MA, #14190144) supplemented with 0.05% TritonX-100 (Sigma-Aldrich, #T8787) and 1%BSA (Sigma-Aldrich #A9576)) for 30 minutes. A murine antibody binding the N-terminus of human and murine dystrophin (not cross-reacting with utrophin) and a rat anti-laminin-2 antibody (Sigma-Aldrich #L0663) were diluted in blocking buffer. Blocking buffer was then removed with a vacuum aspirator and 200µl of

primary antibody solution was added to each slide. Following a one hour incubation at room temperature, slides were washed 3 times in DPBS. A secondary antibody solution was prepared for the detection of the primary antibodies by diluting an Alexafluor 546 goat anti-rat antibody (ThermoFisher, Waltham, MA, #A11077) and an Alexafluor 647 goat anti-mouse IgG2b antibody (ThermoFisher, Waltham, MA, #A21242) in blocking buffer (see above). DAPI was also included in the secondary antibody solution to counterstain nuclei in the tissue. 200µl of secondary antibody solution was added to the tissue and incubated for 30 minutes at room temperature. Following the staining protocol, slides were washed 3 times with DPBS followed by a rinse with diH₂O. One drop of ProLong diamond antifade mountant (ThermoFisher, Waltham, MA, #P36962) was added to each slide and each slide was then sealed with a coverslip. Slides were stored at 4°C for imaging the next day.

Immuno-fluorescence image acquisition - Fluorescence image acquisition of fluorescently labeled tissue sections was conducted on a Leica SP8 confocal microscope (Leica Microsystems; see FIGS. 12, 17, 19 and 20) or an Opera Phenix™ HCS imager (PerkinElmer, Waltham, MA; see FIGS. 15 and 16) equipped with a laser microlens confocal and large 4.7 Mpixel CMOS camera. Fluorescent dyes used for labeling tissues were matched with appropriate laser excitation light sources and complementary emission filters (Nuclei (DAPI): ex 375 nm, em 435-480 nm; miniaturized dystrophin (AF647): ex 640nm, em 650-760 nm; laminin (AF546): ex 561 nm, em 570-630 nm). The software package Harmony 4.9 was used for image acquisition. The software first performed a low magnification scan at 5x to identify the region of interest (ROI). A second round of multi-color image acquisition on the ROI was performed using a water objective lens at 20x magnification. A montage image of the ROI was captured with 20% overlap between fields of view. Images were imported into the Columbus™ Image Data Storage and Analysis System (PerkinElmer, Waltham, MA) for analysis and quantitation.

Immuno-fluorescence image analysis - A building block analysis routine was created in the Columbus™ Image Data Storage and Analysis System to identify muscle fibers in both heart and skeletal muscle tissue and quantitate the amount of miniaturized dystrophin staining. A global image of the entire tissue was created. Each field of view was inverted so that the software could identify “cells” that were outlined by laminin staining.

Size and intensity filters were applied to identify only true muscle fibers. The outer membrane identified by laminin staining was dilated and the miniaturized dystrophin intensity inside this region was calculated. Intensities were calculated for all tissues for all animal groups. Intensity cutoffs for “cells” or muscle fibers positive for miniaturized dystrophin were determined from the tissue of DMD mice, using a mean intensity plus 3 standard deviations. The proportion of laminin⁺ muscle fibers also positive for mini-dystrophin protein and the proportion of laminin⁺ muscle fibers with central nuclei were determined.

Standard histology - Tissue slides prepared as described above were also used for standard histology.

Example 9: Analysis of the dystrophin-glycoprotein complex (DGC) in muscle fibers of *mdx^{scsn}* mice untreated and treated with miniaturized dystrophins

To test if miniaturized dystrophin restored components of the dystrophin-glycoprotein complex (DGC), the diaphragm muscles from *mdx^{scsn}* mice and *mdx^{scsn}* mice treated with either BXA-220931 or with BXA-212374 miniaturized dystrophin as described in Example 8 were analyzed by immune-fluorescence histology, in principle as described in Example 8. Briefly, frozen OCT sections were incubated in blocking buffer (1x PBS, 1% BSA, 0.05% Triton) for 30min, then incubated with primary antibodies to nNOS (1:200; ThermoFisher, Waltham, MA, #61-7000), β -sarcoglycan (1:20; Novus Biologicals, Centennial, CO, #NBP1-90300), or β -syntrophin (1:200; Novus Biologicals, Centennial, CO, NB600-1294) for 1hr, washed three times in 1x PBS, and then incubated with secondary antibodies conjugated to Alexa-488 (1:800 ThermoFisher, Waltham, MA) for 30 min, washed three times in 1x PBS, and coverslipped with prolong gold mounting medium with DAPI. The data showed that BXA-220931 restored dystrophin glycoprotein

complex (DGC) components including nNOS to the sarcolemma of treated mice, whereas BXA-212374 was unable to restore nNOS to the sarcolemma (see FIG. 17).

5 Example 10: Analysis of muscle mass in *mdx^{scsn}* mice untreated and treated with miniaturized dystrophins

Typically, muscle mass is heavier in *mdx^{scsn}* mice due to the significant muscle degeneration and regeneration. The mass of tibialis anterior muscle in untreated and treated mice (as described in Example 8) was determined. Our analysis showed that *mdx^{scsn}* mice treated with both BXA-220931 or BXA-212374 miniaturized dystrophins maintained
10 normal muscle mass (see FIG. 18).

Example 11: Analysis of costameres in muscles of *mdx^{scsn}* mice untreated and treated with miniaturized dystrophins

To immunostain costameres in muscles of *mdx^{scsn}* mice untreated and treated with
15 miniaturized dystrophins as described in Example 8, a method similar to Williams M.W. and Bloch R.J., *Extensive but coordinated reorganization of the membrane skeleton in myofibers of dystrophic (mdx) mice*, *J. Cell. Biol.* 144(6):1259-70 (1999), was used. Briefly, the *mdx^{scsn}* mice were anesthetized and perfusion fixed with 2% paraformaldehyde in 1xPBS. The gastrocnemius muscles were then dissected, placed in
20 20% sucrose in 1x PBS for 2 hours at 4°C, placed in a cryovial, and finally snap frozen in liquid N₂. 40µm longitudinal sections were cut from the 3rd digit of the extensor digitorum longus muscle similar to a previously described protocol (Banks G.B. *et al.*, *Muscle structure influences utrophin expression in mdx mice*, *PLoS Genet.* 10(6):e1004431 (2010)) and the tissue was immune-stained with an N-terminal
25 dystrophin antibody (binding both human and murine dystrophin) and an ankyrin G antibody (Santa Cruz Biotechnology, Dallas, TX). The samples were then washed 3 times in 1x PBS and secondary antibodies conjugated to Alexa 488 to label dystrophin and Alexa-594 to label ankyrin G were applied. The samples were then washed again 3 times in 1x PBS and then finally mounted with ProLong™ Gold antifade mountant containing
30 DAPI. Images were gathered using a Leica SP8 confocal microscope (Leica Microsystems). The data showed that miniaturized dystrophins localized to both the Z-

disks and M bands of costameres similar to dystrophin in wild-type muscles (see FIG. 19).

5 Example 12: Analysis of neuromuscular junctions of *mdx^{scsn}* mice untreated and treated with miniaturized dystrophins

The neuromuscular junctions in *mdx^{scsn}* mice untreated and treated with miniaturized dystrophins as described in Example 8 were labelled with α -bungarotoxin in the third digit of the extensor digitorum longus muscles according to Faber R.M. *et al.*, *Myofiber branching rather than myofiber hyperplasia contributes to muscle hypertrophy* 10 *in mdx mice*, *Skelet. Muscle* 4:10 (2014). The analysis of neuromuscular junctions in *mdx^{scsn}* mice by α -bungarotoxin staining showed that the postsynaptic apparatus fragments upon muscle degeneration in muscles of untreated *mdx^{scsn}* mice, but that both BXA-220931 and BXA-212374 miniaturized dystrophins prevented synaptic fragmentation in mice treated with the respective AAV (see FIG. 20).

15

Example 13: Analysis of serum creatine kinase levels in *mdx^{scsn}* mice untreated and treated with miniaturized dystrophins

Creatine kinase as an indicator of muscle damage was measured in serum using commercially available kits. Creatine kinase was measured at 4 weeks of age (2-weeks post virus delivery) and 12 weeks of age. The data indicated that in *mdx^{scsn}* mice treated with 20 AAV9-BXA-220931 or AAV9-BXA-212374 as described in Example 8, serum creatine kinase levels and thus muscle damage were significantly reduced (not shown).

The AAV used herein was AAV9 or AAV8, wherein the ITRs were AAV2.

25

Example 14: Functional *in vivo* studies

Dystrophin-deficient skeletal muscles produce less specific force (force per area) and are highly susceptible to contraction-induced injury. Restoration of dystrophin expression can mitigate these disorders. Dystrophic *mdx* mice are systemically treated with 30 about 2e14vg/kg AAV9-C5-12(T)-BXA-220931 (SEQ ID NO: 83) at 2 weeks of age by retro-orbital injection. The limb muscle physiology is examined at 8 weeks of age. Briefly,

the mouse knee is clamped and the foot is placed in a stirrup, and the stirrup is moved while the muscles are maximally contracted with a needle electrode. This assay measures the peak twitch and tetanic muscle force production and contraction-induced injury.

The tibialis anterior (TA) muscle contractile properties are tested by an *in vivo* (foot plate) apparatus as per manufacturer's instructions (Aurora Scientific). Briefly, the peak tetanic contraction is achieved at 150Hz in force frequency curve (force is measured in Torque as Newton Meters). The peak tetanic contraction is the same in wild-type, *mdx^{scsn}* and *mdx^{scsn}* mice treated with BXA-220931. However, the TA muscle mass is greater in *mdx^{scsn}* mice, such that peak tetanic force normalized to TA mass is reduced in *mdx^{scsn}* mice, whereas it is at wild-type levels in the treated *mdx^{scsn}* mice.

The right tibialis anterior muscle is examined for strength and resistance to contraction-induced injury similar to previously described protocols (Khairallah *et. al.*, *Science Signaling* 5(236) (2012). The tibialis anterior (TA) muscle injury is measured by an *in vivo* (foot plate) apparatus as per manufacturers instructions (Aurora Scientific). During peak tetanic contraction at 150Hz (maximum isometric torque), the foot plate is rotated from 90° to 135° degrees to strain the muscles. This contraction is repeated every minute for 20 contractions as previously described (Khairallah *et. al.*, 2012). The maximum isometric torque immediately prior to strain is significantly reduced with each contraction in *mdx^{scsn}* mice. In contrast, BXA-220931 prevents the contraction-induced injury similar to wild-type levels. The data are to show that the miniaturized dystrophin design protects the TA muscles from contraction-induced injury.

In vitro and *in vivo* expression of miniaturized dystrophin constructs is under the control of a C5-12(T) promoter (see, e.g., US 2004/0175727). The AAV used is AAV9 or AAV8, wherein the ITRs are AAV2.