

(19) World Intellectual Property  
Organization

International Bureau

(43) International Publication Date  
16 January 2025 (16.01.2025)



(10) International Publication Number  
**WO 2025/015036 A9**

(51) International Patent Classification:

*A61K 38/17* (2006.01) *C12N 15/62* (2006.01)  
*A61K 38/43* (2006.01) *C12N 15/864* (2006.01)  
*A61K 48/00* (2006.01) *C12P 21/00* (2006.01)  
*A61P 21/00* (2006.01) *A61K 31/7088* (2006.01)  
*C07K 14/00* (2006.01)

(21) International Application Number:

PCT/US2024/037366

(22) International Filing Date:

10 July 2024 (10.07.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/525,819 10 July 2023 (10.07.2023) US

(71) Applicant: **KATE THERAPEUTICS, INC.** [US/US];  
10675 Sorrento Valley Road, Suite 150, San Diego, CA  
92121 (US).

(72) Inventors: **TABEBORDBAR, Mohammadsharif**, 10675  
Sorrento Valley Road, Suite 150, San Diego, CA 92121

(US). **VELEZ, Danielle**, 10675 Sorrento Valley Road,  
Suite 150, San Diego, CA 92121 (US). **FIELDEN, Mark**;  
10675 Sorrento Valley Road, Suite 150, San Diego, CA  
92121 (US).

(74) Agent: **SCHOEN, Adam, M.** et al.; Brown Rudnick LLP,  
One Financial Center, Boston, MA 02111 (US).

(81) Designated States (*unless otherwise indicated, for every  
kind of national protection available*): AE, AG, AL, AM,  
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,  
CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM,  
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG,  
KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,  
MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA,  
NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO,  
RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,  
TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,  
ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every  
kind of regional protection available*): ARIPO (BW, CV,  
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,

(54) Title: AAV GENE THERAPY APPROACH FOR TREATING LGMD2B/DYSFERLINOPATHY

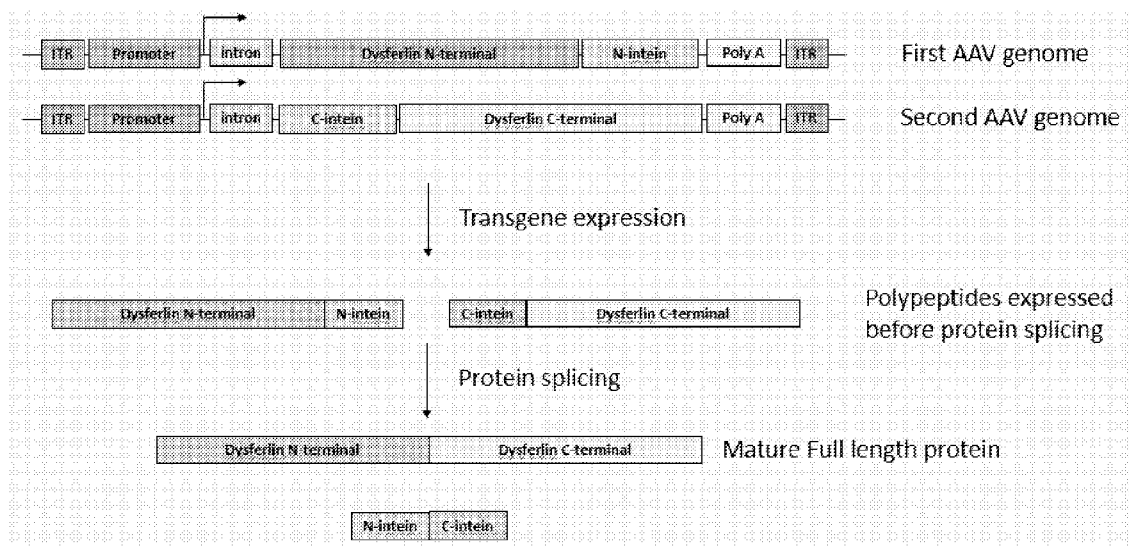


FIG. 1

(57) Abstract: This disclosure relates to compositions and methods for the treatment of dysferlinopathies.



SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

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

**(48) Date of publication of this corrected version:**

29 January 2026 (29.01.2026)

**(15) Information about Correction:**

see Notice of 29 January 2026 (29.01.2026)

## AAV GENE THERAPY APPROACH FOR TREATING LGMD2B/DYSFERLINOPATHY

### FIELD OF DISCLOSURE

This disclosure relates to compositions for the treatment of dysferlinopathies.

5

### BACKGROUND

Mutations in the human dysferlin gene ( DYSF) cause autosomal recessive muscular dystrophies characterized by degeneration and weakness of proximal and/or distal muscles: limb girdle muscular dystrophy type 2B (LGMD2B) and Miyoshi myopathy (MM). The majority of mutations in DYSF lead to the absence or severe reduction in the amount of dysferlin protein present in the muscle cells.

10

The lack of dysferlin in muscle cells leads to lack of muscle fiber repair and eventually muscle cell death. Most individuals with dysferlinopathy eventually require the assistance of a wheelchair within 10-20 years of the onset of muscle weakness. Everyday tasks, including standing from a seating position, can become very difficult or nearly impossible.

15

There is no cure for dysferlinopathy and few effective treatments. Physical therapy is often required simply to retain muscle strength and mobility for as long as possible. As muscles deteriorate, a ventilator may be required to aid breathing.

### SUMMARY

20

The present invention provides compositions and methods for the treatment of dysferlinopathies that overcome the packaging capacity limits of adeno-associated viral (AAV) vectors. Aspects of the invention provide a composition comprising a first nucleic acid molecule each encoding a separate portion of the dysferlin protein and an intein. For example, the first nucleic acid molecule may comprise a sequence encoding a first portion of the Dysferlin protein and a first portion of a split intein, and the second nucleic acid molecule may comprise a sequence encoding second portion of the Dysferlin protein and a second portion of a split intein. The first nucleic acid molecule and second nucleic acid molecule each encode different portions of the Dysferlin protein that form a functional Dysferlin protein with fused together through protein splicing.

25

Advantageously, each nucleic acid may be packaged into a viral vector, for example an AAV vector.

5           Typical AAV vectors have a packaging capacity of ~4.7Kb. Nucleic acids are almost universally packaged with two terminal ITRs, each about 0.14 Kb in length. Additionally regulatory elements, for example promoter, intron, and poly(A) elements, may be about 0.6-1 Kb. As a result, the maximal allowable cDNA length must be less than 3.8 Kb. Gene therapies are restricted by the packaging capacity limit of adeno-associated viral (AAV) vectors. As a consequence, the size of the DYSF coding sequence (~6,243 bp) prevents packaging of the transgene into traditional AAV serotypes, including AAV9. Dual vector AAV-mediated transgene approaches propose expressing parts of a protein from separate vectors followed up homologous recombination or RNA trans-splicing. These approaches, however, are ineffective in skeletal muscle. Without being limited to a mechanism of action, 10 this is thought to be because such approaches require both AAV vectors to enter the same nucleus, where RNA splicing and homologous recombination need to occur. Given that there are hundreds of nuclei in each muscle fiber, efficiency of homologous recombination in post-mitotic tissues (for example, skeletal muscle) is extremely low.

Without being limited to a mechanism of action, aspects of the present invention 20 overcome these limitations by allowing each portion of the dysferlin protein to be expressed in different nuclei of multi-nucleated cells. Once expressed, the two portions of the Dysferlin protein may be joined by splicing of the respective inteins within the cytoplasm of the muscle cells. Splicing of the respective inteins in the cytoplasm results in formation of a functional Dysferlin protein.

25           In aspects of the invention provide, the first portion of the dysferlin protein may be between 1100 and 1150 amino acids in length. The second portion of the dysferlin protein may be between 930 and 980 amino acids in length. For example, the first portion of the dysferlin protein may comprise the C2A-C2C, Fer domains, and dysF domains of the DYSF gene and the second portion of the dysferlin protein may comprise the C2D-C2G domains, as 30 well as the transmembrane domain. The dysferlin protein may be sliced between amino acids 1100 and 1136, counted from the N terminus of the dysferlin protein.

In aspects of the invention, the first nucleic acid molecule may be packaged in a first vector and the second nucleic acid molecule is packaged in a second vector. Each vector may be an adeno-associated viral (AAV) vector. The first nucleic acid molecule may comprise in order, an ITR region, a promoter, the sequence encoding a first portion of the Dysferlin protein, the first intein, an ITR region. The second nucleic acid molecule may comprise, in order, an ITR region, a promoter, the second intein, the sequence encoding a second portion of the Dysferlin protein, an ITR region. Accordingly, the first nucleic acid molecule may have an intein N-terminal to the sequence encoding the first portion of the dysferlin protein and the second nucleic acid molecule may have an intein C-terminal to the sequence encoding the second portion of the dysferlin protein.

Further advantageously, each AAV vector may comprise a capsid protein having at least one modification that results in reduced liver-tropism of the AAV vector and/or preferential targeting of the AAV vector to muscle tissue. For example, the AAV vector may comprise a capsid protein selected from the sequences in Tables 1-4

The present invention also provides methods of treating a subject suffering from a dysferlinopathy comprising providing to the subject a composition comprising a first nucleic acid and second nucleic acid molecule each encoding a separate portion of the dysferlin protein and an intein. The dysferlinopathy may be any dysferlinopathy, for example miyoshi myopathy (MM), Limb-girdle muscular dystrophy type 2B (LGMD2B), and Scapuloperoneal syndrome.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram showing methods of the invention.

FIG. 2 is a crystal structure for dysferlin.

FIG. 3 is a diagram of dysferlin protein and its domains.

FIG. 4 is a diagram of the nucleic acid molecules of the invention.

FIG. 5 is a western blot of cells transfected with the dysferlin-intein plasmids using an anti-HA tag antibody.

FIG. 6 is a western blot of cells transfected with the dysferlin-intein plasmids using an anti-FLAG tag antibody.

FIG. 7 is a diagram of the nucleic acid molecules of the invention where the Dysferlin-intein is expressed under the control of a muscle-specific CK8 promoter.

FIG. 8 is a western blot of C2C12 myotubes transduced with the AAV-dysferlin-intein using an anti-HA tag antibody.

5        FIG. 9 is a western blot of C2C12 myotubes transduced with the AAV-dysferlin-intein using an anti-FLAG tag antibody.

FIG. 10 is a western blot of dysferlin knock out C2C12 myotubes transduced with the AAV-dysferlin-intein using an anti-HA tag antibody.

10       FIG. 11 is a western blot of dysferlin knock out C2C12 myotubes transduced with the AAV-dysferlin-intein using an anti-FLAG tag antibody.

FIG. 12 is a western blot of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein using an anti-dysferlin N-terminal antibody.

FIG. 13 is a bar graph of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein.

15       FIG. 14 is a western blot of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein using an anti-dysferlin C-terminal antibody.

FIG. 15 is a bar graph of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein.

FIG. 16 is a schematic of the split intein strategy.

20       FIG. 17 is western blot a screen of dysferlin inteins.

FIG. 18A-B shows capillary electrophoresis results following split intein transduction in mouse myotubes.

FIG. 19A-D show the expression of dysferlin in knock out mice following vector administration.

25       FIG. 20 show images of dysferlin protein expression in the Tibialis Anterior muscle following vector administration.

FIG. 21 shows images of membrane repair in mice following vector administration.

FIG. 22A-B are graphs showing quantification of membrane repair in mice following vector administration.

## DETAILED DESCRIPTION

The present invention provides compositions and methods for the treatment of dysferlinopathies that overcome the packaging capacity limits of adeno-associated viral (AAV) vectors. Aspects of the invention provide a composition comprising a first nucleic acid and second nucleic acid molecule each encoding a separate portion of the dysferlin protein and an intein. For example, the first nucleic acid molecule may comprise a sequence encoding a first portion of the Dysferlin protein and a first portion of a split intein and the second nucleic acid molecule may comprise a sequence encoding second portion of the Dysferlin protein and a second portion of a split intein. The first nucleic acid molecule and second nucleic acid molecule each encode different portions of the Dysferlin protein that form a functional Dysferlin protein when fused together through protein splicing.

Only by the present invention was it discovered that dysferlin can be packaged and reconstituted by trans-intein splicing. The present invention recognizes that trans-splicing of large proteins is often impossible. For example, intein-mediated protein trans-splicing has been attempted reconstitute dystrophin in muscle and the L-type calcium channel in cardiomyocytes, however these attempts were met with limited efficacy. *See, e.g. Tornabene (2019) "Intein-mediated protein trans-splicing expands adeno-associated virus transfer capacity in the retina", Sci Transl Med. 11(492), the contents of which are incorporated by reference herein in their entirety.* Construct design needs to take into account junction points of amino acid residues needed for efficient protein trans-splicing and splitting of the proteins outside of structural domains to avoid incorrect polypeptide folding.

**FIG. 1** is a diagram showing methods of the invention. As shown a first DNA molecule encodes an N-terminal portion of the dysferlin protein with a N-terminal portion of a split intein and a second DNA molecule encodes a C-terminal portion of the dysferlin protein with a C-terminal portion of a split intein. Following transcription of the DNA, precursor polypeptides are generated having the N-terminal portion of the dysferlin protein fused with the N-terminal portion of the split intein, and the C-terminal portion of the dysferlin protein fused with the C-terminal portion of the split intein. The precursor polypeptides are joined at their respective intein insertion sites and the inteins are then spliced leaving the mature full-length protein. Advantageously, splicing of the inteins can occur within the cytoplasm of cells, without being restricted to processing in the nucleus.

**FIG. 2** is a crystal structure for dysferlin showing an exemplary site of the invention for splitting dysferlin.

**FIG. 3** is a diagram of dysferlin protein and its domains showing an exemplary site of the invention for splitting dysferlin.

5       As indicated by the respective arrows, the dysferlin protein may be sliced between amino acids 1100 and 1136, counted from the N terminus of the dysferlin protein. As shown, this region of the protein is surface-exposed and structurally disordered.

### Dysferlinopathies

10       Dysferlin is a protein that in humans is encoded by the DYSF gene. Dysferlin is linked with stabilization of calcium signaling and muscle fiber repair. Defects in the DYSF gene can result in several types of muscular dystrophy referred to as dysferlinopathies. Dysferlinopathies are characterized by a slow progression of muscle weakness and atrophy or wasting. Dysferlinopathies include Miyoshi myopathy (MM), generally characterized by  
15       defects in the muscles of distal legs close to the feet, Limb-girdle muscular dystrophy type 2B (LGMD2B), characterized by defects in the muscles of the pelvic and shoulder girdles, and Scapuloperoneal syndrome, where weakness and atrophy affect both the distal legs and shoulder girdle muscles.

20       The human dysferlin protein is a 237 kilodalton (2080 aa) type-II transmembrane protein. The cytosolic domain of dysferlin is composed of seven highly conserved C2 domains (C2A-G). Dysferlin also contains “Fer” and “dysF” domains between the C2C and C2D domains. Mutations in each of these domains can cause dysferlinopathy.

### Adeno Associated Virus Vectors

25       AAVs are particularly appropriate viral vectors for delivery of genetic material into mammalian cells. AAVs are not known to cause disease in mammals and cause a very mild immune response. Additionally, AAVs are able to infect cells in multiple stages whether at rest or in a phase of the cell replication cycle. Advantageously, AAV DNA is not regularly inserted into the host's genome at random sites, reducing the oncogenic properties of this  
30       vector.

AAVs have been engineered to deliver a variety of treatments, especially for genetic disorders caused by single nucleotide polymorphisms (“SNP”). Genetic diseases that have been studied in conjunction with AAV vectors include Cystic fibrosis, hemophilia, arthritis, macular degeneration, muscular dystrophy, Parkinson’s disease, congestive heart failure, and Alzheimer’s disease. The AAV can be used as a vector to deliver engineered nucleic acid to a host and utilize the host’s own ribosomes to transcribe that nucleic acid into the desired proteins. See, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); and Muzyczka, *J. Clin. Invest.* 94:1351 (1994). AAVs have some deficiency in their replication and/or pathogenicity and thus can be safer than adenoviral vectors. In some embodiments, the AAV can integrate into a specific site on chromosome 19 of a human cell with no observable side effects. In some embodiments, the capacity of the AAV vector, system thereof, and/or AAV particles can be up to about 4.7 kb. The AAV vector or system thereof can include one or more engineered capsid polynucleotides described herein.

AAVs are small, replication-defective, nonenveloped viruses that infect humans and other primate species and have a linear single-stranded DNA genome. Naturally occurring AAV serotypes exhibit liver tropism. As a result, transfection of non-liver tissue with traditional AAV vectors is impeded by the virus’s natural liver tropism. Moreover, because the liver acts to break down substances delivered to a subject, transfection of non-liver tissue with unmodified AAV vectors requires higher dosing to provide sufficient viral load to overcome the liver and reach non-liver tissue. More than 30 naturally occurring serotypes of AAV are available. Many natural variants in the AAV capsid exist. AAV serotypes include, but are not limited to, AAV serotypes AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV 12, AAV13. AAVs may be engineered using conventional molecular biology techniques, making it possible to optimize these particles, for example, for cell specific delivery, for minimizing immunogenicity, for tuning stability and particle lifetime, for efficient degradation, for accurate delivery to the nucleus. AAV vectors can be specifically targeted to one or more types of cells by choosing the appropriate combination of AAV serotype, promoter, and delivery method.

Previous approaches to identify AAV sequences correlated with tropism have relied upon the comparison of highly related extant serotypes with distinct characteristics, random

domain swaps between unrelated serotypes, or consideration of higher-order structure, to identify motifs that define liver tropism. For example, mapping determinants of AAV tropism have been carried out by comparing highly related serotypes. One such example is the single-amino acid change (E531K) between AAV1 and AAV6 that improves murine liver transduction in AAV1. See Wu et al. (2006) J. Virol., 80(22):11393-7, incorporated by  
5 reference herein. Another example is a reciprocal domain swap between AAV2 and AAV8 that alters tropism, but fails to define any robust specific tissue-targeting motifs. See Raupp et al. (201) J. Virol., 86(17):9396-408, incorporated by reference herein. Further, global consideration of structure has only highlighted gross differences between better- or worse-  
10 liver-transducers that are more observational than useful in practice. Nam et al (2007) J. Virol., 81(22):12260-71.

AAVs exhibiting modified tissue tropism that may be used with the present invention are described in U.S. Patent No. 9,695,220, U.S. Patent No. 9,719,070; U.S. Patent No. 10,119,125; U.S. Patent No. 10,526,584; U.S. Patent Application Publication No. 2018-  
15 0369414; U.S. Patent Application Publication No. 2020-0123504; U.S. Patent Application Publication No. 2020-0318082; PCT International Patent Application Publication No. WO 2015/054653; PCT International Patent Application Publication No. WO 2016/179496; PCT International Patent Application Publication No. WO 2017/100791; and PCT International Patent Application Publication No. WO 2019/217911, the entirety of the contents of each of  
20 which are incorporated by reference herein.

The AAV vector or system thereof may include one or more regulatory molecules, such as promoters, enhancers, repressors and the like. In some embodiments, the AAV vector or system thereof can include one or more polynucleotides that can encode one or more regulatory proteins. In some embodiments, the one or more regulatory proteins can be  
25 selected from Rep78, Rep68, Rep52, Rep40, variants thereof, and combinations thereof. In some embodiments, the muscle specific promoter can drive expression of an engineered AAV capsid polynucleotide.

The AAV vector or system thereof can include one or more polynucleotides that can encode one or more capsid proteins, such as the engineered AAV capsid proteins described  
30 elsewhere herein. The engineered capsid proteins can be capable of assembling into a protein

shell (an engineered capsid) of the AAV virus particle. The engineered capsid can have a cell-, tissue-, and/or organ-specific tropism.

The AAV vector or system thereof can be configured to produce AAV particles having a specific serotype. In some embodiments, the serotype can be AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, AAV-9 or any combinations thereof. In some  
5     embodiments, the AAV can be AAV1, AAV-2, AAV-5, AAV-9 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5, 9 or a hybrid capsid AAV-1, AAV-2, AAV-5, AAV-9 or any combination thereof for targeting brain and/or neuronal cells; and one can select AAV-4 for  
10    targeting cardiac tissue; and one can select AAV-8 for delivery to the liver. Thus, in some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting the brain and/or neuronal cells can be configured to generate AAV particles having serotypes 1, 2, 5 or a hybrid capsid AAV-1, AAV-2, AAV-5 or any combination thereof. In some embodiments, an AAV vector or system thereof capable of producing AAV  
15    particles capable of targeting cardiac tissue can be configured to generate an AAV particle having an AAV-4 serotype. In some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting the liver can be configured to generate an AAV having an AAV-8 serotype. See also Srivastava. 2017. Curr. Opin. Virol. 21:75-80.

It will be appreciated that while the different serotypes can provide some level of cell,  
20    tissue, and/or organ specificity, each serotype still is multi-tropic and thus can result in tissue-toxicity if using that serotype to target a tissue that the serotype is less efficient in transducing. Thus, in addition to achieving some tissue targeting capacity via selecting an AAV of a particular serotype, it will be appreciated that the tropism of the AAV serotype can be modified by an engineered AAV capsid described herein. As described elsewhere herein,  
25    variants of wild-type AAV of any serotype can be generated via a method described herein and determined to have a particular cell-specific tropism, which can be the same or different as that of the reference wild-type AAV serotype. In some embodiments, the cell, tissue, and/or specificity of the wild-type serotype can be enhanced (e.g., made more selective or specific for a particular cell type that the serotype is already biased towards). For example,  
30    wild-type AAV-9 is biased towards muscle and brain in humans (see e.g., Srivastava. 2017. Curr. Opin. Virol. 21:75-80.) By including an engineered AAV capsid and/or capsid protein

variant of wild-type AAV-9 as described herein, the tropism for nervous cells might be reduced or eliminated and/or the muscle specificity increased such that the nervous specificity appears reduced in comparison, thus enhancing the specificity for muscle as compared to the wild-type AAV-9. As previously mentioned, inclusion of an engineered capsid and/or capsid protein variant of a wild-type AAV serotype can have a different tropism than the wild-type reference AAV serotype. For example, an engineered AAV capsid and/or capsid protein variant of AAV-9 can have specificity for a tissue other than muscle or brain in humans.

In some embodiments, the AAV vector is a hybrid AAV vector or system thereof.

Hybrid AAVs are AAVs that include genomes with elements from one serotype that are packaged into a capsid derived from at least one different serotype. For example, if it is the rAAV2/5 that is to be produced, and if the production method is based on the helper-free, transient transfection method discussed above, the 1st plasmid and the 3rd plasmid (the adeno helper plasmid) will be the same as discussed for rAAV2 production. However, the 2nd plasmid, the pRepCap will be different. In this plasmid, called pRep2/Cap5, the Rep gene is still derived from AAV2, while the Cap gene is derived from AAV5. The production scheme is the same as the above-mentioned approach for AAV2 production. The resulting rAAV is called rAAV2/5, in which the genome is based on recombinant AAV2, while the capsid is based on AAV5. It is assumed the cell or tissue-tropism displayed by this AAV2/5 hybrid virus should be the same as that of AAV5. It will be appreciated that wild-type hybrid AAV particles suffer the same specificity issues as with the non-hybrid wild-type serotypes previously discussed.

Advantages achieved by the wild-type based hybrid AAV systems can be combined with the increased and customizable cell-specificity that can be achieved with the engineered AAV capsids can be combined by generating a hybrid AAV that can include an engineered AAV capsid described elsewhere herein. It will be appreciated that hybrid AAVs can contain an engineered AAV capsid containing a genome with elements from a different serotype than the reference wild-type serotype that the engineered AAV capsid is a variant of. For example, a hybrid AAV can be produced that includes an engineered AAV capsid that is a variant of an AAV-9 serotype that is used to package a genome that contains components (e.g., rep elements) from an AAV-2 serotype. As with wild-type based hybrid AAVs previously

discussed, the tropism of the resulting AAV particle will be that of the engineered AAV capsid.

In some embodiments, the AAV vector or system thereof is configured as a “gutless” vector, similar to that described in connection with a retroviral vector. In some embodiments, the “gutless” AAV vector or system thereof can have the cis-acting viral DNA elements involved in genome amplification and packaging in linkage with the heterologous sequences of interest (e.g., the engineered AAV capsid polynucleotide(s)).

The vectors described herein can be constructed using any suitable process or technique. In some embodiments, one or more suitable recombination and/or cloning methods or techniques can be used to the vector(s) described herein. Suitable recombination and/or cloning techniques and/or methods can include, but not limited to, those described in U.S. Application publication No. US 2004-0171156 A1. Other suitable methods and techniques are described elsewhere herein.

Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989). Any of the techniques and/or methods can be used and/or adapted for constructing an AAV or other vector described herein. AAV vectors are discussed elsewhere herein.

In some embodiments, the vector can have one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a “cloning site”). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors.

Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of an engineered AAV capsid system described herein are as used in the foregoing documents, such as International Patent Application Publications WO 2021/050974 and WO 2021/077000 and PCT International Application No. PCT/US2021/042812, the contents of which are incorporated by reference herein.

Additional AAV vectors are described in International Patent Application Publication WO 2019/2071632, the contents of which are incorporated by reference herein.

Further AAV vectors are described in International Patent Application Publications WO 2020/086881 and WO 2020/235543, the contents of each of which are incorporated by  
5 reference herein.

Further AAV vectors are described in International Patent Application Publications WO 2005/033321; WO 2006/110689; WO 2007/127264; WO 2008/027084; WO 2009/073103; WO 2009/073104; WO 2009/105084; WO 2009/134681; WO 2009/136977; WO 2010/051367; WO 2010/138675; WO 2001/038187; WO 2012/112832; WO  
10 2015/054653; WO 2016/179496; WO 2017/100791; WO 2017/019994; WO 2018/209154; WO 2019/067982; WO 2019/195701; WO 2019/217911; WO 2020/041498; WO 2020/210839; U.S. Patent No. 7,906,111; U.S. Patent No. 9,737,618; U.S. Patent No. 10,265,417; U.S. Patent No. 10,485,883; U.S. Patent No. 10,695,441; U.S. Patent No. 10,722,598; U.S. Patent No. 8,999,678; U.S. Patent No. 10,301,648; U.S. Patent No.  
15 10,626,415; U.S. Patent No. 9,198,984; U.S. Patent No. 10,155,931; U.S. Patent No. 8,524,219; U.S. Patent No. 9,206,238; U.S. Patent No. 8,685,387; U.S. Patent No. 9,359,618; U.S. Patent No. 8,231,880; U.S. Patent No. 8,470,310; U.S. Patent No. 9,597,363; U.S. Patent No. 8,940,290; U.S. Patent No. 9,593,346; U.S. Patent No. 10,501,757; U.S. Patent No. 10,786,568; U.S. Patent No. 10,973,928; U.S. Patent No. 10,519,198; U.S. Patent No.  
20 8,846,031; U.S. Patent No. 9,617,561; U.S. Patent No. 9,884,071; U.S. Patent No. 10,406,173; U.S. Patent No. 9,596,220; U.S. Patent No. 9,719,010; U.S. Patent No. 10,117,125; U.S. Patent No. 10,526,584; U.S. Patent No. 10,881,548; U.S. Patent No. 10,738,087; U.S. Patent Publication No. 2011-023353; U.S. Patent Publication No. 2019-0015527; U.S. Patent Publication No. 2020-155704; U.S. Patent Publication No. 2017-0191079; U.S. Patent Publication No. 2019-0218574; U.S. Patent Publication No. 2020-0208176; U.S. Patent Publication No. 2020-0325491; U.S. Patent Publication No. 2019-0055523; U.S. Patent Publication No. 2020-0385689; U.S. Patent Publication No. 2009-0317417; U.S. Patent Publication No. 2016-0051603; U.S. Patent Publication No. 2016-00244783; U.S. Patent Publication No. 2017-0183636; U.S. Patent Publication No. 2020-0263201; U.S. Patent Publication No. 2020-0101099; U.S. Patent Publication No. 2020-0318082; U.S. Patent Publication No. 2018-0369414; U.S. Patent Publication No. 2019-

0330278; U.S. Patent Publication No. 2020-0231986, the contents of each of which are incorporated by reference herein.

MyoAAV Capsid Variants

5

Table: MyoAAV (eMyoAAV) Capsid Variants

| MyoAAV Capsid Variants |             |             |                                   |             |
|------------------------|-------------|-------------|-----------------------------------|-------------|
| RANK                   | N-MER MOTIF | SEQ ID NO.: | ENCODING SEQUENCE                 | SEQ ID NO.: |
| 1                      | AEGRGDQYTR  | 177         | GCTGAGGGTAGAGGAGACCAGTATACTCGT    | 181         |
| 2                      | GPGRGDQTTL  | 178         | GGGCCGGGTAGAGGAGACCAGACTACGTTG    | 182         |
| 3                      | AEGRGDQYTR  | 177         | GCGGAAGGCAGAGGAGACCAATACACAAGG    | 183         |
| 4                      | ATGRGDLGQA  | 179         | GCGACTGGTAGAGGAGACCTGGGTTCAGGCT   | 184         |
| 5                      | AVARGDQGLI  | 180         | GCGGTGGCGAGAGGAGACCAGGGTCTTATT    | 185         |
| 6                      | NISRGDQGYQ  | 6           | AACATCTCCAGAGGAGACCAAGGTTACCAA    | 186         |
| 7                      | APARGDQGSQ  | 7           | GCGCCGGCTAGAGGAGACCAGGGGAGTCAG    | 187         |
| 8                      | AVSRGDRMEF  | 8           | GCCGTTAGCAGAGGAGACCGGATGGAATTC    | 188         |
| 9                      | QMGRGDMGIK  | 10          | CAGATGGGTAGAGGAGACATGGGGATTAAG    | 189         |
| 10                     | EYRRGDKADI  | 11          | GAGTATCGGAGAGGAGACAAGGCGGATATT    | 190         |
| 11                     | ESRRGDKEPL  | 12          | GAGAGTCGGAGAGGAGACAAGGAGCCGCTG    | 191         |
| 12                     | AHMRGDLGGT  | 13          | GCACACATGAGAGGAGACCTAGGCGGCACG    | 192         |
| 13                     | VWQRGDKMDM  | 14          | GTGTGGCAAAGAGGAGACAAAATGGACATG    | 193         |
| 14                     | SIGRGDTGHM  | 15          | TCTATTGGGAGAGGAGACACGGGTCATATG    | 194         |
| 15                     | NIARGDAGQY  | 16          | AATATTGCTAGAGGAGACGCTGGTCAGTAT    | 195         |
| 16                     | AVARGDQGLI  | 180         | GCAGTAGCAAGAGGAGACCAAGGCTTAATC    | 196         |
| 17                     | SVSRGDQGLH  | 17          | TCGGTCTCGAGAGGAGACCAAGGATTGCAC    | 197         |
| 18                     | EYRRGDKADI  | 11          | GAATACAGGAGAGGAGACAAAGCAGACATC    | 198         |
| 19                     | AHMRGDLGGT  | 13          | GCGCATATGAGAGGAGACTTGGGGGGGACT    | 199         |
| 20                     | QIGRGDITHG  | 18          | CAGATTGGTAGAGGAGACATTACTCATGGG    | 200         |
| 21                     | EVRRGDLHGT  | 19          | GAAGTCAGAAGAGGAGACTTGCACGGGACA    | 201         |
| 22                     | SVSRGDVHTM  | 20          | AGTGTCTCAAGAGGAGACGTGCACACGATG    | 202         |
| 23                     | MVTRGDLGTR  | 21          | ATGGTGACTIONAGAGGAGACCTTGGTACGCGG | 203         |
| 24                     | NGGRGDTTHF  | 22          | AACGGCGGGAGAGGAGACACGACGCACTTC    | 204         |
| 25                     | TMGRGDMNSL  | 23          | ACGATGGGGAGAGGAGACATGAATTTCGCTT   | 205         |
| 26                     | AISRGDQGLS  | 24          | GCGATTAGTAGAGGAGACCAGGGTCTTTCT    | 206         |
| 27                     | NDARGDKSTY  | 25          | AACGACGCAAGAGGAGACAAATCCACATAC    | 207         |

|    |            |     |                                 |     |
|----|------------|-----|---------------------------------|-----|
| 28 | VGLRGDLTGS | 26  | GTGGGGCTGAGAGGAGACTTGACGGGTTCG  | 208 |
| 29 | EMRRGDLGTS | 27  | GAGATGAGGAGAGGAGACCTGGGGACGAGT  | 209 |
| 30 | AMSRGDMGMA | 28  | GCGATGAGTAGAGGAGACATGGGGATGGCG  | 210 |
| 31 | ATGRGDLGQA | 179 | GCAACGGGCAGAGGAGACTTAGGCCAAGCA  | 211 |
| 32 | NVARGDQVNY | 29  | AACGTAGCAAGAGGAGACCAAGTTAACTAC  | 212 |
| 33 | SIGRGDTGHM | 15  | AGCATCGGAAGAGGAGACACCGGCCACATG  | 213 |
| 34 | ASVRGDLSGS | 30  | GCATCCGTAAGAGGAGACCTATCAGGCTCA  | 214 |
| 35 | SAARGDTERL | 31  | AGCGCCGCGAGAGGAGACACTGAACGGCTA  | 215 |
| 36 | AMSRGDMGMA | 28  | GCCATGTCAAGAGGAGACATGGGTATGGCT  | 216 |
| 37 | SPSRGDQGR  | 9   | AGTCCATCGAGAGGAGACCAAGGACGCACT  | 217 |
| 38 | VPGRGDLNTM | 32  | GTGCCTGGGAGAGGAGACCTTAATACTATG  | 218 |
| 39 | GPGRGDQTTL | 178 | GGTCCTGGGAGAGGAGACCAAACGACCCTT  | 219 |
| 40 | TSVRGDHGT  | 33  | ACATCGGTCAGAGGAGACCACGGAACATTG  | 220 |
| 41 | TPSRGDLGQT | 34  | ACGCCTTCGAGAGGAGACCTTGGGCAGACT  | 221 |
| 42 | AGHRGDTGVI | 35  | GCTGGACACAGAGGAGACACAGGAGTCATC  | 222 |
| 43 | APARGDQGS  | 36  | GCGCCGGCGAGAGGAGACCAGGGTAGTATG  | 223 |
| 44 | MSLRGDLNGS | 37  | ATGTCGTTGAGAGGAGACTTGAATGGGTCG  | 224 |
| 45 | EAKRGDVHSI | 38  | GAGGCTAAGAGAGGAGACGTGCATTCTATT  | 225 |
| 46 | SAQRGDVQAV | 39  | AGTGCTCAGAGAGGAGACGTGCAGGCGGTG  | 226 |
| 47 | STARGDQGDR | 40  | AGTACGGCGAGAGGAGACCAGGGGGATAGG  | 227 |
| 48 | QISRGDLGIN | 41  | CAGATTAGTAGAGGAGACCTGGGGATTAAT  | 228 |
| 49 | TFTRGDMTMN | 42  | ACGTTCAACAAGAGGAGACATGACGATGAAC | 229 |
| 50 | QEGRGDLNMR | 43  | CAAGAAGGAAGAGGAGACCTGAACATGAGG  | 230 |
| 51 | AHARGDTSSL | 44  | GCACACGCAAGAGGAGACACCAGCTCCCTG  | 231 |
| 52 | TNARGDAGTL | 45  | ACGAATGGGAGAGGAGACGCTGGGACTCTG  | 232 |
| 53 | AFGRGDQQL  | 46  | GCTTTCGGCAGAGGAGACCAAGGGCAACTA  | 233 |
| 54 | ADGRGDRSSL | 47  | GCTGATGGGAGAGGAGACCGTTTCGTCTCTG | 234 |
| 55 | DGRRGDGHS  | 48  | GACGGACGCAGAGGAGACGGACACAGCCTT  | 235 |
| 56 | EKLRGDLHST | 49  | GAGAAGTTGAGAGGAGACCTTCATTCGACT  | 236 |
| 57 | EARRGDASAM | 50  | GAGGCGCGTAGAGGAGACGCTTCGGCGATG  | 237 |
| 58 | MGTRGDKMDF | 51  | ATGGGAACGAGAGGAGACAAAATGGACTTC  | 238 |
| 59 | STSRGDRESY | 52  | TCGACTTCGAGAGGAGACCGGGAGTCGTAT  | 239 |
| 60 | CQPRGDTTRC | 53  | TGTCAGCCGAGAGGAGACACGACTCGGTGT  | 240 |
| 61 | MHTRGDKMDF | 54  | ATGCACACGAGAGGAGACAAAATGGACTTC  | 241 |
| 62 | YSARGDTSGL | 55  | TACAGTGCCAGAGGAGACACAAGTGGCCTG  | 242 |
| 63 | STMRGDHEKL | 56  | TCGACGATGAGAGGAGACCATGAGAAGTTG  | 243 |
| 64 | TQMRGDGVTL | 57  | ACCCAAATGAGAGGAGACGGGGTCACTA    | 244 |
| 65 | TTRRGDMGDN | 58  | ACTACTCGTAGAGGAGACATGGGGGATAAT  | 245 |
| 66 | ANGRGDRLEL | 59  | GCGAATGGTAGAGGAGACAGGCTGGAGTTG  | 246 |

|     |            |    |                                  |     |
|-----|------------|----|----------------------------------|-----|
| 67  | AVSRGDRMEF | 8  | GCTGTGTCTAGAGGAGACAGGATGGAGTTT   | 247 |
| 68  | ENQRGDLGR  | 60 | GAGAATCAGAGAGGAGACTTGTCTGGGTCGG  | 248 |
| 69  | STGRGDLGQA | 61 | TCGACTGGTAGAGGAGACCTGGGGCAGGCT   | 249 |
| 70  | MHTRGDKMDF | 54 | ATGCATACTAGAGGAGACAAGATGGATTTT   | 250 |
| 71  | SSARGDYSEV | 62 | AGTAGTGCGAGAGGAGACTATAGTGAGGTG   | 251 |
| 72  | AFGRGDQGQL | 46 | GCGTTTGGGAGAGGAGACCAGGGTCAGCTT   | 252 |
| 73  | STARGDAATY | 63 | AGTACGGCGAGAGGAGACGCGGCGACTTAT   | 253 |
| 74  | VNTRGDTQKL | 64 | GTCAACACAAGAGGAGACACTCAAAAACCTT  | 254 |
| 75  | AMSRGDHASL | 65 | GCGATGAGTAGAGGAGACCATGCGTCGCTT   | 255 |
| 76  | QAARGDVNKL | 66 | CAGGCTGCTAGAGGAGACGTTAATAAGCTG   | 256 |
| 77  | RGTRGDTVLE | 67 | AGGGGTACGAGAGGAGACACGGTTGAGTTG   | 257 |
| 78  | TTRRGDMGDN | 58 | ACCACCCGGAGAGGAGACATGGGAGACAAC   | 258 |
| 79  | ASTRGDYAGV | 68 | GCTAGTACGAGAGGAGACTATGCGGGTGTT   | 259 |
| 80  | DDRRGDKLPL | 69 | GACGACAGGAGAGGAGACAACTGCCCTT     | 260 |
| 81  | SGSRGDLNAV | 70 | TCGGGGTCGAGAGGAGACCTGAATGCGGTG   | 261 |
| 82  | EYRRGDQQIQ | 71 | GAGTATAGGAGAGGAGACCAGCAGATTCAG   | 262 |
| 83  | TYVRGDRAEV | 72 | ACGTACGTCAGAGGAGACAGAGCAGAAGTG   | 263 |
| 84  | QSLRGDLNGS | 73 | CAATCACTAAGAGGAGACTTGAACGGATCC   | 264 |
| 85  | EGRRGDTTSL | 74 | GAGGGTTCGGAGAGGAGACACGACGTCGTTG  | 265 |
| 86  | AENRGDRSSL | 75 | GCGGAGAATAGAGGAGACCGGTCGAGTTTG   | 266 |
| 87  | VAGRGDRSEL | 76 | GTGGCTGGGAGAGGAGACCGTAGTGAGCTG   | 267 |
| 88  | DPARGDIGAR | 77 | GATCCGGCGAGAGGAGACATTGGTGCGCGT   | 268 |
| 89  | EMRRGDLGTS | 27 | GAAATGCGCAGAGGAGACTTGGGAACCTCA   | 269 |
| 90  | VWQRGDKMDM | 14 | GTTTGGCAGAGAGGAGACAAGATGGATATG   | 270 |
| 91  | VIGRGDKALQ | 78 | GTGATTGGGAGAGGAGACAAGGCGCTTCAG   | 271 |
| 92  | SFERGDKNSL | 79 | AGCTTCGAAAGAGGAGACAAAACTCTCTT    | 272 |
| 93  | QIARGDIASV | 80 | CAGATTGCTAGAGGAGACATTGCTAGTGTG   | 273 |
| 94  | TISRGDLGGA | 81 | ACGATTAGTAGAGGAGACCTGGGTGGGGCT   | 274 |
| 95  | AMIRGDMTHS | 82 | GCGATGATTAGAGGAGACATGACGCATAGT   | 275 |
| 96  | TSVRGDHGT  | 33 | ACGAGTGTGAGAGGAGACCATGGGACGCTG   | 276 |
| 97  | TMRRGDLNDS | 83 | ACGATGCGGAGAGGAGACCTGAATGATAGT   | 277 |
| 98  | AYSRGDLGNH | 84 | GCCTACTCCAGAGGAGACCTAGGCAACCAC   | 278 |
| 99  | EERRGDTHRL | 85 | GAGGAGCGGAGAGGAGACACTCATCGGCTG   | 279 |
| 100 | SQLRGDAGTI | 86 | TCACAACTAAGAGGAGACGCGGGGACAATC   | 280 |
| 101 | AMVRGDAHQL | 87 | GCGATGGTGAGAGGAGACGCGCATCAGCTG   | 281 |
| 102 | TTSRGDLNSV | 88 | ACGACTAGTAGAGGAGACTTGAATTGCGTT   | 282 |
| 103 | WEQRGDMIGK | 89 | TGGGAGCAGAGAGGAGACATGATTGGGAAG   | 283 |
| 104 | YTTRGDLQSN | 90 | TACACCACCAGAGGAGACCTCCAATCAAAC   | 284 |
| 105 | MSQRGDLSHQ | 91 | ATGTGCGCAGAGAGGAGACCTTTTCGCATCAG | 285 |

|     |            |     |                                 |     |
|-----|------------|-----|---------------------------------|-----|
| 106 | ASTRGDTTHM | 92  | GCGAGTACTAGAGGAGACACGACGCATATG  | 286 |
| 107 | SDTRGDTHRL | 93  | TCAGACACCAGAGGAGACACCCACAGACTG  | 287 |
| 108 | TLGRGDQYTS | 94  | ACGCTGGGTAGAGGAGACCAGTATACGAGT  | 288 |
| 109 | SYARGDVREV | 95  | TCGTATGCGAGAGGAGACGTTAGGGAGGTT  | 289 |
| 110 | ETRRGDLSQN | 96  | GAGACTCGTAGAGGAGACCTGAGTCAGAAT  | 290 |
| 111 | TIGRGDLMDK | 97  | ACAATCGGGAGAGGAGACCTGATGGACAAA  | 291 |
| 112 | SYTRGDLSQN | 98  | AGTTACACCAGAGGAGACCTTAGCCAAAAC  | 292 |
| 113 | SKLRGDASEI | 99  | TCGAAGCTGAGAGGAGACGCTAGTGAGATT  | 293 |
| 114 | SMARGDKAEI | 100 | TCGATGGCGAGAGGAGACAAGGCGGAGATT  | 294 |
| 115 | GAVRGDHSSL | 101 | GGAGCCGTCAGAGGAGACCACTCATCGTTA  | 295 |
| 116 | METRGDMSNR | 102 | ATGGAGACGAGAGGAGACATGAGTAATCGT  | 296 |
| 117 | AYTRGDKDAL | 103 | GCGTACACAAGAGGAGACAAAGACGCACTA  | 297 |
| 118 | ENRRGDITQS | 104 | GAGAATCGGAGAGGAGACATTACGCAGTCG  | 298 |
| 119 | STGRGDLGQA | 61  | TCCACGGGCAGAGGAGACCTTGGACAAGCA  | 299 |
| 120 | GAVRGDHSSL | 101 | GGTGCTGTAGAGGAGACCATAGTAGTCTT   | 300 |
| 121 | SQLRGDAGTI | 86  | TCTCAGCTTAGAGGAGACGCTGGTACGATT  | 301 |
| 122 | MLLRGDKNDF | 105 | ATGCTCCTAAGAGGAGACAAAACGACTTC   | 302 |
| 123 | KVQRGDASDW | 106 | AAGGTGCAGAGAGGAGACGCGAGTGATTGG  | 303 |
| 124 | AFTRGDMQTA | 107 | GCGTTTACGAGAGGAGACATGCAGACTGCT  | 304 |
| 125 | QPKRGDVQAQ | 108 | CAGCCTAAGAGAGGAGACGTTCAAGGCTCAG | 305 |
| 126 | ASLRGDMGVS | 109 | GCTAGTCTGAGAGGAGACATGGGGGTGAGT  | 306 |
| 127 | AYSRGDKESF | 110 | GCCTACTCCAGAGGAGACAAAGAATCGTTC  | 307 |
| 128 | RFQRGDLTDA | 111 | CGGTTCCAAAGAGGAGACTTGACCGACGCT  | 308 |
| 129 | SDARGDTMKL | 112 | TCTGATGCTAGAGGAGACACGATGAAGCTG  | 309 |
| 130 | SGARGDTVSL | 113 | AGCGGGGCCAGAGGAGACACAGTAAGTCTC  | 310 |
| 131 | DEKRGDQKHL | 114 | GATGAGAAGAGAGGAGACCAGAAGCATCTT  | 311 |
| 132 | NGVRGDVQNF | 115 | AACGGAGTAAGAGGAGACGTCCAAAACCTTC | 312 |
| 133 | MTLRGDLGGS | 116 | ATGACGTTGAGAGGAGACCTGGGTGGGTCG  | 313 |
| 134 | EARRGDASAM | 50  | GAGGCTAGGAGAGGAGACGCTAGTGCGATG  | 314 |
| 135 | GQERGDLGTR | 117 | GGGCAGGAGAGAGGAGACCTGGGGACTCGG  | 315 |
| 136 | RGARGDLVDA | 118 | AGAGGCGCCAGAGGAGACCTAGTCGACGCT  | 316 |
| 137 | QYSRGDHTDL | 119 | CAATACTCCAGAGGAGACCACACCGACCTT  | 317 |
| 138 | MNTRGDVHAM | 120 | ATGAACACCAGAGGAGACGTGCACGCTATG  | 318 |
| 139 | TVRRGDLATE | 121 | ACGGTACGGAGAGGAGACCTAGCAACCGAA  | 319 |
| 140 | NIARGDAGQY | 16  | AATATTGCGAGAGGAGACGCTGGTCAGTAT  | 320 |
| 141 | SVGRGDKADI | 122 | TCTGTTGGGAGAGGAGACAAGGCTGATATT  | 321 |
| 142 | SHNRGDTGTM | 123 | TCACACAACAGAGGAGACACCGGAACCATG  | 322 |
| 143 | TNTRGDKESV | 124 | ACGAACACCAGAGGAGACAAAGAATCAGTA  | 323 |
| 144 | SIGRGDQYTI | 125 | TCGATTGGTAGAGGAGACCAGTATACGATT  | 324 |

|     |             |     |                                  |     |
|-----|-------------|-----|----------------------------------|-----|
| 145 | QTSRGDAGSW  | 126 | CAGACGAGTAGAGGAGACGCGGGGTCTTGG   | 325 |
| 146 | AENRGDRSSL  | 75  | GCTGAAAACAGAGGAGACCGAAGCAGCCTA   | 326 |
| 147 | AI SRGDVQSL | 127 | GCGATTTTCGAGAGGAGACGTTTCAGTCGTTG | 327 |
| 148 | NAQRGDHGQL  | 128 | AACGCACAAAGAGGAGACCACGGGCAACTG   | 328 |
| 149 | LNSRGDQASV  | 129 | CTCAACTCGAGAGGAGACCAAGCCTCCGTC   | 329 |
| 150 | SGNRGDIGTF  | 130 | TCGGGAAACAGAGGAGACATAGGTACATTC   | 330 |
| 151 | NVGRGDQATM  | 131 | AACGTGGGAAGAGGAGACCAAGCCACAATG   | 331 |
| 152 | SYSRGDTGRL  | 132 | TCGTATTCTAGAGGAGACACTGGTTCGGCTT  | 332 |
| 153 | SHLRGDQALA  | 133 | AGCCACTTAAGAGGAGACCAAGCGTTGGCC   | 333 |
| 154 | YFARGDMTMN  | 134 | TACTTCGCCAGAGGAGACATGACTATGAAC   | 334 |
| 155 | QSQRGDLNPM  | 135 | CAGTCGCAGAGAGGAGACCTTAATCCTATG   | 335 |
| 156 | EQRRGDKTEL  | 136 | GAGCAGAGGAGAGGAGACAAGACGGAGCTG   | 336 |
| 157 | VSARGDAGQL  | 137 | GTGTCTGCGAGAGGAGACGCGGGTCAGTTG   | 337 |
| 158 | SPSRGDQGRT  | 9   | TCTCCGTCTAGAGGAGACCAGGGGCGGACG   | 338 |
| 159 | EASRGDKGTH  | 138 | GAGGCTAGTAGAGGAGACAAGGGTACGCAT   | 339 |
| 160 | LSTRGDMGMQ  | 139 | CTGTCTACTAGAGGAGACATGGGTATGCAG   | 340 |
| 161 | ANMRGDQLHT  | 140 | GCGAATATGAGAGGAGACCAGCTGCATACT   | 341 |
| 162 | ENTRGDLTHA  | 141 | GAGAATACGAGAGGAGACCTTACTCATGCG   | 342 |
| 163 | IERRGDSAGL  | 142 | ATTGAGCGTAGAGGAGACTCGGCGGGGCTG   | 343 |
| 164 | SAMRGDGVSL  | 143 | AGTGCGATGAGAGGAGACGGAGTGAGCTTG   | 344 |
| 165 | TVSRGDQGLS  | 144 | ACGGTTTCGAGAGGAGACCAGGGGTTGTCT   | 345 |
| 166 | SIVRGDLNST  | 145 | TCGATCGTCAGAGGAGACCTGAACTCCACG   | 346 |
| 167 | DVARGDKTNF  | 146 | GACGTAGCGAGAGGAGACAAAACAACTTC    | 347 |
| 168 | WSQRGDLSGT  | 147 | TGGAGCCAAAGAGGAGACCTGAGTGGAACC   | 348 |
| 169 | LNQRGDVSNM  | 148 | CTGAACCAAAGAGGAGACGTTTCCAACATG   | 349 |
| 170 | AALRGDLSGS  | 149 | GCCGCCCTAAGAGGAGACCTATCCGGCTCA   | 350 |
| 171 | RVARGDITDI  | 150 | AGGGTTGCGAGAGGAGACATTACGGATATT   | 351 |
| 172 | STNRGDLNQV  | 151 | TCGACGAACAGAGGAGACCTCAACCAAGTT   | 352 |
| 173 | DSRRGDSVSL  | 152 | GATTCGCGGAGAGGAGACTCTGTGAGTCTT   | 353 |
| 174 | SVGRGDQSQM  | 153 | AGTGTTGGGAGAGGAGACCAGTCTCAGATG   | 354 |
| 175 | SNTRGDTNSL  | 154 | TCCAACACGAGAGGAGACACAACTCCCTA    | 355 |
| 176 | DNRRGDGTTM  | 155 | GATAATCGGAGAGGAGACGGTACGACTATG   | 356 |
| 177 | TPSRGDQGRL  | 156 | ACGCCTTCGAGAGGAGACCAAGGAAGACTA   | 357 |
| 178 | NYSRGDSMTL  | 157 | AACTACAGCAGAGGAGACTCAATGACGCTT   | 358 |
| 179 | MQARGDAGTL  | 158 | ATGCAGGCTAGAGGAGACGCTGGGACTCTG   | 359 |
| 180 | TQSRGDLSGA  | 159 | ACGCAGTCGAGAGGAGACTTGTCTGGTGCG   | 360 |
| 181 | FAQRGDLTGV  | 160 | TTTGCGCAGAGAGGAGACTTGACTGGGGTT   | 361 |
| 182 | APVRGDLIGT  | 161 | GCGCCGGTGAGAGGAGACCTGATTGGTACG   | 362 |
| 183 | SNTRGDTNSL  | 154 | TCGAATACTAGAGGAGACACGAATAGTTTG   | 363 |

|     |            |     |                                |     |
|-----|------------|-----|--------------------------------|-----|
| 184 | EPKRGDLSNT | 162 | GAGCCGAAGAGAGGAGACTTGAGTAATACG | 364 |
| 185 | SVGRGDTYPL | 163 | TCGGTTGGGAGAGGAGACACTTATCCTCTG | 365 |
| 186 | QTARGDMTGH | 164 | CAAACAGCCAGAGGAGACATGACTGGCCAC | 366 |
| 187 | AGARGDLENR | 165 | GCTGGTGCGAGAGGAGACTTGGAGAATCGG | 367 |
| 188 | MSTRGDLNNV | 166 | ATGAGCACTAGAGGAGACCTAAACAACGTC | 368 |
| 189 | EKQRGDLNSM | 167 | GAAAAACAAAGAGGAGACCTCAACAGCATG | 369 |
| 190 | IKTRGDLGHE | 168 | ATTAAGACGAGAGGAGACCTTGGTCATGAG | 370 |
| 191 | IGRRGDMELS | 169 | ATTGGGAGGAGAGGAGACATGGAGTTGTCT | 371 |
| 192 | TTGRGDLKEV | 170 | ACTACGGGTAGAGGAGACTTGAAGGAGGTT | 372 |
| 193 | TDKRGDQNTV | 171 | ACGGATAAGAGAGGAGACCAGAATACTGTG | 373 |
| 194 | SNVRGDREAV | 172 | TCGAATGTTAGAGGAGACCGGGAGGCGGTT | 374 |
| 195 | SVGRGDQSQM | 153 | TCAGTCGGCAGAGGAGACCAAAGTCAAATG | 375 |
| 196 | YANRGDLSHQ | 173 | TACGCCAACAGAGGAGACCTGAGTCACCAA | 376 |
| 197 | QISRGDLGIN | 41  | CAAATATCGAGAGGAGACCTAGGTATCAAC | 377 |
| 198 | SMMRGDAGQL | 174 | TCGATGATGAGAGGAGACGCTGGGCAGTTG | 378 |
| 199 | EARRGDSQGL | 175 | GAGGCGCGGAGAGGAGACAGTCAGGGGCTG | 379 |
| 200 | MPTRGDQTKF | 176 | ATGCCCACAAGAGGAGACCAAACCAAATTC | 380 |

Table: Enhanced MyoAAV (eMyoAAV) Capsid Variants

| Enhanced MyoAAV (eMyoAAV) Capsid Variants |              |             |                                  |             |
|-------------------------------------------|--------------|-------------|----------------------------------|-------------|
| RANK                                      | N-MER INSERT | SEQ ID NO.: | ENCODING SEQUENCE                | SEQ ID NO.: |
| 1                                         | SNSRGDYNLS   | 381         | TCAAACCTCCAGAGGAGACTACAACCTCCTTG | 557         |
| 2                                         | SSARGDRDYL   | 397         | TCGTCGGCTAGAGGAGACCGTGATTATCTG   | 558         |
| 3                                         | STVRGDYTSV   | 398         | AGCACGGTCAGAGGAGACTACACGTCCGTG   | 559         |
| 4                                         | QERRGDYTSM   | 382         | CAGGAGCGGAGAGGAGACTATACGAGTATG   | 560         |
| 5                                         | NGGRGDTTHF   | 386         | AACGGCGGGAGAGGAGACACGACGCACTTC   | 561         |
| 6                                         | ASTRGDHGVL   | 383         | GCGAGTACTAGAGGAGACCATGGTGTGTTG   | 562         |
| 7                                         | SYQRGDQHNH   | 399         | TCGTACCAAAGAGGAGACCAACACAACATG   | 563         |
| 8                                         | NDTRGDRSYM   | 400         | AACGACACAAGAGGAGACCGATCTTACATG   | 564         |
| 9                                         | ENRRGDFNNT   | 384         | GAAAACAGGAGAGGAGACTTCAACAACACT   | 565         |
| 10                                        | SGSRGDLSGH   | 401         | TCCGGCTCGAGAGGAGACTTGTTCAGGACAC  | 566         |
| 11                                        | ENRRGDITQS   | 393         | GAGAATCGGAGAGGAGACATTACGCAGTCG   | 567         |
| 12                                        | DKPRGDRQLL   | 402         | GATAAGCCGAGAGGAGACCGGCAGTTGTTG   | 568         |
| 13                                        | EERRGDTHRL   | 392         | GAGGAGCGGAGAGGAGACACTCATCGGCTG   | 569         |
| 14                                        | AMIRGDMTHS   | 391         | GCGATGATTAGAGGAGACATGACGCATAGT   | 570         |
| 15                                        | SNNRGDYTTM   | 403         | TCAAACAACAGAGGAGACTACACGACAATG   | 571         |
| 16                                        | QGGRGDLSTL   | 404         | CAGGGTGGGAGAGGAGACTTGTTCGACGTTG  | 572         |
| 17                                        | SKDRGDYQTL   | 405         | TCGAAGGATAGAGGAGACTATCAGACTCTT   | 573         |
| 18                                        | ATARGDQMKL   | 406         | GCTACGGCGAGAGGAGACCAGATGAAGCTG   | 574         |
| 19                                        | QHMRGDLTGA   | 407         | CAGCATATGAGAGGAGACTTGACTGGTGCG   | 575         |
| 20                                        | DGTRGDLGRY   | 408         | GACGGCACGAGAGGAGACCTTGGAAGGTAC   | 576         |
| 21                                        | DSRRGDYANL   | 409         | GACTCTCGGAGAGGAGACTACGCAAACCTC   | 577         |
| 22                                        | SAMRGDYTST   | 410         | TCGGCGATGAGAGGAGACTATACTAGTACT   | 578         |
| 23                                        | ENRRGDAQLL   | 411         | GAGAATCGGAGAGGAGACGCGCAGCTGCTG   | 579         |
| 24                                        | TTLRGDHSQL   | 412         | ACGACGCTGAGAGGAGACCATAGTCAGCTT   | 580         |
| 25                                        | DAGRGDMHRM   | 413         | GACGCAGGGAGAGGAGACATGCACCGTATG   | 581         |
| 26                                        | QHTRGDLSS    | 414         | CAGCATAAGAGAGGAGACTTGAGTTCTACT   | 582         |
| 27                                        | QVYRGDRESV   | 415         | CAAGTTTACAGAGGAGACAGGGAATCCGTG   | 583         |
| 28                                        | SSQRGDLAGT   | 416         | TCGAGTCAGAGAGGAGACTTGGCTGGTACT   | 584         |
| 29                                        | TDVRGDRGTF   | 417         | ACGGATGTTAGAGGAGACAGGGGGACGTTT   | 585         |
| 30                                        | SSVRGDREVT   | 418         | TCGTCTGTGAGAGGAGACCGGGAGGTGACG   | 586         |
| 31                                        | ENRRGDLTNA   | 419         | GAGAATAGGAGAGGAGACCTGACTAATGCG   | 587         |
| 32                                        | SYARGDVHSI   | 420         | TCGTATGCGAGAGGAGACGTTCAATTCTATT  | 588         |

|    |            |     |                                 |     |
|----|------------|-----|---------------------------------|-----|
| 33 | DSSRGDLNLR | 421 | GACAGCTCAAGAGGAGACCTCAACCTCCGG  | 589 |
| 34 | SNARGDYSNM | 422 | TCGAACGCGAGAGGAGACTACTCAAACATG  | 590 |
| 35 | SSVRGDHSV  | 423 | TCGTCTGTGAGAGGAGACCATTCCGTGCTT  | 591 |
| 36 | SAFRGDLHAT | 424 | TCGGCTTTTAGAGGAGACCTGCATGCTACT  | 592 |
| 37 | SAMRGDYTST | 410 | TCCGCCATGAGAGGAGACTACACAAGCACG  | 593 |
| 38 | NAQRGDHGQL | 394 | AACGCACAAAGAGGAGACCACGGGCAACTG  | 594 |
| 39 | SDRRGDQLLY | 425 | TCGGATCGGAGAGGAGACCAGCTGTTGTAT  | 595 |
| 40 | SGVRGDRLAV | 426 | AGCGGCGTAAGAGGAGACCGCCTAGCCGTA  | 596 |
| 41 | ASLRGDLST  | 427 | GCCAGCCTAAGAGGAGACCTCAGCTCAACG  | 597 |
| 42 | TNTRGDHLGM | 428 | ACGAACGGAAGAGGAGACCTTCACGGAATG  | 598 |
| 43 | TNTRGDHGML | 429 | ACGAATACGAGAGGAGACCATGGGATGCTG  | 599 |
| 44 | SSARGDRDYL | 397 | AGTTCCGCCAGAGGAGACCGGGACTACCTC  | 600 |
| 45 | EDRRGDLLRT | 430 | GAGGATCGGAGAGGAGACCTTTTGAGGACT  | 601 |
| 46 | SSLRGDLLHS | 431 | TCGTCGCTGAGAGGAGACCTGTTGCATTCT  | 602 |
| 47 | ENRRGDLTNA | 419 | GAAAACCGTAGAGGAGACCTCACGAACGCT  | 603 |
| 48 | DGRRGDRESI | 432 | GACGGCAGGAGAGGAGACCGGGAATCGATC  | 604 |
| 49 | TGTRGDTMTW | 433 | ACTGGAACCAGAGGAGACACAATGACATGG  | 605 |
| 50 | STVRGDYTSV | 398 | TCGACTGTAGAGGAGACTATACTTCTGTT   | 606 |
| 51 | SATRGDHNVL | 434 | TCGGCAACGAGAGGAGACCACAACGTACTG  | 607 |
| 52 | TDRRGDSGTL | 435 | ACGGATCGGAGAGGAGACAGTGGTACTCTG  | 608 |
| 53 | TSARGDLIHT | 436 | ACGAGTGCTAGAGGAGACCTGATTCATACG  | 609 |
| 54 | SGVRGDRLAV | 426 | AGTGGGGTGAGAGGAGACCGGCTGGCGGTG  | 610 |
| 55 | AVERGDLML  | 437 | GCGGTTGAGAGAGGAGACCGTCTTATGCTG  | 611 |
| 56 | DRSRGDTHVL | 438 | GACAGGTCGAGAGGAGACACGCACGTACTC  | 612 |
| 57 | ESTRGDLRVV | 439 | GAGTCGACGAGAGGAGACCTTCGTGTTGTG  | 613 |
| 58 | TTIRGDYREL | 440 | ACGACGATTAGAGGAGACTATCGTGAGTTG  | 614 |
| 59 | TFIRGDLAGA | 441 | ACGTTCATCAGAGGAGACCTGGCAGGTGCG  | 615 |
| 60 | MTSRGDLTTY | 442 | ATGACGAGCAGAGGAGACTTGACAACATAC  | 616 |
| 61 | ADRRGDQLYS | 443 | GCGGATCGTAGAGGAGACCAGCTTTATTCG  | 617 |
| 62 | AHMRGDLGGT | 385 | GCACACATGAGAGGAGACCTAGGCGGCACG  | 618 |
| 63 | GYVRGDLGQH | 444 | GGGTACGTCAGAGGAGACCTGGGGCAACAC  | 619 |
| 64 | DASRGDRSL  | 445 | GATGCGAGTAGAGGAGACAGGACTAGTTTG  | 620 |
| 65 | TEVRGDRSQV | 446 | ACGGAGGTGAGAGGAGACAGGTTCGACGGTT | 621 |
| 66 | SHLRGDQALA | 395 | TCGCATCTGAGAGGAGACCAGGCTCTTGCG  | 622 |
| 67 | SEVRGDRLSI | 447 | TCGGAGGTGAGAGGAGACCGGTTGAGTATT  | 623 |
| 68 | SAARGDTERL | 387 | AGCGCCGCGAGAGGAGACACTGAACGGCTA  | 624 |
| 69 | SDRRGDLSGA | 448 | TCAGACCGTAGAGGAGACCTGTGCGGCGCT  | 625 |
| 70 | SVTRGDRVVI | 449 | AGCGTCACAAGAGGAGACAGAGTAGTCATC  | 626 |
| 71 | SEGRGDRMAL | 450 | TCGGAGGGGAGAGGAGACAGGATGGCTCTT  | 627 |

|     |             |     |                                 |     |
|-----|-------------|-----|---------------------------------|-----|
| 72  | SSTRGDHLSL  | 451 | TCGTCTACTAGAGGAGACCATCTTTTCGCTG | 628 |
| 73  | AMIRGDMTHS  | 391 | GCTATGATCAGAGGAGACATGACCCACTCC  | 629 |
| 74  | SSTRGDQIYV  | 452 | TCGTCGACGAGAGGAGACCAGATTTATGTG  | 630 |
| 75  | EERRGDTHRL  | 392 | GAGGAGCGGAGAGGAGACACTCATCGGTTG  | 631 |
| 76  | SNSRGDYNLSL | 381 | TCGAATTCGAGAGGAGACTATAATAGTCTT  | 632 |
| 77  | SAQRDRMGSP  | 453 | TCCGCACAAAGGGATCGTATGGGTTCCTCCG | 633 |
| 78  | TGTRGDIATF  | 454 | ACGGGTACGAGAGGAGACATTGCGACTTTT  | 634 |
| 79  | SMQRGDLLST  | 455 | TCGATGCAGAGAGGAGACCTGCTTTCTACT  | 635 |
| 80  | EDVRGDRSKL  | 456 | GAGGATGTGAGAGGAGACAGGAGTAAGCTT  | 636 |
| 81  | AHMRGDLGGT  | 385 | GCGCATATGAGAGGAGACTTGGGGGGGACT  | 637 |
| 82  | TSVRGDQSQY  | 457 | ACGAGTGTGAGAGGAGACCAGAGTCAGTAT  | 638 |
| 83  | ADRRGDTGLF  | 458 | GCGGATCGGAGAGGAGACACGGGGTTGTTT  | 639 |
| 84  | NTYRGDRDSL  | 459 | AACACTTACAGAGGAGACCGAGACAGCCTG  | 640 |
| 85  | TDVRGDRGTF  | 417 | ACGGATGTTAGAGGAGACCGGGGGACGTTT  | 641 |
| 86  | SRRGDNLNDL  | 460 | TCTCGTGGGAGAGGAGACCTGAATGATTTG  | 642 |
| 87  | SNVRGDREAV  | 396 | AGCAACGTAAGAGGAGACAGGGAAGCCGTC  | 643 |
| 88  | MDRRGDASIL  | 461 | ATGGACAGGAGAGGAGACGCCTCCATACTG  | 644 |
| 89  | SIWRGDRTEV  | 462 | TCGATATGGAGAGGAGACCGTACAGAAGTC  | 645 |
| 90  | NSQRGDYSGM  | 463 | AACTCACAAAGAGGAGACTACAGTGGAATG  | 646 |
| 91  | SSVRGDIGGI  | 464 | TCTAGCGTAAGAGGAGACATCGGGGGGATC  | 647 |
| 92  | ENQRGDLSGR  | 390 | GAGAATCAGAGAGGAGACTTGTCGGGTCGG  | 648 |
| 93  | EIRRGDLLAQ  | 465 | GAGATTTCGGAGAGGAGACCTTCTGGCTCAG | 649 |
| 94  | SNIRGDHAVM  | 466 | TCGAACATCAGAGGAGACCACGCAGTCATG  | 650 |
| 95  | LDTRGDRSQV  | 467 | CTGGATACGAGAGGAGACCGGAGTCAGGTT  | 651 |
| 96  | ANTRGDHQN   | 468 | GCGAATACGAGAGGAGACCATCAGAATTTT  | 652 |
| 97  | QVYRGDRESV  | 415 | CAGGTGTATAGAGGAGACCGTGAGTCTGTT  | 653 |
| 98  | TSVRGDLSMN  | 469 | ACGTCGGTGAGAGGAGACCTTTTCGATGAAT | 654 |
| 99  | ASTRGDHGVL  | 383 | GCCTCAACGAGAGGAGACCACGGAGTTCTC  | 655 |
| 100 | TTTRGDYLN   | 470 | ACGACGACGAGAGGAGACTATCTTAATACG  | 656 |
| 101 | AERRGDIKEY  | 471 | GCGGAGAGGAGAGGAGACATTAAGGAGTAT  | 657 |
| 102 | SHARGDLGST  | 472 | TCGCATGCTAGAGGAGACTTGGGTTCGACT  | 658 |
| 103 | TSQRGDYVSL  | 473 | ACGTCTCAGAGAGGAGACTATGTTTCTCTT  | 659 |
| 104 | GPGRGDQTTL  | 556 | GGGCCGGGTAGAGGAGACCAGACTACGTTG  | 660 |
| 105 | VNVRGDRGEV  | 474 | GTGAATGTGAGAGGAGACCGTGGTGAGGTG  | 661 |
| 106 | SYRRGDHDQL  | 475 | TCGTATCGTAGAGGAGACCATGATCAGCTT  | 662 |
| 107 | SGTRGDRVDL  | 476 | AGCGGCACAAGAGGAGACAGGGTAGACCTC  | 663 |
| 108 | SHFRGDLHTS  | 477 | TCCCACCTTCAGAGGAGACTTGACACGTCC  | 664 |
| 109 | SDRRGDLSVP  | 478 | TCGGATCGTAGAGGAGACCTTTTCGGTGCCT | 665 |
| 110 | SGTRGDTLIL  | 479 | AGCGGCACCAGAGGAGACACCTTATACTC   | 666 |

|     |             |     |                                  |     |
|-----|-------------|-----|----------------------------------|-----|
| 111 | ENTRGDLHGK  | 480 | GAAAACACCAGAGGAGACCTCCACGGCAA    | 667 |
| 112 | DRARGDTHVL  | 481 | GATCGGGCGAGAGGAGACACGCATGTGCTG   | 668 |
| 113 | ASGRGDIAGL  | 482 | GCGTCTGGTAGAGGAGACATTGCGGGGCTT   | 669 |
| 114 | EGRRGDLFQA  | 483 | GAGGGTCGTAGAGGAGACCTGTTTCAGGCG   | 670 |
| 115 | RSERGDRLEI  | 484 | AGATCCGAAAGAGGAGACAGGCTAGAAATC   | 671 |
| 116 | AAYRGDAHVL  | 485 | GCGGCCTACAGAGGAGACGCACACGTGCTC   | 672 |
| 117 | NDSRGDQHRL  | 486 | AACGACTCGAGAGGAGACCAACACCGATTG   | 673 |
| 118 | SSARGDHSQL  | 487 | TCATCGGCAAGAGGAGACCACTCTCAACTG   | 674 |
| 119 | TNERGDRLSI  | 488 | ACGAATGAGAGAGGAGACCGTCTTTTCGATT  | 675 |
| 120 | SNSRGDYTSV  | 489 | AGCAACTCCAGAGGAGACTACACCAGTGTC   | 676 |
| 121 | NVTRGDHTVM  | 490 | AACGTCACAAGAGGAGACCACACCGTTATG   | 677 |
| 122 | TMVRGDVKGL  | 491 | ACGATGGTGAGAGGAGACGTTAAGGGGCTT   | 678 |
| 123 | TGTRGDIATF  | 454 | ACCGGGACAAGAGGAGACATAGCCACGTTC   | 679 |
| 124 | VQLRGDLAST  | 492 | GTGCAGCTGAGAGGAGACTTGCTTCTACT    | 680 |
| 125 | MNRRGDYSEQ  | 493 | ATGAACCGGAGAGGAGACTACTCGGAACAA   | 681 |
| 126 | ANTRGDLSPV  | 494 | GCGAATACTAGAGGAGACCTGAGTCCGTT    | 682 |
| 127 | TGMRGDAGSL  | 495 | ACGGGTATGAGAGGAGACGCTGGGAGTCTT   | 683 |
| 128 | LDRRGDFSSA  | 496 | CTAGACCGCAGAGGAGACTTCAGCTCCGCG   | 684 |
| 129 | MYHRGDMTSV  | 497 | ATGTACCACAGAGGAGACATGACTTCGGTG   | 685 |
| 130 | GSMRGDLHST  | 498 | GGGTTCGATGAGAGGAGACTTGCAATTCGACG | 686 |
| 131 | SSSRGDFSSV  | 499 | TCGTCGTCCAGAGGAGACTTCAGCTCTGTT   | 687 |
| 132 | TGTRGDLHTY  | 500 | ACGGGTACGAGAGGAGACTTGCACTATTAT   | 688 |
| 133 | YTQRGDLATV  | 501 | TACACCCAAAGAGGAGACCTTGCCACCGTC   | 689 |
| 134 | TDVRGDRMYV  | 502 | ACGGATGTTAGAGGAGACCGGATGTATGTG   | 690 |
| 135 | TSVRGDHGT   | 388 | ACATCGGTCAGAGGAGACCACGGAACATTG   | 691 |
| 136 | SNSRGDLSVN  | 503 | TCGAACTCCAGAGGAGACCTATCGGTAAAC   | 692 |
| 137 | QSGRGDQMTL  | 504 | CAGTCTGGGAGAGGAGACCAGATGACGTTG   | 693 |
| 138 | MNTRGDLQQS  | 505 | ATGAACACCAGAGGAGACTTGCAACAATCG   | 694 |
| 139 | AGSRGDLQTV  | 506 | GCGGGATCAAGAGGAGACCTACAAACGGTG   | 695 |
| 140 | SNMRGDQTYV  | 507 | TCGAACATGAGAGGAGACCAAACCTTACGTC  | 696 |
| 141 | TPSRGDMSNV  | 508 | ACGCCAAGTAGAGGAGACATGTCCAACGTC   | 697 |
| 142 | NSTRGDLHMA  | 509 | AACAGCACAAGAGGAGACCTACACATGGCA   | 698 |
| 143 | ANTRGDL SVS | 510 | GCTAACACCAGAGGAGACCTCTCAGTCAGT   | 699 |
| 144 | TDRRGDTGLL  | 511 | ACCGACCGAAGAGGAGACACCGGCTTACTG   | 700 |
| 145 | DGVRGDRATW  | 512 | GACGGCGTAAGAGGAGACCGTGCAACATGG   | 701 |
| 146 | ENRRGDFNNT  | 384 | GAGAATCGTAGAGGAGACTTTAATAATACG   | 702 |
| 147 | EPSRGDRMVL  | 513 | GAGCCTAGTAGAGGAGACAGGATGGTGTG    | 703 |
| 148 | SGVRGDRIII  | 514 | AGCGGCGTTAGAGGAGACCGGATCATAATC   | 704 |
| 149 | TSMRGDYTHV  | 515 | ACGTCGATGAGAGGAGACTATACGCATGTT   | 705 |

|     |            |     |                                |     |
|-----|------------|-----|--------------------------------|-----|
| 150 | ASSRGDLTVI | 516 | GCGTCGAGTAGAGGAGACTTGACTGTTATT | 706 |
| 151 | ETRRGDLSLQ | 517 | GAGACGCGTAGAGGAGACCTGTGCTTCAG  | 707 |
| 152 | NNARGDQSMH | 518 | AACAACGCAAGAGGAGACCAATCAATGCAC | 708 |
| 153 | TDRRGDSSAL | 519 | ACGGATAGGAGAGGAGACTCTAGTGCGCTT | 709 |
| 154 | SSQRGDLTGV | 520 | TCAAGCCAAAGAGGAGACTTGACAGGTGTG | 710 |
| 155 | QTARGDVMVH | 521 | CAGACGGCTAGAGGAGACGTTATGGTTCAT | 711 |
| 156 | VSQRGDLNAV | 522 | GTGTCGCAGAGAGGAGACCTGAATGCTGTT | 712 |
| 157 | TMTRGDLAAN | 523 | ACGATGACCAGAGGAGACCTCGCTGCTAAC | 713 |
| 158 | NASRGDHSSL | 524 | AATGCTTCGAGAGGAGACCATTCGTCGTTG | 714 |
| 159 | SNTRGDMGLT | 525 | TCAAACACCAGAGGAGACATGGGCCTCACG | 715 |
| 160 | MYSRGDTHSL | 526 | ATGTATTCGAGAGGAGACACTCATAGTCTG | 716 |
| 161 | ADQRGDRAPL | 527 | GCGGATCAGAGAGGAGACAGGGCTCCGCTT | 717 |
| 162 | QERRGDYTSM | 382 | CAAGAAAGGAGAGGAGACTACACCTCTATG | 718 |
| 163 | SQNRGDLANT | 528 | TCCCAAACAGAGGAGACCTAGCCAACACG  | 719 |
| 164 | EKLRGDLHST | 389 | GAGAAGTTGAGAGGAGACCTTCATTGCTG  | 720 |
| 165 | QGGRGDLSTL | 404 | CAAGGCGGAAGAGGAGACCTGAGTACACTG | 721 |
| 166 | SNRRGDTEMQ | 529 | TCGAATAGGAGAGGAGACACTGAGATGCAG | 722 |
| 167 | SGSRGDVSAL | 530 | TCGGGTAGTAGAGGAGACGTGAGTGCTTTG | 723 |
| 168 | ESTRGDRGTL | 531 | GAGTCGACGAGAGGAGACCGTGGTACGCTG | 724 |
| 169 | SAVRGDAALH | 532 | TCGGCGGTGAGAGGAGACGCGGCGCTTCAT | 725 |
| 170 | LSSRGDVNRL | 533 | CTGAGCAGCAGAGGAGACGTAAACCGCCTT | 726 |
| 171 | EQRRGDIQTI | 534 | GAGCAGAGGAGAGGAGACATTCAGACTATT | 727 |
| 172 | AAYRGDAHVL | 485 | GCTGCGTATAGAGGAGACGCGCATGTTCTT | 728 |
| 173 | SPVRGDHGAL | 535 | TCGCCGGTGAGAGGAGACCATGGGGCTTTG | 729 |
| 174 | DSRRGDYANL | 409 | GATTCGAGGAGAGGAGACTATGCGAATCTG | 730 |
| 175 | LSRRGDYQEL | 536 | TTGAGTCGGAGAGGAGACTATCAGGAGTTG | 731 |
| 176 | SYARGDVHSI | 420 | TCCTACGCAAGAGGAGACGTCCACTCCATC | 732 |
| 177 | DSRRGDASYH | 537 | GACTCACGCAGAGGAGACGCGTCGTACCAC | 733 |
| 178 | NHQRGDLSSS | 538 | AACCACCAAAGAGGAGACCTGAGCTCGAGT | 734 |
| 179 | SSTRGDHLSL | 451 | TCCTCAACGAGAGGAGACCACCTGTCTTTG | 735 |
| 180 | SWNRGDISGL | 539 | TCGTGGAACAGAGGAGACATATCTGGCCTT | 736 |
| 181 | TGTRGDLGTM | 540 | ACGGGTACTAGAGGAGACCTGGGGACGATG | 737 |
| 182 | QQLRGDTHTL | 541 | CAGCAGCTTAGAGGAGACACGCATACTCTT | 738 |
| 183 | AGLRGDRDSL | 542 | GCGGGGTTGAGAGGAGACCGTGATTGCTT  | 739 |
| 184 | QVYRGDRDQL | 543 | CAGGTTTATAGAGGAGACCGGGATCAGTTG | 740 |
| 185 | AGVRGDRVTI | 544 | GCAGGAGTAAGAGGAGACAGAGTGACCATC | 741 |
| 186 | NSARGDLLHS | 545 | AACTCAGCCAGAGGAGACCTTCTGCACTCC | 742 |
| 187 | NHSRGDLTGV | 546 | AACCACAGTAGAGGAGACCTGACAGGCGTT | 743 |
| 188 | NAYRGDTSAF | 547 | AACGCTTACAGAGGAGACACATCCGCGTTC | 744 |

|     |            |     |                                 |     |
|-----|------------|-----|---------------------------------|-----|
| 189 | SANRGDIMHT | 548 | TCGGCTAACAGAGGAGACATAATGCACACG  | 745 |
| 190 | QSARGDLVSY | 549 | CAATCCGCCAGAGGAGACCTCGTCAGTTAC  | 746 |
| 191 | DSRRGDASYH | 537 | GATAGTAGGAGAGGAGACGCTAGTTATCAT  | 747 |
| 192 | SFVRGDVRTL | 550 | TCGTTTGTGAGAGGAGACGTTTCGTACGCTG | 748 |
| 193 | SDMRGDRSVY | 551 | AGCGACATGAGAGGAGACCGATCTGTGTAC  | 749 |
| 194 | SGIRGDRYPI | 552 | TCCGGAATCAGAGGAGACCGCTACCCAATA  | 750 |
| 195 | SNVRGDMAHS | 553 | TCCAACGTCAGAGGAGACATGGCTCACAGC  | 751 |
| 196 | TEVRGDRSQV | 446 | ACTGAGGTTAGAGGAGACCGTTTCGCAGGTT | 752 |
| 197 | ASSRGDLQQT | 554 | GCGAGCAGCAGAGGAGACCTGCAACAAACG  | 753 |
| 198 | QSARGDLVSY | 549 | CAGTCTGCTAGAGGAGACCTTGTGTCTTAT  | 754 |
| 199 | DASRGDRTSL | 445 | GACGCATCTAGAGGAGACCGAACATCCCTC  | 755 |
| 200 | NSVRGDLNQT | 555 | AACAGTGTCAGAGGAGACCTCAACCAAAC   | 756 |

Table: Top Ranking Skeletal Muscle Specific n-mer inserts and/or RGD Motifs

| Top Ranking Skeletal Muscle Specific n-mer inserts and/or RGD Motifs |             |             |     |             |     |
|----------------------------------------------------------------------|-------------|-------------|-----|-------------|-----|
| Variant                                                              | SEQ ID NO.: |             |     |             |     |
| TTGRGDQARL                                                           | 757         | ENRRGDAQLL  | 791 | QHYRGDLSTT  | 833 |
| TTLRGDHSQL                                                           | 758         | TVYRGDREQH  | 792 | METRGRDQQQL | 834 |
| TNLRGDYQLT                                                           | 759         | LDRRGDLLNS  | 793 | EALRGDKVVL  | 835 |
| STSRGDYVTH                                                           | 760         | ESVRGDLRST  | 794 | NAVRGDQSYV  | 836 |
| MPARGDLVQT                                                           | 761         | DKYRGDQRGL  | 795 | SPLRGDNRAL  | 837 |
| SGVRGDR LAV                                                          | 762         | TDRRGDAHIL  | 796 | ENRRGDTHAL  | 838 |
| SNERGDQHRL                                                           | 763         | SQTRGDTSLM  | 797 | NSQRGDLTGT  | 839 |
| STLRGDYAVS                                                           | 764         | DATRGDTRML  | 798 | QYAPSQGTV   | 840 |
| SNARGDIMTT                                                           | 765         | SMQRGDLTHA  | 799 | DNSRGDHSRL  | 841 |
| EASRGDLQMR                                                           | 766         | SGTRGDTLIL  | 800 | NTHRGDYASI  | 842 |
| TYARGDLSGH                                                           | 767         | ANTRGDHQN F | 801 | HTTRGDQQPL  | 843 |
| DSSRGDRSSV                                                           | 768         | TSQRGDHTQL  | 802 | EWRRGDQTAL  | 844 |
| SQSRGDLVPL                                                           | 769         | SHWRGDTSSL  | 803 | TAVRGDQTTV  | 845 |
| SATRGDYSIL                                                           | 770         | ISGRGDRQEL  | 804 | EYLRGDRISQ  | 846 |
| QATRGDLQHT                                                           | 771         | SVLRGDIGTI  | 805 | SRYRGDQETI  | 847 |
| HDDRGRDRML                                                           | 772         | QDSRGDLSGK  | 806 | ASLRGDTVTL  | 848 |
| SSHRGDLLAT                                                           | 773         | KPIRGDGAEL  | 807 | VGVRGD SHVL | 849 |
| SYQRGDQHNM                                                           | 774         | TYMRGDLQHT  | 808 | TNSRGDHTSL  | 850 |
| SHQRGDYN TL                                                          | 775         | TELRGDLSST  | 809 | VDQRGDQRAL  | 851 |
| MDQRGDMSV V                                                          | 776         | SAMRGDHAVL  | 810 | ASVRGDRLDV  | 852 |
| QIARGDLHVP                                                           | 777         | SWVRGDASTL  | 811 | TSYRGDLSSN  | 853 |
| TYTRGDQHMM                                                           | 778         | QYARGDPHAL  | 812 | TSDRGDSSNR  | 854 |
| DASRGDR TSL                                                          | 779         | STNRGDIVVH  | 813 | SYMRGDLDMR  | 855 |
| SDRRGDLHPQ                                                           | 780         | VNIRGDLHQT  | 814 | TDVRGDRSMI  | 856 |
| SATRGDHSTL                                                           | 781         | REYRGDSSIL  | 815 | SYVRGDSSTV  | 857 |
| AYQRGDFTSV                                                           | 782         | STLRGDGSMM  | 816 | QVYRGDRESV  | 858 |
| TGNRGDLQSV                                                           | 783         | SSQRGDITAH  | 817 | TDRRGDAVSF  | 859 |
| AWRRGDLENA                                                           | 784         | TEFRGDRSTV  | 818 | TNSRGDLHTT  | 860 |
| NAARGDHGLL                                                           | 785         | TSTRGDYVTV  | 819 | EAIRGDFKTL  | 861 |
| SYSRGDHQAL                                                           | 786         | NHLRGDVSAI  | 820 | TGIRGDTAGL  | 862 |
| DGVRGDR TVV                                                          | 787         | LERRGDLVST  | 821 | TSSRGDVQSL  | 863 |
| SSERGDKMST                                                           | 788         | NEYRGDHSHP  | 822 | SSRRGDTNEL  | 864 |
| TSMRGDYTHV                                                           | 789         | NWQRGDTSTL  | 823 | SNARGDHTPL  | 865 |
| SGVRGDR III                                                          | 790         | TDSRGDRTVL  | 824 | SELRGDRSPY  | 866 |
|                                                                      |             | YAVRGDQHTI  | 825 | VHARGDTSIM  | 867 |
|                                                                      |             | FGVRGDTRDL  | 826 | SALRGDHGTL  | 868 |
|                                                                      |             | SSARGDTGML  | 827 | QTGRGDIAGM  | 869 |
|                                                                      |             | SAMRGDQMGQ  | 828 | TVARGDQIRL  | 870 |
|                                                                      |             | SHTRGDLVTT  | 829 | TNSRGDSVAL  | 871 |
|                                                                      |             | VTTRGDTAIL  | 830 | QNARGDSAVI  | 872 |
|                                                                      |             | EPTRGDRHSI  | 831 | HYLRGDTTTL  | 873 |
|                                                                      |             | GSQRGDRLVL  | 832 | ANARGDLQNT  | 874 |

|             |     |
|-------------|-----|
| QNIRGDIRTV  | 875 |
| AALRGDHGQM  | 876 |
| TTDRGDL SAR | 877 |
| LNVRGDLVHT  | 878 |
| YSTRGDHTVL  | 879 |
| TNARGDLTVS  | 880 |
| TNARGDY TAL | 881 |
| AEGRGDTRGS  | 882 |
| TNIRGDQSSV  | 883 |
| VDRRGDTMTM  | 884 |
| QIGRGDVTPL  | 885 |
| MQARGDLLGV  | 886 |
| NTQRGDYASV  | 887 |
| SDTRGDRTQ   | 888 |
| SAIRGDYRDV  | 889 |
| VGVRGDVHPL  | 890 |
| TVLRGDAERL  | 891 |
| ASTRGDL SGQ | 892 |
| DSRRGDGQNL  | 893 |
| IDKRGDIQGL  | 894 |
| TVTRGDLLVN  | 895 |
| DSTRGDLRLGL | 896 |
| AANRGDYASL  | 897 |
| RDYRGDTSSL  | 898 |
| SNIRGDASLH  | 899 |
| NLQRGDR TDF | 900 |
| NASRGDHSSM  | 901 |
| TSMRGDYTHV  | 902 |
| TSRRGDLNDV  | 903 |
| QSHRGDVQSL  | 904 |
| STTRGDYQSV  | 905 |
| YSTRGDHAML  | 906 |
| TSARGDLIHT  | 907 |
| ATARGDRESF  | 908 |
| ASTRGDYSTM  | 909 |
| SNNRGDVQRL  | 910 |
| ERLRGDQHII  | 911 |
| SSIRGDHATL  | 912 |
| NPTRGDIATV  | 913 |
| MSTRGDL SVN | 914 |
| MDRRGDITYA  | 915 |
| NATRGDLMSA  | 916 |
| SDHRGDRTYQ  | 917 |

|             |     |
|-------------|-----|
| NWVRGDLSTA  | 918 |
| SSHRGDL SAT | 919 |
| NATRGDYREI  | 920 |
| DTLRGDLQTA  | 921 |
| IAVRGDLTPV  | 922 |
| TLLRGDLAHS  | 923 |
| MNTRGDHTVL  | 924 |
| TGHRGDI AVY | 925 |
| TSNRGDQMSN  | 926 |
| SATRGDHTAI  | 927 |
| SSVRGDREHA  | 928 |
| TNSRGDLV VH | 929 |
| ETRRGDLSSQ  | 930 |
| TYARGDITLT  | 931 |
| SASRGDTLHL  | 932 |
| DNQRGDLSSR  | 933 |
| TQSRGDLRDA  | 934 |
| ENRRGDLTNA  | 935 |
| TSIRGDISSV  | 936 |
| ISARGDTMTL  | 937 |
| SRVRGDQDHW  | 938 |
| SGNRGDIASY  | 939 |
| NEQRGDRYNH  | 940 |
| EALRGDV RAL | 941 |
| VDRRGDYTQM  | 942 |
| NNVRGDYTSV  | 943 |
| QTYRGDASVL  | 944 |
| SNYRGD HGSL | 945 |
| TSSRGDLREV  | 946 |
| QYTRGDVSNL  | 947 |
| SDLRGDV RER | 948 |
| QNARGD TTAL | 949 |
| NSLRGDRDTM  | 950 |
| SNSRGDQAQV  | 951 |
| EATRGDR STL | 952 |
| VPQRGDITNL  | 953 |
| DKIRGDLRSI  | 954 |
| SAARGDY SIV | 955 |
| VGLRGDLIGQ  | 956 |
| QAGRGDLTVI  | 957 |
| SKTRGDFQEA  | 958 |
| GTARGDY SML | 959 |
| ISTRGDLVTI  | 960 |

|             |      |
|-------------|------|
| TGTRGDREYM  | 961  |
| MTTRGDL SIT | 962  |
| STSRGDMTSV  | 963  |
| GPARGDFTSS  | 964  |
| VDRRGDLGHV  | 965  |
| NSGRGDMSQL  | 966  |
| SQGRGDLERY  | 967  |
| VFGRGDQYTL  | 968  |
| IVRGD LHLQ  | 969  |
| SYSRGDTSRL  | 970  |
| MSSRGDRATI  | 971  |
| MSSRGDLTRV  | 972  |
| QHTRGDFGNL  | 973  |
| ETRRGDLSSQ  | 974  |
| TGTRGDLRAI  | 975  |
| QMARGDLSTA  | 976  |
| LVERGDRIMN  | 977  |
| GLVRGDLNAR  | 978  |
| NNPRGDLSSV  | 979  |
| HEKRGDLSNS  | 980  |
| SQARGDLHQT  | 981  |
| FTGRGDLRTA  | 982  |
| QSIRGDTGHL  | 983  |
| SAGRGDMSVL  | 984  |
| SERRGDEHKW  | 985  |
| TTGRGDLVSL  | 986  |
| LGLRGDQIGQ  | 987  |
| VALRGDLTHQ  | 988  |
| MIGRGDQAPL  | 989  |
| FQARGDH STL | 990  |
| NSVRGDHSSL  | 991  |
| WPNRGDLQST  | 992  |
| SASRGDYQTL  | 993  |
| TMQRGDLHTI  | 994  |
| TALRGDLAGT  | 995  |
| VNARGDFSQT  | 996  |
| GWIRGDR TSM | 997  |
| TFSRGDHTVL  | 998  |
| SGARGDLVGV  | 999  |
| NNVRGDYTSV  | 1000 |
| QTARGDHVVL  | 1001 |
| TLIRGDLTRS  | 1002 |
| VTTRGDRLEL  | 1003 |

Table: Top Ranking Skeletal Muscle

Specific n-mer inserts and/or RGD Motifs

| <b>Top Ranking Skeletal Muscle Specific n-mer inserts and/or RGD Motifs</b> |             |
|-----------------------------------------------------------------------------|-------------|
| Variant                                                                     | SEQ ID NO.: |
| EASRGDLQMR                                                                  | 1004        |
| TYARGDLSGH                                                                  | 1005        |
| DSSRGDRSSV                                                                  | 1006        |
| SQSRGDLVPL                                                                  | 1007        |
| MPARGDLVQT                                                                  | 1008        |
| SGVRGDR LAV                                                                 | 1009        |
| SNERGDQHRL                                                                  | 1010        |
| STLRGDYAVS                                                                  | 1011        |
| SNARGDIMTT                                                                  | 1012        |
| TTGRGDQARL                                                                  | 1013        |
| TTLRGDHSQL                                                                  | 1014        |
| TNLRGDYQLT                                                                  | 1015        |
| STSRGDYVTH                                                                  | 1016        |
| SATRGDYSIL                                                                  | 1017        |
| QATRGDLQHT                                                                  | 1018        |
| HDDRGRDRML                                                                  | 1019        |
| SSHRGDL LAT                                                                 | 1020        |
| SYQRGDQHNM                                                                  | 1021        |
| SHQRGDYN TL                                                                 | 1022        |
| MDQRGDMSV V                                                                 | 1023        |
| QIARGDLHVP                                                                  | 1024        |
| TYTRGDQHMM                                                                  | 1025        |
| DASRGDR TSL                                                                 | 1026        |
| SDRRGDLHPQ                                                                  | 1027        |
| SATRGDHSTL                                                                  | 1028        |
| AYQRGDFTSV                                                                  | 1029        |
| TGNRGDLQSV                                                                  | 1030        |
| AWRRGDLENA                                                                  | 1031        |
| NAARGDHGLL                                                                  | 1032        |
| SYSRGDHQAL                                                                  | 1033        |
| DGVRGDR TVV                                                                 | 1034        |

|             |      |
|-------------|------|
| SSERGDKMST  | 1035 |
| TSMRGDYTHV  | 1036 |
| SGVRGDR III | 1037 |
| ENRRGDAQLL  | 1038 |
| TVYRGDREQH  | 1039 |
| LDRRGDLLNS  | 1040 |
| ESVRGDLRST  | 1041 |
| DKYRGDQRGL  | 1042 |
| TDRRGDAHIL  | 1043 |
| SQTRGDTSLM  | 1044 |
| DATRGDTRML  | 1045 |
| SMQRGDLTHA  | 1046 |
| SGTRGDTLIL  | 1047 |
| ANTRGDHQNF  | 1048 |
| TSQRGDHTQL  | 1049 |
| SHWRGDTSSL  | 1050 |
| ISGRGDRQEL  | 1051 |
| SVLRGDIGTI  | 1052 |
| QDSRGDLSGK  | 1053 |
| KPIRGDGAEL  | 1054 |
| TYMRGDLQHT  | 1055 |
| TELRGDL SST | 1056 |
| SAMRGDHA VL | 1057 |
| SWVRGDASTL  | 1058 |
| QYARGDPHAL  | 1059 |
| STNRGDIVVH  | 1060 |
| VNIRGDLHQT  | 1061 |
| REYRGDSSIL  | 1062 |
| STLRGDGSMM  | 1063 |
| SSQRGDITAH  | 1064 |
| TEFRGDRSTV  | 1065 |
| TSTRGDYVTV  | 1066 |
| NHLRGDVSAI  | 1067 |
| LERRGDLVST  | 1068 |
| NEYRGDHSHP  | 1069 |
| NWQRGDTSTL  | 1070 |
| TDSRGDR TVL | 1071 |
| YAVRGDQHTI  | 1072 |
| FGVRGDTRDL  | 1073 |

|              |      |
|--------------|------|
| SSARGDTGML   | 1074 |
| SAMRGDQMGQ   | 1075 |
| SHTRGDLVTT   | 1076 |
| VTTRGD TAIL  | 1077 |
| EPTRGDRHSI   | 1078 |
| GSQRGDR LVL  | 1079 |
| QHYRGDLSTT   | 1080 |
| METRGRD RQL  | 1081 |
| EALRGDKVVL   | 1082 |
| NAVRGDQSYV   | 1083 |
| SPLRGDNRAL   | 1084 |
| ENRRGDTHAL   | 1085 |
| NSQRGDLTGT   | 1086 |
| DNSRGDHSRL   | 1087 |
| NTHRGDYASI   | 1088 |
| HTTRGDQ QPL  | 1089 |
| EWRRGDQTAL   | 1090 |
| TAVRGDQTTV   | 1091 |
| EYLRGDRISQ   | 1092 |
| SRYRGDQETI   | 1093 |
| ASLRGDTVTL   | 1094 |
| VGVRGD SHVL  | 1095 |
| TNSRGDHTSL   | 1096 |
| VDQRGDQ RAL  | 1097 |
| ASVRGDR L DV | 1098 |
| TSYRGDLSSN   | 1099 |
| TSDRGDSSNR   | 1100 |
| SYMRGDL DMR  | 1101 |
| TDVRGDRSMI   | 1102 |
| SYVRGDSSTV   | 1103 |
| QVYRGDRESV   | 1104 |
| TDRRGDAVSF   | 1105 |
| TNSRGDLHTT   | 1106 |
| EAIRGDFKTL   | 1107 |
| TGIRGDTAGL   | 1108 |
| TSSRGDVQSL   | 1109 |
| SSRRGDTNEL   | 1110 |
| SNARGDHTPL   | 1111 |
| SELRGDRSPY   | 1112 |

|             |      |
|-------------|------|
| VHARGDTSIM  | 1113 |
| SALRGDHGTL  | 1114 |
| QTGRGDIAGM  | 1115 |
| TVARGDQIRL  | 1116 |
| TNSRGDSVAL  | 1117 |
| QNARGDSAVI  | 1118 |
| HYLRGDTTTL  | 1119 |
| ANARGDLQNT  | 1120 |
| QNIRGDIRTV  | 1121 |
| AALRGDHGQM  | 1122 |
| TTDRGDL SAR | 1123 |
| LNVRGDLVHT  | 1124 |
| YSTRGDHTVL  | 1125 |
| TNARGDLTVS  | 1126 |
| TNARGDYTAL  | 1127 |
| AEGRGDTRGS  | 1128 |
| TNIRGDQSSV  | 1129 |
| VDRRGDTMTM  | 1130 |
| QIGRGDVTPL  | 1131 |
| MQARGDLLGV  | 1132 |
| NTQRGDYASV  | 1133 |
| SDTRGDRLTQ  | 1134 |
| SAIRGDYRDV  | 1135 |
| VGVRGDVHPL  | 1136 |
| TVLRGDAERL  | 1137 |
| ASTRGDL SGQ | 1138 |
| DSRRGDGQNL  | 1139 |
| IDKRGDIQGL  | 1140 |
| TVTRGDL LVN | 1141 |
| DSTRGDLRGL  | 1142 |
| AANRGDYASL  | 1143 |
| RDYRGDTSSL  | 1144 |
| SNIRGDASLH  | 1145 |
| NLQRGDR TDF | 1146 |
| NASRGDHSSM  | 1147 |
| TSMRGDYTHV  | 1148 |
| TSRRGDLNDV  | 1149 |
| QSHRGDVQSL  | 1150 |
| STTRGDYQSV  | 1151 |
| YSTRGDHAML  | 1152 |
| TSARGDLIHT  | 1153 |
| ATARGDRESF  | 1154 |
| ASTRGDYSTM  | 1155 |

|             |      |
|-------------|------|
| SNNRGDVQRL  | 1156 |
| ERLRGDQHII  | 1157 |
| SSIRGDHATL  | 1158 |
| NPTRGDIATV  | 1159 |
| MSTRGDLSVN  | 1160 |
| MDRRGDITYA  | 1161 |
| NATRGDLMSA  | 1162 |
| SDHRGDRTYQ  | 1163 |
| NWVRGDLSTA  | 1164 |
| SSHRGDL SAT | 1165 |
| NATRGDYREI  | 1166 |
| DTLRGDLQTA  | 1167 |
| IAVRGDLTPV  | 1168 |
| TLLRGDLAHS  | 1169 |
| MNTRGDHTVL  | 1170 |
| TGHRGDI AVY | 1171 |
| TSNRGDQMSN  | 1172 |
| SATRGDHTAI  | 1173 |
| SSVRGDREHA  | 1174 |
| TNSRGDLV VH | 1175 |
| ETRRGDLSSQ  | 1176 |
| TYARGDITLT  | 1177 |
| SASRGDTLHL  | 1178 |
| DNQRGDLSSR  | 1179 |
| TQSRGDLRDA  | 1180 |
| ENRRGDLTNA  | 1181 |
| TSIRGDISSV  | 1182 |
| ISARGDTMTL  | 1183 |
| SRVRGDQDHW  | 1184 |
| SGNRGDIASY  | 1185 |
| NEQRGDRYNH  | 1186 |
| EALRGDV RAL | 1187 |
| VDRRGDYTQM  | 1188 |
| NNVRGDYTSV  | 1189 |
| QTYRGDASVL  | 1190 |
| SNYRGDHGSL  | 1191 |
| TSSRGDLREV  | 1192 |
| QYTRGDVSNL  | 1193 |
| SDLRGDV RER | 1194 |
| QNARGDTTAL  | 1195 |
| NSLRGDRDTM  | 1196 |
| SNSRGDQAQV  | 1197 |
| EATRGDRSTL  | 1198 |

|             |      |
|-------------|------|
| VPQRGDITNL  | 1199 |
| DKIRGDLRSI  | 1200 |
| SAARGDY SIV | 1201 |
| VGLRGDLIGQ  | 1202 |
| QAGRGDLTVI  | 1203 |
| SKTRGDFQEA  | 1204 |
| GTARGDY SML | 1205 |
| ISTRGDLVTI  | 1206 |
| TGTRGDREYM  | 1207 |
| MTTRGDL SIT | 1208 |
| STSRGDMTSV  | 1209 |
| GPARGDFTSS  | 1210 |
| VDRRGDLGHV  | 1211 |
| NSGRGDMSQL  | 1212 |
| SQGRGDLERY  | 1213 |
| VFGRGDQYTL  | 1214 |
| IVRGDLHLQ   | 1215 |
| SYSRGDTSRL  | 1216 |
| MSSRGDRATI  | 1217 |
| MSSRGDLTRV  | 1218 |
| QHTRGDFGNL  | 1219 |
| ETRRGDLSSQ  | 1220 |
| TGTRGDLRAI  | 1221 |
| QMARGDLSTA  | 1222 |
| LVERGDRIMN  | 1223 |
| GLVRGDLNAR  | 1224 |
| NNPRGDLSSV  | 1225 |
| HEKRGDLSNS  | 1226 |
| SQARGDLHQT  | 1227 |
| FTGRGDLRTA  | 1228 |
| QSIRGDTGHL  | 1229 |
| SAGRGDMSVL  | 1230 |
| SERRGDEHKW  | 1231 |
| TTGRGDLVSL  | 1232 |
| LGLRGDQIGQ  | 1233 |
| VALRGDLTHQ  | 1234 |
| MIGRGDQAPL  | 1235 |
| FQARGDHSTL  | 1236 |
| NSVRGDHSSL  | 1237 |
| WPNRGDLQST  | 1238 |
| SASRGDYQTL  | 1239 |
| TMQRGDLHTI  | 1240 |
| TALRGDLAGT  | 1241 |

|            |      |
|------------|------|
| VNARGDFSQT | 1242 |
| GWIRGDRTSM | 1243 |
| TFSRGDHTVL | 1244 |

|            |      |
|------------|------|
| SGARGDLVGV | 1245 |
| NNVRGDYTSV | 1246 |
| QTARGDHVVL | 1247 |

|            |      |
|------------|------|
| TLIRGDLTRS | 1248 |
| VTTRGDRLEL | 1249 |

### Promoter

The invention may contain a muscle specific promoter or another promoter. The promoter may be linked to the nucleic acid sequence so that the transcription preferably occurs within myocytes. Promoter regions enable the host cells to replicate the AAV delivered nucleic acid only in those cell types and tissues or organs in which the desired protein should be created. Here, the muscle specific promoter is included because it is principally desired that the proteins only be translated in myocytes. Specificity of the cell type into which the nucleic acid is delivered and thus the proteins translated is desired because of the adverse effects that may ensue from delivering the nucleic acid and having it translated in cells in which that nucleic acid and thus protein is not needed.

In some embodiments, the muscle specific promoter yields increased muscle cell potency, muscle cell specificity, reduced immunogenicity, or any combination thereof. As used herein the terms “muscle-specific”, “muscle cell specificity”, “muscle cell potency,” “myocyte specific” and the like, refer to the increased specificity, selectivity, or potency, of the muscle-specific targeting moieties and compositions incorporating said muscle-specific targeting moieties of the present invention for myocytes relative to non-muscle cells. In some embodiments, the cell specificity, or selectivity, or potency, or a combination thereof of a muscle-specific targeting moiety or composition incorporating a muscle-specific targeting moiety described herein is at least 2 to at least 500 times more specific, selective, and/or potent for/in a muscle cell relative to a non-muscle cell.

In some embodiments, the myocyte-selective promoter utilized is MHCK7. MHCK7 is a 770 base pair length promoter that is small enough to be included in an AAV vector. MHCK7 directs expression in fast and slow skeletal and cardiac muscle, with low expression in the liver, lung, and spleen. It is less active in smooth muscle. The MHCK7 promoter is associated with high levels of expression in skeletal muscles, including the diaphragm, and includes an enhancer to especially drive expression in the heart, whereas expression in off-target tissues is minimal.

In some embodiments, the promoters described herein are inserted into an AAV protein (e.g., an AAV capsid protein) that has reduced specificity (or no detectable, measurable, or clinically relevant interaction) for one or more non-muscle cell types. Exemplary non-muscle cell types include, but are not limited to, liver, kidney, lung, heart,

spleen, central or peripheral nervous system cells, bone, immune, stomach, intestine, eye, skin cells and the like. In some embodiments, the non-muscle cells are liver cells.

The term “operably linked” refers to the association of two or more nucleic acid molecules on a single nucleic acid fragment so that the function of one is affected by the other.

Further exemplary tissue specific promoters include U6 promoter sequence, MHCK7 promoter sequence, CK6 promoter sequence, tMCK promoter sequence, CK5 promoter sequence, MCK promoter sequence, HAS promoter sequence, MPZ promoter sequence, desmin promoter sequence, APOA2 promoter sequence, hAAT promoter sequence, INS promoter sequence, IRS2 promoter sequence, MYH6 promoter sequence, MYL2 promoter sequence, TNNI3 promoter sequence, SYN1 promoter sequence, GFAP promoter sequence, NES promoter sequence, MBP promoter sequence, or TH promoter sequence.

Muscle specific promoters are described in International Patent Application Publications WO 2020/006458 and WO 2021/126880, the contents of each of which are incorporated by reference herein.

Further muscle specific promoters are described in U.S. Patent No. 9,133,482; U.S. Patent No. 10,105,453; U.S. Patent No. 10,301,367; U.S. Patent Publication No. 2020-0360534; PCT International Patent Publication Nos. WO 2020/006458; WO 2021/035120; WO 2021/053124; and WO 2021/077000, the contents of each of which are incorporated by reference herein.

It may be convenient to use an RNA polymerase II or III promoter; these are known to the person skilled in the art and reviewed in e.g. Kornberg 1999. However, transcripts from an RNA II polymerase often have complex transcription terminators and transcripts are polyadenylated; this may hamper with the requirements of the miRNA strand which because both its 5' and 3' ends need to be precisely defined in order to achieve the required secondary structure to produce a functional molecule. These drawbacks can however be circumvented. In case an RNA polymerase II or III promoter is used, the polynucleotide encoding the miRNA strand may also encode self-processing ribozymes and may be operably linked to an RNA polymerase II or III promoter; as such the polynucleotide encodes a pre- miRNA strand comprising the miRNA strand and self-processing ribozymes, wherein, when transcribed, the

miRNA strand is released by the self-processing ribozymes from the pre- miRNA strand de transcript.

Preferably, in a composition according to the present invention the AAV vector is comprised of an RNA polymerase II promoter or III promoter, and encodes a pre- miRNA strand comprising the miRNA strand and self-processing ribozymes, wherein, when transcribed, the miRNA strand is released by the self-processing ribozymes from the pre-miRNA strand transcript. Conveniently, multiple pre-miRNA strands and multiple self-processing ribozymes may be encoded by a single polynucleotide, operably linked to one or more RNA polymerase II promoters.

RNA polymerase II or III promoters that are inducible and/or tissue-specific have been previously described. RNA polymerase promoters are known in the art and further described in U.S. Patent Publication 11,149,288, the contents of which is incorporated by reference herein.

#### Capsid Protein

The capsid protein is the shell or coating of the virus that enables its delivery into the host. Without the protein, the nucleic acids would be destroyed by the host without entering into the host cells and beginning transcription and translation. The capsid protein may be in the natural conformation of a naturally occurring AAV, or it may be modified.

In certain example embodiments, the AAV capsid protein is an engineered AAV capsid protein having reduced or eliminated uptake in a non-muscle cell as compared to a corresponding wild-type AAV capsid polypeptide.

In some embodiments, the engineered AAV capsid encoding polynucleotide can be included in a polynucleotide that is configured to be an AAV genome donor in an AAV vector system that can be used to generate engineered AAV particles described elsewhere herein. In some embodiments, the engineered AAV capsid encoding polynucleotide can be operably coupled to a poly adenylation tail. In some embodiments, the poly adenylation tail can be an SV40 poly adenylation tail. In some embodiments, the AAV capsid encoding polynucleotide can be operably coupled to a promoter. In some embodiments, the promoter can be a tissue specific promoter. In some embodiments, the tissue specific promoter is specific for muscle (e.g., cardiac, skeletal, and/or smooth muscle), neurons and supporting

cells (e.g., astrocytes, glial cells, Schwann cells, etc.), fat, spleen, liver, kidney, immune cells, spinal fluid cells, synovial fluid cells, skin cells, cartilage, tendons, connective tissue, bone, pancreas, adrenal gland, blood cell, bone marrow cells, placenta, endothelial cells, and combinations thereof. In some embodiments, the promoter can be a constitutive promoter.

5     Suitable tissue specific promoters and constitutive promoters are discussed elsewhere herein and are generally known in the art and can be commercially available. Suitable muscle specific promoters include, but are not limited to CK8, MHCK7, Myoglobin promoter (Mb), Desmin promoter, muscle creatine kinase promoter (MCK) and variants thereof, and SPc5-12 synthetic promoter.

10             Described herein are various embodiments of engineered viral capsids, such as adeno-associated virus (AAV) capsids, that can be engineered to confer cell-specific tropism, such as muscle specific tropism, to an engineered viral particle. Engineered viral capsids can be lentiviral, retroviral, adenoviral, or AAV capsids. The engineered capsids can be included in an engineered virus particle (e.g., an engineered lentiviral, retroviral, adenoviral, or AAV  
15     virus particle), and can confer cell-specific tropism, reduced immunogenicity, or both to the engineered viral particle. The engineered viral capsids described herein can include one or more engineered viral capsid proteins described herein. The engineered viral capsids described herein can include one or more engineered viral capsid proteins described herein that can contain a muscle-specific targeting moiety containing or composed of an n-mer  
20     motif described elsewhere herein.

           The engineered viral capsid and/or capsid proteins can be encoded by one or more engineered viral capsid polynucleotides. In some embodiments, the engineered viral capsid polynucleotide is an engineered AAV capsid polynucleotide, engineered lentiviral capsid polynucleotide, engineered retroviral capsid polynucleotide, or engineered adenovirus capsid  
25     polynucleotide. In some embodiments, an engineered viral capsid polynucleotide (e.g., an engineered AAV capsid polynucleotide, engineered lentiviral capsid polynucleotide, engineered retroviral capsid polynucleotide, or engineered adenovirus capsid polynucleotide) can include a 3' polyadenylation signal. The polyadenylation signal can be an SV40 polyadenylation signal.

30             The engineered viral capsids can be variants of wild-type viral capsid. For example, in some embodiments, the engineered AAV capsids can be variants of wild-type AAV

capsids. In some embodiments, the wild-type AAV capsids can be composed of VP1, VP2, VP3 capsid proteins or a combination thereof. In other words, the engineered AAV capsids can include one or more variants of a wild-type VP1, wild-type VP2, and/or wild-type VP3 capsid proteins. In some embodiments, the serotype of the reference wild-type AAV capsid  
5 can be AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, AAV-9 or any combination thereof. In some embodiments, the serotype of the wild-type AAV capsid can be AAV-9. The engineered AAV capsids can have a different tropism than that of the reference wild-type AAV capsid.

The engineered viral capsid can contain 1-60 engineered capsid proteins. In some  
10 embodiments, the engineered viral capsids can contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 engineered capsid proteins. In some embodiments, the engineered viral capsid can contain 0-59 wild-type viral capsid proteins. In some embodiments, the engineered viral capsid can  
15 contain 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, or 59 wild-type viral capsid proteins.

In some embodiments, the engineered AAV capsid can contain 1-60 engineered capsid proteins. In some embodiments, the engineered AAV capsids can contain 1, 2, 3, 4, 5,  
20 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 engineered capsid proteins. In some embodiments, the engineered AAV capsid can contain 0-59 wild-type AAV capsid proteins. In some embodiments, the engineered AAV capsid can contain 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,  
25 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, or 59 wild-type AAV capsid proteins.

In some embodiments, the engineered viral capsid protein can have an n-mer amino acid motif, where n can be at least 3 amino acids. In some embodiments, n can be 3, 4, 5, 6,  
30 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids. In some embodiments, an engineered AAV capsid can have a 6-mer or 7-mer amino acid motif. In some embodiments, the n-mer amino

acid motif can be inserted between two amino acids in the wild-type viral protein (VP) (or capsid protein). In some embodiments, the *n*-mer motif can be inserted between two amino acids in a variable amino acid region in a viral capsid protein.

In some embodiments, the *n*-mer motif can be inserted between two amino acids in a variable amino acid region in an AAV capsid protein. The core of each wild-type AAV viral protein contains an eight-stranded beta-barrel motif (betaB to betaI) and an alpha-helix (alphaA) that are conserved in autonomous parvovirus capsids (see e.g., DiMattia et al. 2012. J. Virol. 86(12):6947-6958). Structural variable regions (VRs) occur in the surface loops that connect the beta-strands, which cluster to produce local variations in the capsid surface.

AAVs have 12 variable regions (also referred to as hypervariable regions) (see e.g., Weitzman and Linden. 2011. "Adeno-Associated Virus Biology." In Snyder, R.O., Moullier, P. (eds.) Totowa, NJ: Humana Press). In some embodiments, one or more *n*-mer motifs can be inserted between two amino acids in one or more of the 12 variable regions in the wild-type AAV capsid proteins. In some embodiments, the one or more *n*-mer motifs can be each inserted between two amino acids in VR-I, VR-II, VR-III, VR-IV, VR-V, VR-VI, VR-VII, VR-VIII, VR-IX, VR-X, VR-XI, VR-XII, or a combination thereof. In some embodiments, the *n*-mer can be inserted between two amino acids in the VR-III of a capsid protein. In some embodiments, the engineered capsid can have an *n*-mer inserted between any two contiguous amino acids between amino acids 262 and 269, between any two contiguous amino acids between amino acids 327 and 332, between any two contiguous amino acids between amino acids 382 and 386, between any two contiguous amino acids between amino acids 452 and 460, between any two contiguous amino acids between amino acids 488 and 505, between any two contiguous amino acids between amino acids 545 and 558, between any two contiguous amino acids between amino acids 581 and 593, between any two contiguous amino acids between amino acids 704 and 714 of an AAV9 viral protein. In some embodiments, the engineered capsid can have an *n*-mer inserted between amino acids 588 and 589 of an AAV9 viral protein. In some embodiments, the engineered capsid can have a 7-mer motif inserted between amino acids 588 and 589 of an AAV9 viral protein. In other embodiments, the motif inserted is a 10-mer motif, with replacement of amino acids 586-88 and an insertion before 589. SEQ ID NO. 1 is a reference AAV9 capsid sequence for at least referencing the insertion sites discussed above. It will be appreciated that *n*-mers can

be inserted in analogous positions in AAV viral proteins of other serotypes. In some embodiments as previously discussed, the *n*-mer(s) can be inserted between any two contiguous amino acids within the AAV viral protein and in some embodiments the insertion is made in a variable region.

5           In some embodiments, the first 1, 2, 3, or 4 amino acids of an *n*-mer motif can replace 1, 2, 3, or 4 amino acids of a polypeptide into which it is inserted and preceding the insertion site. In some embodiments, the amino acids of the *n*-mer motif that replace 1 or more amino acids of the polypeptide into which the *n*-mer motif is inserted come before or immediately before an “RGD” in an *n*-mer motif. For example, in one or more of the 10-mer inserts, the  
10   first three amino acids shown can replace 1-3 amino acids into a polypeptide to which they may be inserted. Using an AAV as another non-limiting example, one or more of the *n*-mer motifs can be inserted into e.g., and AAV9 capsid polypeptide between amino acids 588 and 589 and the insert can replace amino acids 586, 587, and 588 such that the amino acid immediately preceding the *n*-mer motif after insertion is residue 585. It will be appreciated  
15   that this principle can apply in any other insertion context and is not necessarily limited to insertion between residues 588 and 589 of an AAV9 capsid or equivalent position in another AAV capsid. It will further be appreciated that in some embodiments, no amino acids in the polypeptide into which the *n*-mer motif is inserted are replaced by the *n*-mer motif.

          In some embodiments, the AAV capsids or other viral capsids or compositions can be  
20   muscle-specific. In some embodiments, muscle-specificity of the engineered AAV or other viral capsid or other composition is conferred by a muscle specific *n*-mer motif incorporated in the engineered AAV or other viral capsid or other composition described herein. While not intending to be bound by theory, it is believed that the *n*-mer motif confers a 3D structure to or within a domain or region of the engineered AAV capsid or other viral capsid or other  
25   composition such that the interaction of the viral particle or other composition containing the engineered AAV capsid or other viral capsid or other composition described herein has increased or improved interactions (e.g., increased affinity) with a cell surface receptor and/or other molecule on the surface of a muscle cell. In some embodiments, the cell surface receptor is AAV receptor (AAVR). In some embodiments, the cell surface receptor is a  
30   muscle cell specific AAV receptor. In some embodiments, the cell surface receptor or other molecule is a cell surface receptor or other molecule selectively expressed on the surface of a

muscle cell. In some embodiments, the cell surface receptor or molecule is an integrin or dimer thereof. In some embodiments, the cell surface receptor or molecule is an Vb6 integrin heterodimer.

In some embodiments, a muscle specific engineered viral particle or other composition described herein containing the muscle-specific capsid, n-mer motif, or muscle-specific targeting moiety described herein can have an increased uptake, delivery rate, transduction rate, efficiency, amount, or a combination thereof in a muscle cell as compared to other cells types and/or other virus particles (including but not limited to AAVs) and other compositions that do not contain the muscle-specific n-mer motif of the present invention.

First- and second-generation muscle specific AAV capsids were developed using a muscle specific promoter and the resulting capsid libraries were screened in mice and non-human primates as described elsewhere herein and/or in e.g., U.S. Provisional Application Serial Nos. 62/899,453, 62/916,207, 63/018,454, 63/242,008, and No. 63/345,14.

#### Pharmaceutical Composition

Some embodiments of the invention may include any acceptable form of providing the AAV vector to a subject. For example, the AAV vector may be provided to the subject in the form of a composition or formulation comprising the AAV vector. The expression vector of this invention can be formulated and administered to treat a variety of disease states by any means that produces contact of the active ingredient with the agent's site of action in the body of the subject. The compositions, polynucleotides, polypeptides, particles, cells, vector systems and combinations thereof described herein can be contained in a formulation, such as a pharmaceutical formulation. In some embodiments, the formulations can be used to generate polypeptides and other particles that include one or more muscle-specific targeting moieties described herein. In some embodiments, the formulations can be delivered to a subject in need thereof. In some embodiments, component(s) of the engineered AAV capsid system, engineered cells, engineered AAV capsid particles, and/or combinations thereof described herein can be included in a formulation that can be delivered to a subject or a cell. In some embodiments, the formulation is a pharmaceutical formulation. One or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein can be provided to a subject in need thereof or a cell alone or as an active ingredient, such as in a

pharmaceutical formulation. As such, also described herein are pharmaceutical formulations containing an amount of one or more of the polypeptides, polynucleotides, vectors, cells, or combinations thereof described herein. In some embodiments, the pharmaceutical formulation can contain an effective amount of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein. The pharmaceutical formulations described herein can be administered to a subject in need thereof or a cell.

In some embodiments, the amount of the one or more of the polypeptides, polynucleotides, vectors, cells, virus particles, nanoparticles, other delivery particles, and combinations thereof described herein contained in the pharmaceutical formulation can range from about 1 pg/kg to about 10 mg/kg based upon the bodyweight of the subject in need thereof or average bodyweight of the specific patient population to which the pharmaceutical formulation can be administered. The amount of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein in the pharmaceutical formulation can range from about 1 pg to about 10 g, from about 10 nL to about 10 mL. In embodiments where the pharmaceutical formulation contains one or more cells, the amount can range from about 1 cell to  $1 \times 10^2$ ,  $1 \times 10^3$ ,  $1 \times 10^4$ ,  $1 \times 10^5$ ,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$  or more cells. In embodiments where the pharmaceutical formulation contains one or more cells, the amount can range from about 1 cell to  $1 \times 10^2$ ,  $1 \times 10^3$ ,  $1 \times 10^4$ ,  $1 \times 10^5$ ,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$  or more cells per nL,  $\mu$ L, mL, or L.

In embodiments, where engineered AAV capsid particles are included in the formulation, the formulation can contain 1 to  $1 \times 10^2$ ,  $1 \times 10^3$ ,  $1 \times 10^4$ ,  $1 \times 10^5$ ,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$ ,  $1 \times 10^{13}$ ,  $1 \times 10^{14}$ ,  $1 \times 10^{15}$ ,  $1 \times 10^{16}$ ,  $1 \times 10^{17}$ ,  $1 \times 10^{18}$ ,  $1 \times 10^{19}$ , or  $1 \times 10^{20}$  transducing units (TU)/mL of the engineered AAV capsid particles. In some embodiments, the formulation can be 0.1 to 100 mL in volume and can contain 1 to  $1 \times 10^2$ ,  $1 \times 10^3$ ,  $1 \times 10^4$ ,  $1 \times 10^5$ ,  $1 \times 10^6$ ,  $1 \times 10^7$ ,  $1 \times 10^8$ ,  $1 \times 10^9$ ,  $1 \times 10^{10}$ ,  $1 \times 10^{11}$ ,  $1 \times 10^{12}$ ,  $1 \times 10^{13}$ ,  $1 \times 10^{14}$ ,  $1 \times 10^{15}$ ,  $1 \times 10^{16}$ ,  $1 \times 10^{17}$ ,  $1 \times 10^{18}$ ,  $1 \times 10^{19}$ , or  $1 \times 10^{20}$  transducing units (TU)/mL of the engineered AAV capsid particles.

Pharmaceutically Acceptable Carriers and Auxiliary Ingredients and Agents

In embodiments, the pharmaceutical formulation containing an amount of one or more of the polypeptides, polynucleotides, vectors, cells, virus particles, nanoparticles, other delivery particles, and combinations thereof described herein can further include a

5 pharmaceutically acceptable carrier. Suitable pharmaceutically acceptable carriers include, but are not limited to, water, salt solutions, alcohols, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxy methylcellulose, and polyvinyl pyrrolidone, which do not deleteriously react with the active  
10 composition.

The pharmaceutical formulations can be sterilized, and if desired, mixed with auxiliary agents, such as lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances, and the like which do not deleteriously react with the active composition.

15 In some embodiments, the pharmaceutical formulations described herein may be in a dosage form. The dosage forms can be adapted for administration by any appropriate route. Appropriate routes include, but are not limited to, oral (including buccal or sublingual), rectal, epidural, intracranial, intraocular, inhaled, intranasal, topical (including buccal, sublingual, or transdermal), vaginal, intraurethral, parenteral, intracranial, subcutaneous,  
20 intramuscular, intravenous, intraperitoneal, intradermal, intraosseous, intracardiac, intraarticular, intracavernous, intrathecal, intravitreal, intracerebral, gingival, subgingival, intracerebroventricular, and intradermal. Such formulations may be prepared by any method known in the art.

Dosage forms adapted for oral administration can be discrete dosage units such as  
25 capsules, pellets or tablets, powders or granules, solutions, or suspensions in aqueous or non-aqueous liquids; edible foams or whips, or in oil-in-water liquid emulsions or water-in-oil liquid emulsions. In some embodiments, the pharmaceutical formulations adapted for oral administration also include one or more agents which flavor, preserve, color, or help disperse the pharmaceutical formulation. Dosage forms prepared for oral administration can also be in  
30 the form of a liquid solution that can be delivered as foam, spray, or liquid solution. In some embodiments, the oral dosage form can contain about 1 ng to 1000 g of a pharmaceutical

formulation containing a therapeutically effective amount or an appropriate fraction thereof of the targeted effector fusion protein and/or complex thereof or composition containing the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein. The oral dosage form can be administered to a subject in need thereof.

5           Where appropriate, the dosage forms described herein can be microencapsulated.

          The dosage form can also be prepared to prolong or sustain the release of any ingredient. In some embodiments, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein can be the ingredient whose release is delayed. In other embodiments, the release of an optionally included auxiliary ingredient is delayed. Suitable methods for delaying the release of an ingredient include, but are not limited to, coating or embedding the ingredients in material in polymers, wax, gels, and the like. Delayed release dosage formulations can be prepared as described in standard references such as "Pharmaceutical dosage form tablets," eds. Liberman et. al. (New York, Marcel Dekker, Inc., 1989), "Remington - The science and practice of pharmacy", 20th ed., 10           Lippincott Williams & Wilkins, Baltimore, MD, 2000, and "Pharmaceutical dosage forms and drug delivery systems", 6th Edition, Ansel et al., (Media, PA: Williams and Wilkins, 15           1995). These references provide information on excipients, materials, equipment, and processes for preparing tablets and capsules and delayed release dosage forms of tablets and pellets, capsules, and granules. The delayed release can be anywhere from about an hour to 20           about 3 months or more.

          Examples of suitable coating materials include, but are not limited to, cellulose polymers such as cellulose acetate phthalate, hydroxypropyl cellulose, hydroxypropyl methylcellulose, hydroxypropyl methylcellulose phthalate, and hydroxypropyl methylcellulose acetate succinate; polyvinyl acetate phthalate, acrylic acid polymers and 25           copolymers, and methacrylic resins that are commercially available under the trade name EUDRAGIT® (Roth Pharma, Westerstadt, Germany), zein, shellac, and polysaccharides.

          Coatings may be formed with a different ratio of water-soluble polymer, water insoluble polymers, and/or pH dependent polymers, with or without water insoluble/water soluble non-polymeric excipient, to produce the desired release profile. The coating is either 30           performed on the dosage form (matrix or simple) which includes, but is not limited to, tablets (compressed with or without coated beads), capsules (with or without coated beads), beads,

particle compositions, "ingredient as is" formulated as, but not limited to, suspension form or as a sprinkle dosage form.

Dosage forms adapted for topical administration can be formulated as ointments, creams, suspensions, lotions, powders, solutions, pastes, gels, sprays, aerosols, or oils. In some embodiments for treatments of the eye or other external tissues, for example the mouth or the skin, the pharmaceutical formulations are applied as a topical ointment or cream. When formulated in an ointment, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein can be formulated with a paraffinic or water-miscible ointment base. In some embodiments, the active ingredient can be formulated in a cream with an oil-in-water cream base or a water-in-oil base. Dosage forms adapted for topical administration in the mouth include lozenges, pastilles, and mouth washes.

Dosage forms adapted for nasal or inhalation administration include aerosols, solutions, suspension drops, gels, or dry powders. In some embodiments, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein is contained in a dosage form adapted for inhalation is in a particle-size-reduced form that is obtained or obtainable by micronization. In some embodiments, the particle size of the size reduced (e.g., micronized) compound or salt or solvate thereof, is defined by a D50 value of about 0.5 to about 10 microns as measured by an appropriate method known in the art.

Dosage forms adapted for administration by inhalation also include particle dusts or mists. Suitable dosage forms wherein the carrier or excipient is a liquid for administration as a nasal spray or drops include aqueous or oil solutions/suspensions of an active ingredient (e.g., the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein and/or auxiliary active agent), which may be generated by various types of metered dose pressurized aerosols, nebulizers, or insufflators.

In some embodiments, the dosage forms can be aerosol formulations suitable for administration by inhalation. In some of these embodiments, the aerosol formulation can contain a solution or fine suspension of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein and a pharmaceutically acceptable aqueous or non-aqueous solvent. Aerosol formulations can be presented in single or multi-dose quantities in sterile form in a sealed container. For some of these embodiments, the

sealed container is a single dose or multi-dose nasal, or an aerosol dispenser fitted with a metering valve (e.g., metered dose inhaler), which is intended for disposal once the contents of the container have been exhausted.

Where the aerosol dosage form is contained in an aerosol dispenser, the dispenser  
5 contains a suitable propellant under pressure, such as compressed air, carbon dioxide, or an organic propellant, including but not limited to a hydrofluorocarbon. The aerosol formulation dosage forms in other embodiments are contained in a pump-atomizer. The pressurized aerosol formulation can also contain a solution or a suspension of one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein. In  
10 further embodiments, the aerosol formulation can also contain co-solvents and/or modifiers incorporated to improve, for example, the stability and/or taste and/or fine particle mass characteristics (amount and/or profile) of the formulation. Administration of the aerosol formulation can be once daily or several times daily, for example 2, 3, 4, or 8 times daily, in which 1, 2, or 3 doses are delivered each time.

15 For some dosage forms suitable and/or adapted for inhaled administration, the pharmaceutical formulation is a dry powder inhalable formulation. In addition to the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein, an auxiliary active ingredient, and/or pharmaceutically acceptable salt thereof, such a dosage form can contain a powder base such as lactose, glucose, trehalose, mannitol, and/or  
20 starch. In some of these embodiments, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein is in a particle-size reduced form. In further embodiments, a performance modifier, such as L-leucine or another amino acid, cellobiose octaacetate, and/or metals salts of stearic acid, such as magnesium or calcium stearate.

25 In some embodiments, the aerosol dosage forms can be arranged so that each metered dose of aerosol contains a predetermined amount of an active ingredient, such as the one or more of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein.

Dosage forms adapted for vaginal administration can be presented as pessaries,  
30 tampons, creams, gels, pastes, foams, or spray formulations. Dosage forms adapted for rectal administration include suppositories or enemas.

Dosage forms adapted for parenteral administration and/or adapted for any type of injection (e.g. intravenous, intraperitoneal, subcutaneous, intramuscular, intradermal, intraosseous, epidural, intracardiac, intraarticular, intracavernous, gingival, subgingival, intrathecal, intravitreal, intracerebral, and intracerebroventricular) can include aqueous and/or non-aqueous sterile injection solutions, which can contain anti-oxidants, buffers, bacteriostats, solutes that render the composition isotonic with the blood of the subject, and aqueous and non-aqueous sterile suspensions, which can include suspending agents and thickening agents. The dosage forms adapted for parenteral administration can be presented in a single- unit dose or multi-unit dose containers, including but not limited to sealed ampoules or vials. The doses can be lyophilized and resuspended in a sterile carrier to reconstitute the dose prior to administration. Extemporaneous injection solutions and suspensions can be prepared in some embodiments, from sterile powders, granules, and tablets.

Dosage forms adapted for ocular administration can include aqueous and/or nonaqueous sterile solutions that can optionally be adapted for injection, and which can optionally contain anti-oxidants, buffers, bacteriostats, solutes that render the composition isotonic with the eye or fluid contained therein or around the eye of the subject, and aqueous and nonaqueous sterile suspensions, which can include suspending agents and thickening agents.

For some embodiments, the dosage form contains a predetermined amount of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein per unit dose. In some embodiments, the predetermined amount of the Such unit doses may therefore be administered once or more than once a day. Such pharmaceutical formulations may be prepared by any of the methods well known in the art.

#### Incorporation by Reference

References and citations to other documents, such as patents, patent applications, patent publications, journals, books, papers, web contents, have been made throughout this disclosure. All such documents are hereby incorporated herein by reference in their entirety for all purposes.

### Equivalents

Various modifications of the invention and many further embodiments thereof, in addition to those shown and described herein, will become apparent to those skilled in the art from the full contents of this document, including references to the scientific and patent literature cited herein. The subject matter herein contains important information, exemplification and guidance that can be adapted to the practice of this invention in its various embodiments and equivalents thereof.

### **EXAMPLES**

#### *in vitro* expression of dysferlin

Dysferlin-intein plasmids were generated for *in vitro* experiments in human embryonic kidney (HEK293T) cells.

**FIG. 4** is a diagram of plasmids used to transfect HEK293T cells where the dysferlin-intein is expressed under the control of an ubiquitously active CMV promoter.. Each nucleic acid is packaged between ITR regions, and comprises regulatory elements, including a cytomegalovirus (CMV) promoter, and a poly-A tail. One nucleic acid provided the N-terminal portion of the dysferlin protein together with the N-terminal portion of a split intein while a second nucleic acid provided the C-terminal portion of the dysferlin protein together with the C-terminal portion of a split intein.

5 different inteins constructs were tested: M86, Rma, Ssp, Npu, and Mja.

**FIG. 5** is an anti-HA western blot of HEK293T transfected with the dysferlin-intein plasmids.

**FIG. 6** is an anti-FLAG western blot of HEK293T transfected with the dysferlin-intein plasmids.

As shown, when both N-terminal and C-terminal nucleic acid molecules were provided *in vitro*, the full length (~230 kDa) dysferlin protein was formed in HEK293T cells, with the Rma construct providing robust results.

The Rma intein construct was further tested for dysferlin protein expression in wild type or dysferlin knock out differentiated mouse myoblast C2C12 cell lines by AAV transduction.

**FIG. 7** is a diagram of constructs use to transduce C2C12 cells. Each nucleic acid is packaged between ITR regions. One nucleic acid provided the N-terminal portion of the dysferlin protein together with the N-terminal portion of the Rma-intein while a second nucleic acid provided the C-terminal portion of the dysferlin protein together with the C-terminal portion of the Rma-intein.

**FIG. 8** is an anti-HA western blot of C2C12 myotubes transduced with the AAV-dysferlin-intein.

**FIG. 9** is an anti-FLAG western blot of C2C12 myotubes transduced with the AAV-dysferlin-intein.

**FIG. 10** is a western blot of dysferlin knock out C2C12 myotubes transduced with the AAV-dysferlin-intein.

**FIG. 11** is a western blot of dysferlin knock out C2C12 myotubes transduced with the AAV-dysferlin-intein.

As shown, when both Rma-intein N-terminal and C-terminal nucleic acid molecules were provided *in vitro*, the full length (~230 kDa) dysferlin protein was formed in dysferlin KO C2C12 myotubes. Remarkably, expression of the full length dysferlin protein was higher than the expression of dysferlin in untransduced wild type cells.

#### *in vivo expression of dysferlin*

Full length dysferlin expression was tested in mice with the progressive muscular dystrophy (prmd) allele from the A/J inbred strain (BLA/J mice) *in vivo* after systemic administration of the MyoAAV-dysferlin-intein construct at 1.6E+13 vg/kg. An antibody recognizing dysferlin was used for protein expression analysis.

The mice were divided into two groups. Group 1 had tissues harvested 2 weeks post-injection. Group 2 had tissues harvested 4 weeks post injection.

**FIG. 12** is a western blot of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein.

**FIG. 13** is a bar graph of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein.

**FIG. 14** is a western blot of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein.

**FIG. 15** is a bar graph of dysferlin expression in BLA/J mice *in vivo* after administration of the AAV-dysferlin-intein.

Full length dysferlin was expressed across all groups and in gastrocnemius tissues. By week 4, mice in Group 2 showed higher dysferlin expression in gastrocnemius tissue than wild type mice.

#### Split-intein expression

Dysferlinopathy (LGMD2B, LGMDR1, Myoshi Myopathy) is a genetic myopathy caused by mutations in the DYSF gene resulting in the absence of dysferlin, a membrane protein involved in muscle membrane repair (membrane patch). Dysferlin deficiency compromises the sarcolemma integrity and leads to progressive muscle degeneration. Gene replacement therapy is a promising therapeutic strategy for dysferlinopathy. However, the dysferlin coding sequence (~6.2 kb) exceeds the packing capacity of a single AAV. To overcome this hurdle, dual AAV strategies based on homologous recombination of two vector genomes in myonuclei have been proposed to express the full-length dysferlin in animal muscles. The requirement for delivering vector genomes from the two AAVs into the same nuclei in muscle fibers, as well as low efficiency of homologous recombination in post mitotic cells results in ineffective production of full length dysferlin with this approach.

The present invention provides an alternative dual AAV strategy utilizing split protein inteins to express full-length dysferlin.

#### Experimental details

A dual vector system was utilized in which two segments of the dysferlin protein were expressed individually and the full-length protein is generated by protein splicing in the muscle fiber cytoplasm.

(1) Multiple split inteins were tested *in vitro*. HEK293T cells were transfected with split dysferlin intein constructs. 72 hours post transfection, cells were harvested and protein lysates were prepared for western blot analysis.

(2) Dysferlin knock-out C2C12 mouse myoblasts were cultured and differentiated for 96 hours. Myotubes were transduced with dual MyoAAV split dysferlin MyoAAV vectors

and harvested 72 hours post transduction. Dysferlin protein expression was evaluated by Jess capillary electrophoresis.

(3) MyoAAV dual dysferlin vectors were administered systemically into dysferlin knock out (Bla/J) mice at a dose of  $1.6 \times 10^{13}$  vg/kg. Dysferlin protein expression was measured  
5 by capillary electrophoresis 2 and 4 weeks post injection.

(4) Dual MyoAAV vectors were administered systemically in Bla/J mice at a dose of  $1 \times 10^{13}$  vg/kg. ~9-10 weeks post injection, the Tibialis Anterior muscle was strained through controlled repetitive cycles of eccentric contraction. 24 hours post strain injury, Evan's blue dye (EBD) was administered by IP injection. 16-24 hours post EBD injection, tissues were  
10 collected for histology. EBD uptake was evaluated by fluorescent imaging and quantified as a measure of lack of membrane repair. Tissues were also analyzed by immunofluorescence to evaluate dysferlin expression.

**FIG. 16** is a schematic of the split intein strategy. The human dysferlin coding sequence was split between two AAV cassettes. Split inteins were added to the end of  
15 dysferlin N terminal and beginning of dysferlin C terminal. Split dysferlin constructs were packaged into MyoAAV. Dual vectors were co-administered/transduced, resulting in expression of dysferlin N-terminal +N-intein and C-intein + dysferlin C-terminal fragments and subsequent protein splicing, yielding full length dysferlin protein expression.

20 *(1) Multiple split inteins tested in vitro*

HEK293T cells were transfected with split dysferlin intein constructs. 72 hours post transfection, cells were harvested and protein lysates were prepared for western blot analysis.

**FIG. 17** is a western blot screen of dysferlin inteins. Cells were transfected with individual dysferlin N-terminal and C-terminal constructs, or co-transfected with both N- and  
25 C- terminal constructs. Protein expression was evaluated by western blot using an antibody specific for dysferlin N-terminal. The *in vitro* screen identified split intein B as the most efficient intein for dysferlin protein splicing. Inclusion of intein B resulted in greatest full-length dysferlin protein (~237 kDa) expression.

*(2) Dysferlin knock-out C2C12 mouse myoblasts*

Dysferlin knock-out C2C12 mouse myoblasts were cultured and differentiated for 96 hours. Myotubes were transduced with dual MyoAAV split dysferlin MyoAAV vectors and harvested 72 hours post transduction. Dysferlin protein expression was evaluated by Jess capillary electrophoresis.

**FIG. 18A-B** shows capillary electrophoresis results following split intein transduction in mouse myotubes. Individual MyoAAV vectors result in expression of dysferlin protein N-terminal and C-terminal fragments detectable by capillary electrophoresis using antibodies specific for dysferlin N-terminal (**FIG18A**) and C-terminal (**FIG18B**). Myotubes transduced with dual vectors containing split intein B express full-length dysferlin (~237 kDa).

*(3) Dual dysferlin vectors administered to dysferlin knock out mice 2 and 4 weeks post injection*

MyoAAV dual dysferlin vectors were administered systemically into dysferlin knock out (Bla/J) mice at a dose of  $1.6 \times 10^{13}$  vg/kg. Dysferlin protein expression was measured by capillary electrophoresis 2 and 4 weeks post injection.

**FIG. 19A-D** show the expression of dysferlin in knock out mice following vector administration. Dysferlin protein expression in gastrocnemius of Bla/J mice treated with dual MyoAAVs (at  $1.6 \times 10^{13}$  vg/kg,  $8 \times 10^{12}$  vg/kg per vector) was ~40-60% compared to endogenous levels of dysferlin protein in wild type (WT) mice 2 weeks post injection and >100% compared to WT at 4 weeks. Full-length dysferlin is detectable using antibodies specific for both dysferlin N-terminal (**FIG19A-B**) and C-terminal (**FIG. 19C-D**).

*(4) Dual dysferlin vectors administered to dysferlin knock out mice 9-10 weeks post injection*

Dual MyoAAV vectors were administered systemically in Bla/J mice at a dose of  $1 \times 10^{13}$  vg/kg. ~9-10 weeks post injection, the Tibialis Anterior muscle was strained through controlled repetitive cycles of eccentric contraction. 24 hours post strain injury, Evan's blue dye (EBD) was administered by IP injection. 16-24 hours post EBD injection, tissues were collected for histology. EBD uptake was evaluated by fluorescent imaging and quantified as

a measure of lack of membrane repair. Tissues were also analyzed by immunofluorescence to evaluate dysferlin expression.

**FIG. 20** show images of dysferlin protein expression in the Tibialis Anterior muscle following vector administration. Widespread dysferlin protein expression was demonstrated throughout the Tibialis Anterior (TA) muscle of treated Bla/J animals. Dysferlin immunofluorescent staining of the TA muscle of wild-type C57BL/6 mice, and Bla/J mice treated with vehicle or dual MyoAAV at  $1\text{E}+13$  vg/kg ( $5\text{E}+12$  vg/kg per vector).

**FIG. 21** shows images of membrane repair in mice following vector administration. Evans blue dye (EBD) fluorescent images of strained (right TA) and uninjured (left TA) of wild-type, vehicle treated Bla/J and dual MyoAAV treated Bla/J mice. The damaged TA of vehicle treated mice contain higher number of EBD+ fibers compared to the control TA, indicative of deficiency in membrane repair after strain injury. EBD staining is similar for the damaged and control TA in wild type and MyoAAV treated Bla/J mice, demonstrating effective membrane repair following injury.

**FIG. 22A-B** are graphs showing quantification of membrane repair in mice following vector administration.

Comparison of percent Evans blue dye (EBD) positive area in the damaged TA compared to the uninjured contralateral TA within each group (**FIG. 22A**) and the damaged TA between all groups (**FIG. 22B**) show a significant difference in EBD+ area between the damaged and control TA for vehicle treated mice (A). There is also a significant difference in the EBD+ area in the damaged TA of the vehicle mice compared to the damaged TA of WT or MyoAAV treated Bla/J animals (B).

### Discussion

Highly efficient production of full-length dysferlin protein after delivery of dual MyoAAV vectors encoding for dysferlin N-terminal and C-terminal fragments conjugated to split inteins was successfully demonstrated *in vitro* and *in vivo*. *In vitro* screening of different split inteins identified intein B as the most effective split intein for full-length dysferlin expression. Based on this data, a dual MyoAAV candidate containing the split intein B was tested in myotubes and showed high efficiency of reconstitution and full-length dysferlin protein production. Systemic delivery of dual MyoAAVs in Bla/J mice at low dose resulted

in physiological or supraphysiological full-length dysferlin expression in skeletal muscles. Immunofluorescence analysis showed widespread dysferlin expression across majority of muscle fibers, which improved membrane repair following strain injury in Bla/J muscles. The results provide evidence for the feasibility of using a dual AAV strategy utilizing split  
5 inteins for effective expression of full-length dysferlin protein in dystrophic muscle following low-dose systemic MyoAAV administration.

Compositions and methods of the invention allow for split-intein expression of dysferlin in muscle cells.

10

## CLAIMS

We claim:

1. A composition comprising:
  - a first nucleic acid molecule comprising:
    - a sequence encoding a first portion of the dysferlin protein; and
    - a first portion of a split intein;
  - a second nucleic acid molecule comprising:
    - a sequence encoding a second portion of the dysferlin protein; and
    - a second portion of a split intein,wherein the first nucleic acid molecule and second nucleic acid molecule encode different portions of the dysferlin protein.
2. The composition of claim 1, wherein the first and second portion of the dysferlin protein together provide a functional dysferlin protein.
3. The composition of claim 2, wherein the first nucleic acid molecule is packaged in a first vector and the second nucleic acid molecule is packaged in a second vector.
4. The composition of claim 3, wherein each vector is an adeno-associated viral (AAV) vector.
5. The composition of claim 4, wherein the first portion of the dysferlin protein is between 1100-1150 amino acids in length.
6. The composition of claim 5, wherein the sequence encoding the first portion of the dysferlin protein comprises the C2A-C2C, Fer domains, and dysF domains of the dysferlin protein and the second portion of the dysferlin protein comprises the C2D-C2G domains and transmembrane domain of the dysferlin protein.

7. The composition of claim 4, wherein the first nucleic acid molecule comprises in order:

- an inverted terminal repeats (ITR) region;
- a promoter;
- the sequence encoding a first portion of the dysferlin protein;
- the first portion of a split intein;
- an ITR region, and

the second nucleic acid molecule comprises in order:

- an ITR region;
- a promoter;
- the second portion of a split intein intein;
- the sequence encoding a second portion of the dysferlin protein;
- an ITR region.

8. The composition of claim 4, wherein each AAV vector comprises a capsid protein having at least one modification that results in reduced liver-tropism of the AAV vector and/or preferential targeting of the AAV vector to muscle tissue.

9. The composition of claim 8, wherein when the composition is provided to a subject, the first nucleic acid molecule and second nucleic acid molecule are each expressed in different nuclei in muscle tissue.

10. The composition of claim 9, wherein once expressed, the two portions of the dysferlin protein are joined by splicing of the respective inteins within the cytoplasm of the muscle cells.

11. A method of treating a subject suffering from a dysferlinopathy, the method comprising:

- providing to the subject:

- a first nucleic acid molecule comprising:

- a sequence encoding a first portion of the dysferlin protein; and

- a first portion of a split intein; and

a second nucleic acid molecule comprising a sequence encoding:

a second portion of the dysferlin protein; and

a second portion of a split intein,

wherein the first nucleic acid molecule and second nucleic acid molecule encode different portions of the dysferlin protein.

12. The method of claim 11, wherein the first nucleic acid molecule is packaged in a first vector and the second nucleic acid molecule is packaged in a second vector.

13. The method of claim 12, wherein each vector is an adeno-associated viral (AAV) vector.

14. The method of claim 13, wherein the first portion of the dysferlin protein is between 1100-1150 amino acids in length.

15. The method of claim 14, wherein the sequence encoding the first portion of the dysferlin protein comprises the C2A-C2C, Fer domains, and dysF domains of the dysferlin protein and the second portion of the dysferlin protein comprises the C2D-C2G domains and the transmembrane domain of the dysferlin protein.

16. The method of claim 13, wherein the first nucleic acid molecule comprises in order:

an inverted terminal repeats (ITR) region;

a promoter;

the sequence encoding a first portion of the dysferlin protein;

the first portion of a split intein;

an ITR region, and

the second nucleic acid molecule comprises in order:

an ITR region;

a promoter;

the second portion of a split intein;

the sequence encoding a second portion of the dysferlin protein;

an ITR region.

17. The method of claim 13, wherein each AAV vector comprises a capsid protein having at least one modification that results in reduced liver-tropism of the AAV vector and/or preferential targeting of the AAV vector to muscle tissue.

18. The method of claim 17, wherein when the composition is provided to a subject, the first nucleic acid molecule and second nucleic acid molecule are each expressed in different nuclei in multi-nucleated muscle cells.

19. The composition of claim 9, wherein once expressed, the two portions of the dysferlin protein are joined by splicing of the respective inteins within the cytoplasm of the muscle cells.

20. The method of claim 19, wherein splicing of the respective inteins results in formation of a functional dysferlin protein.

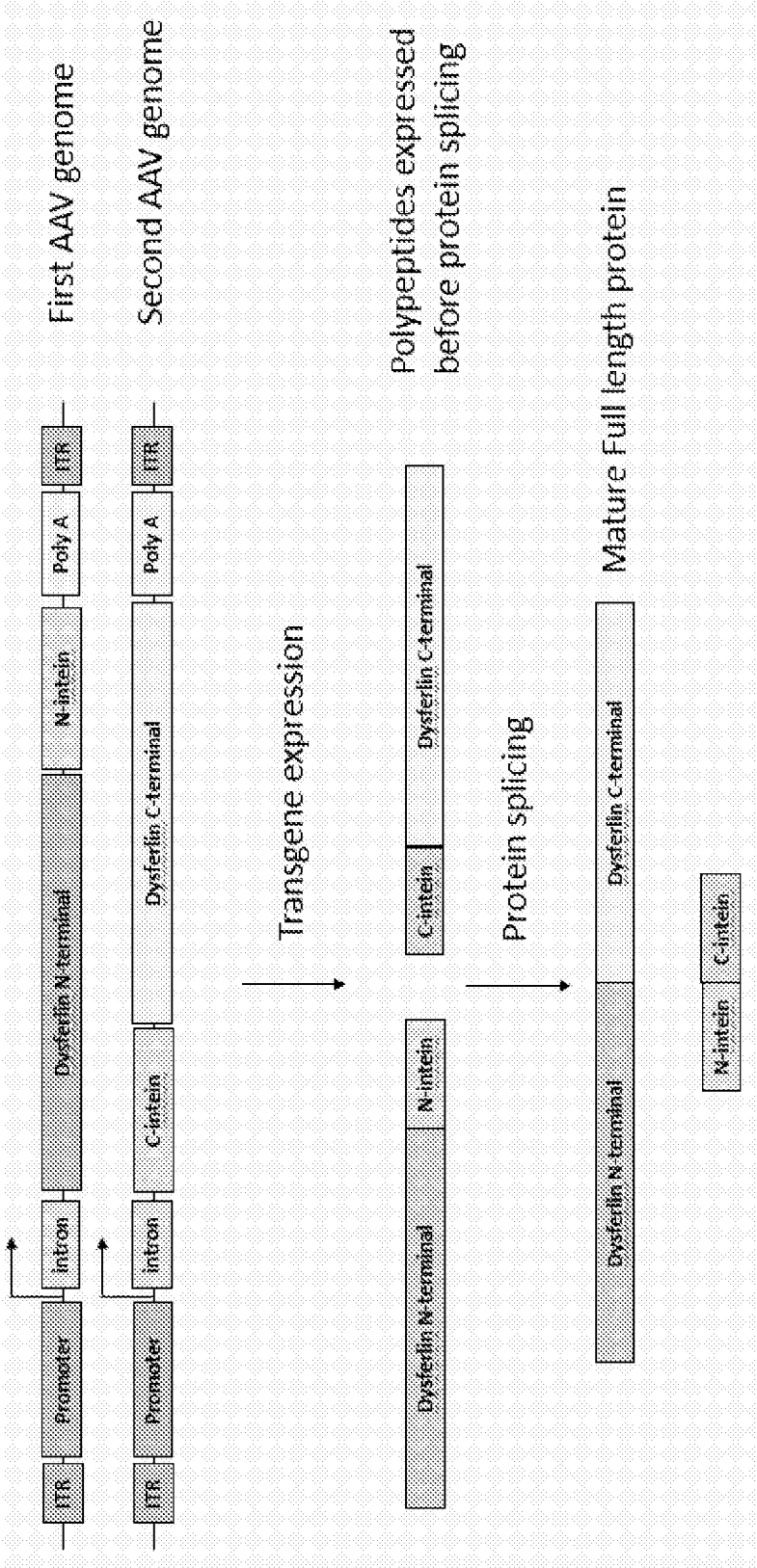


FIG. 1

2/22

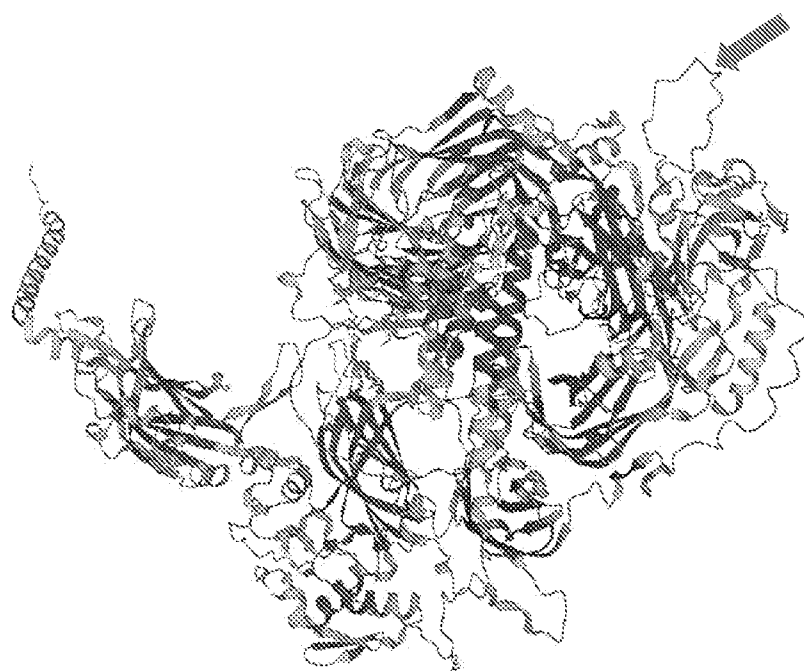


FIG. 2

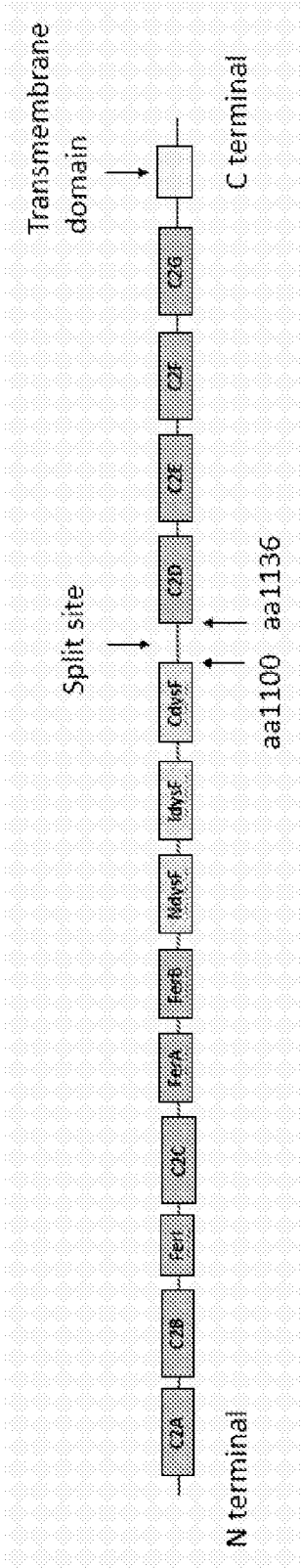


FIG. 3

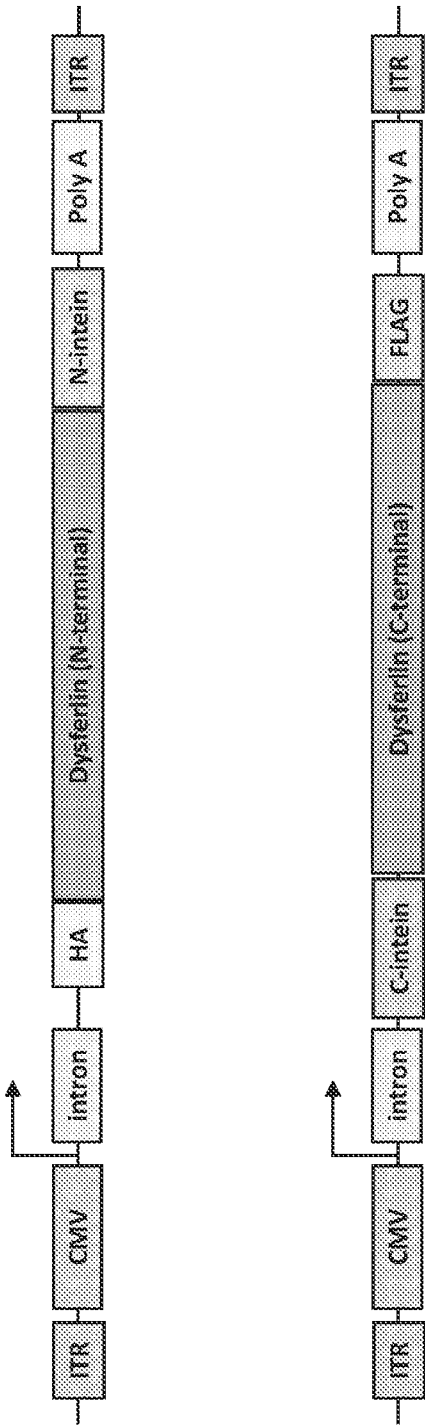
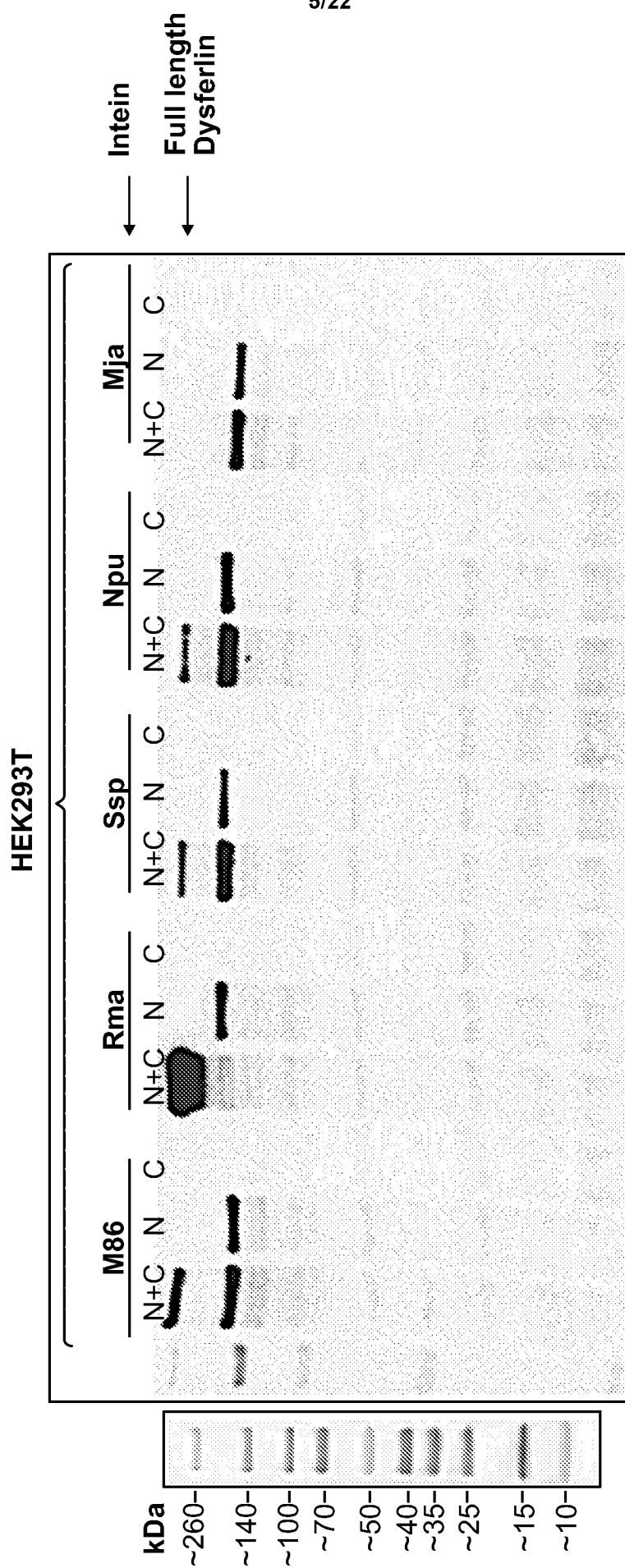
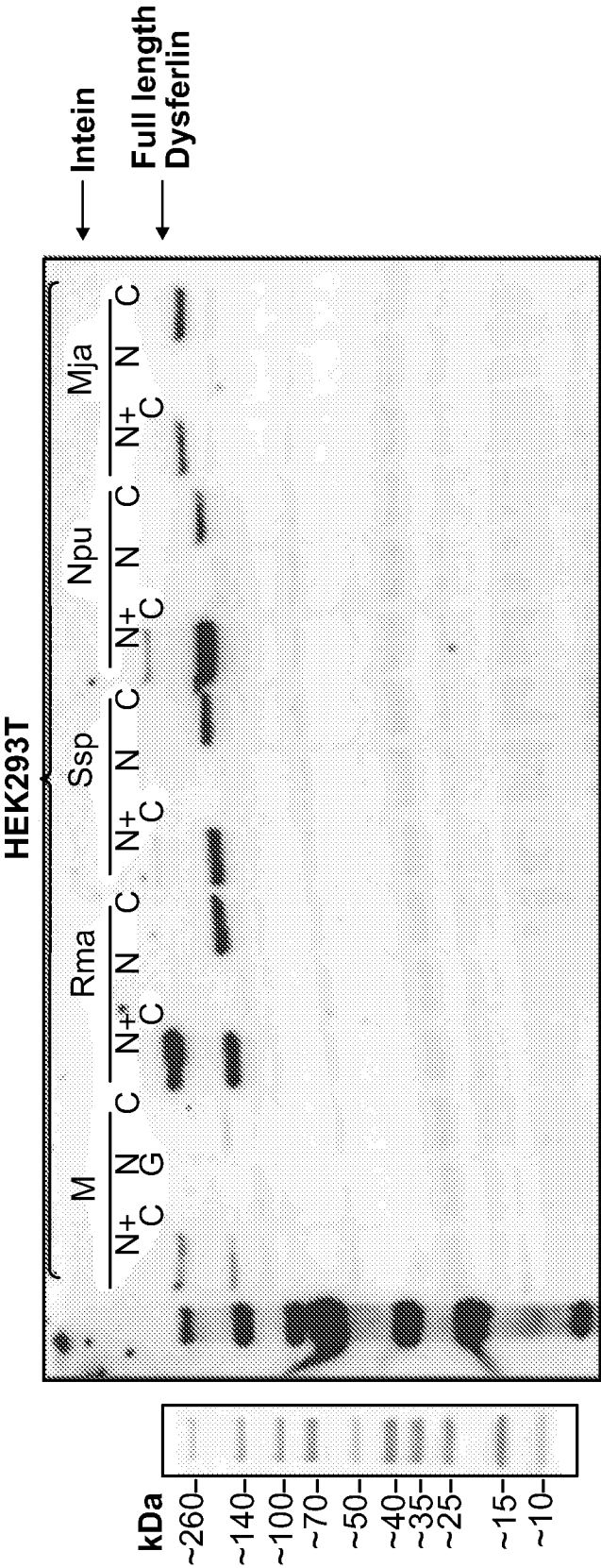


FIG. 4



### Anti-HA western blot

FIG. 5



Anti-HA Western blot

FIG. 6

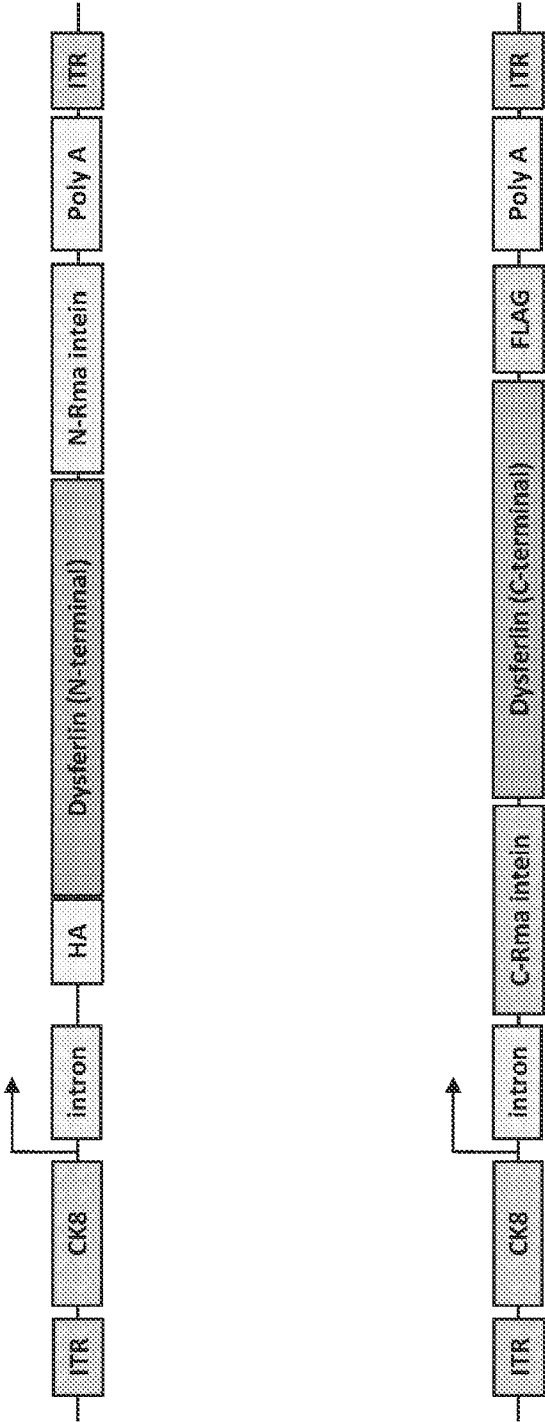


FIG. 7

8/22

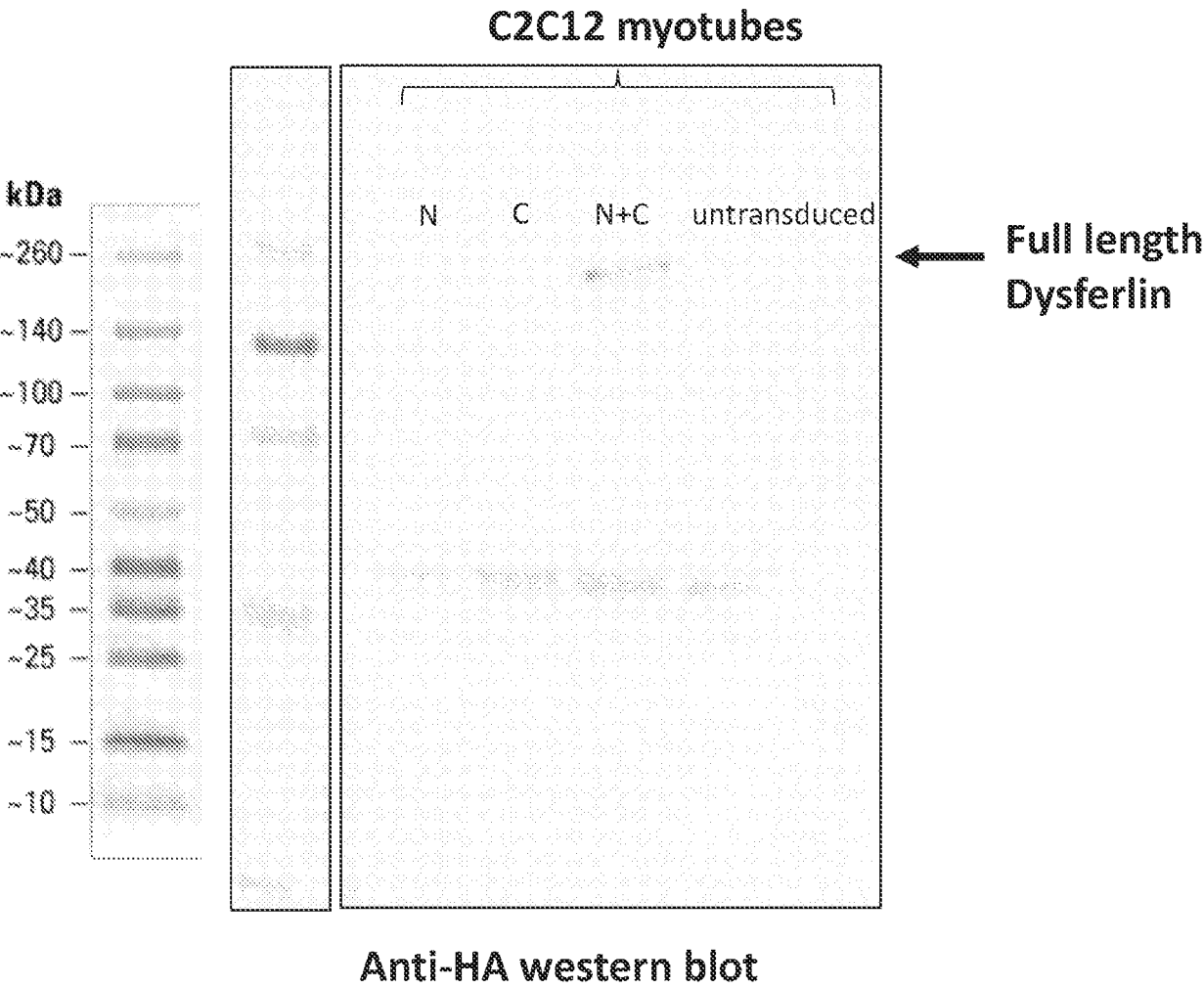


FIG. 8

9/22

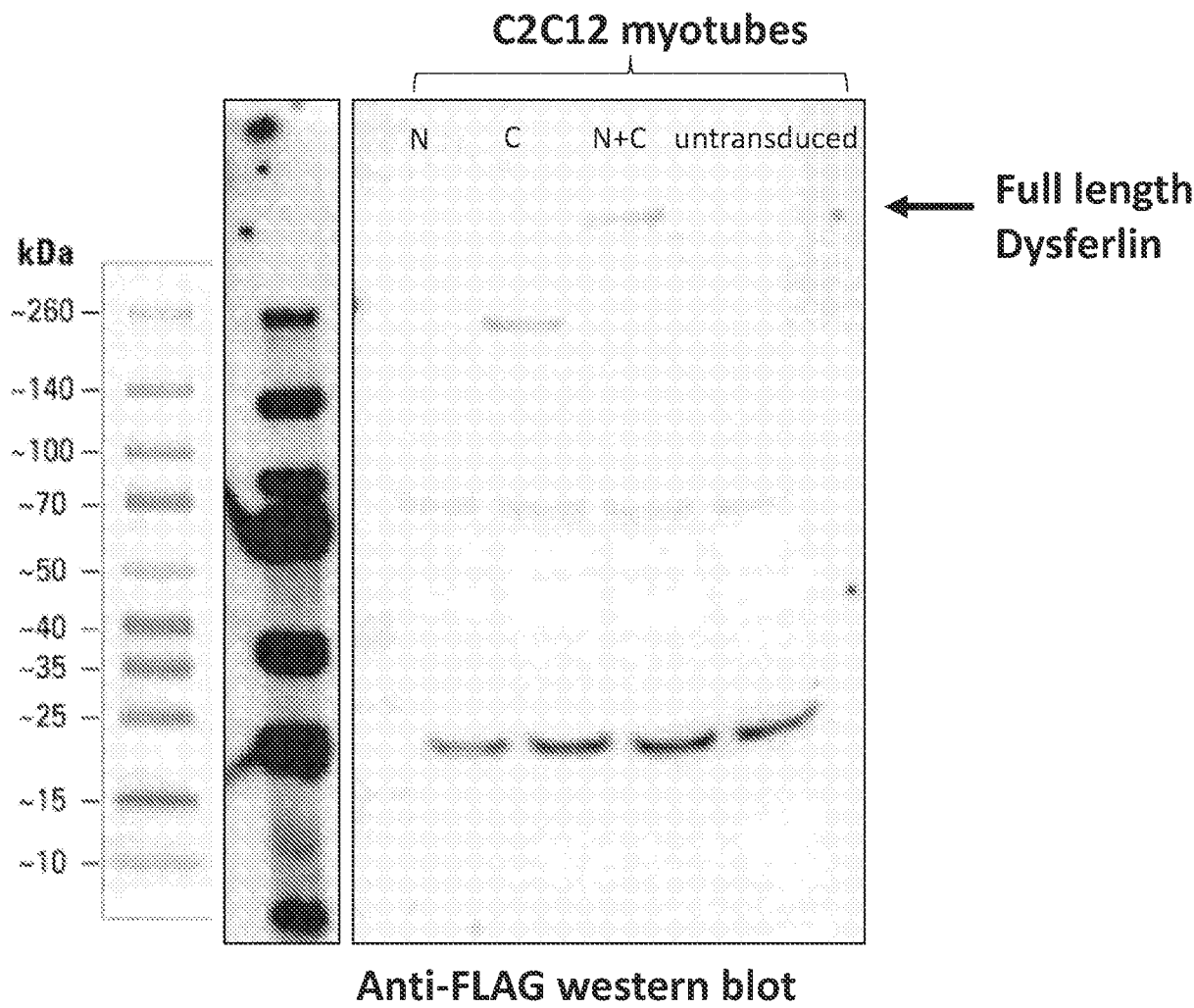


FIG. 9

10/22

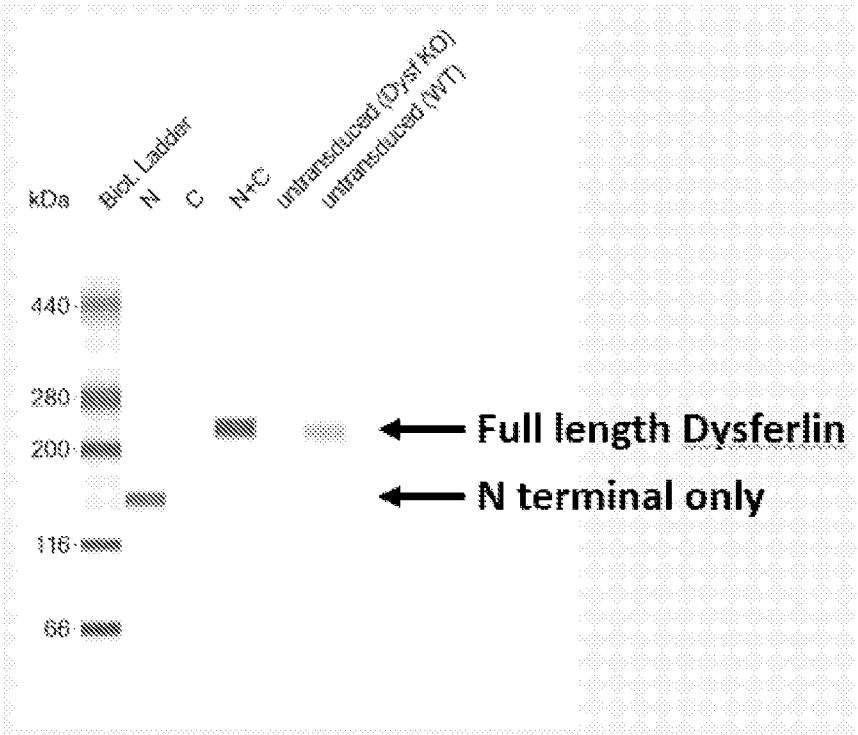


FIG. 10

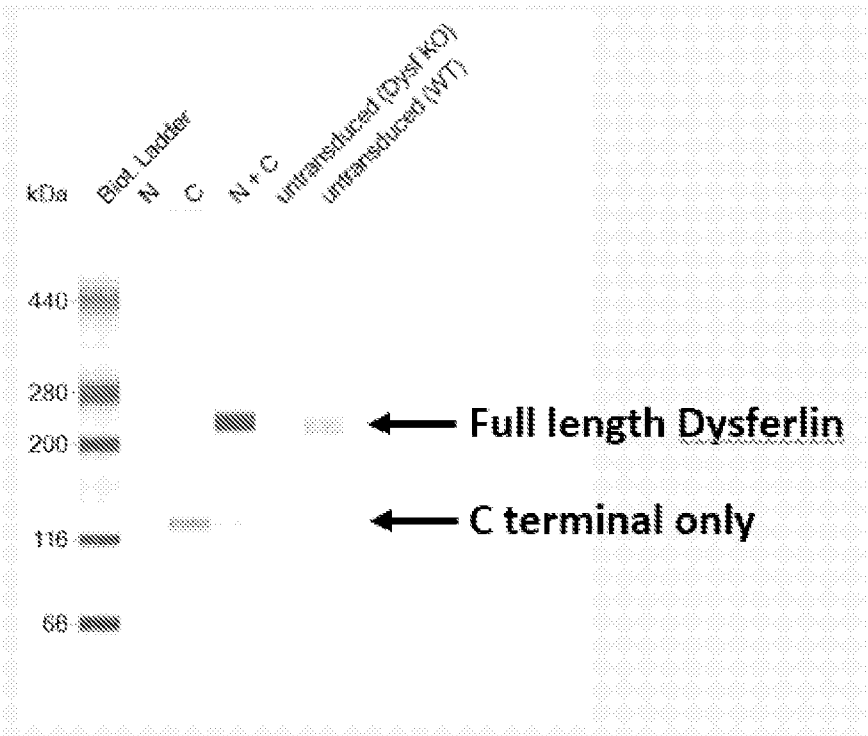


FIG. 11

11/22

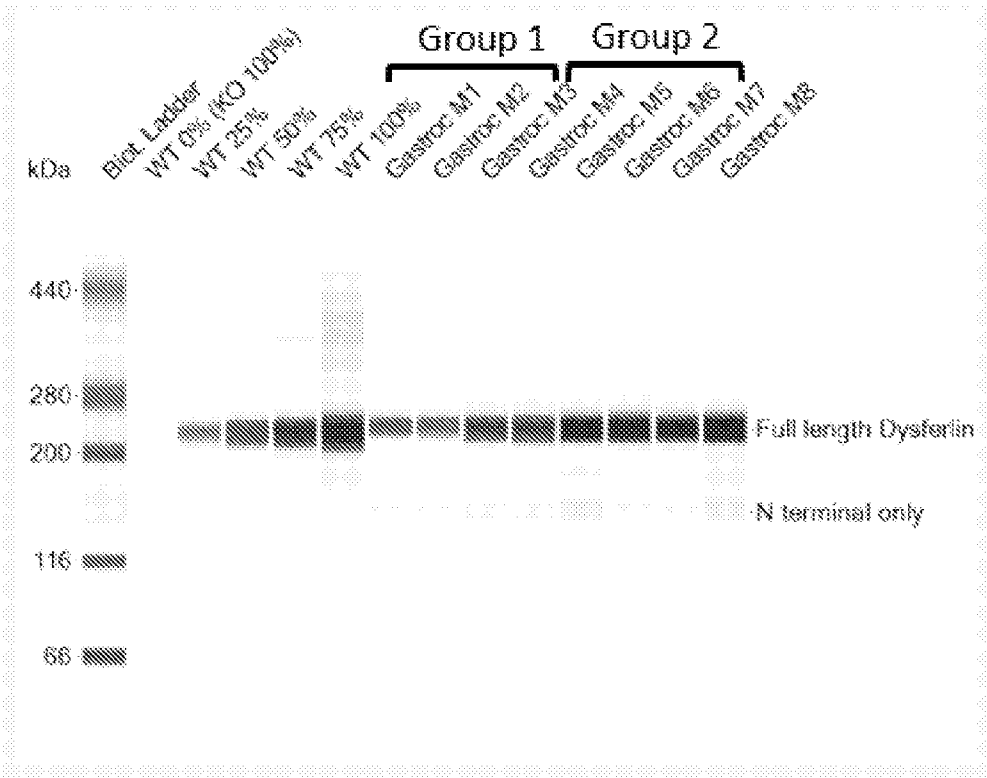


FIG. 12

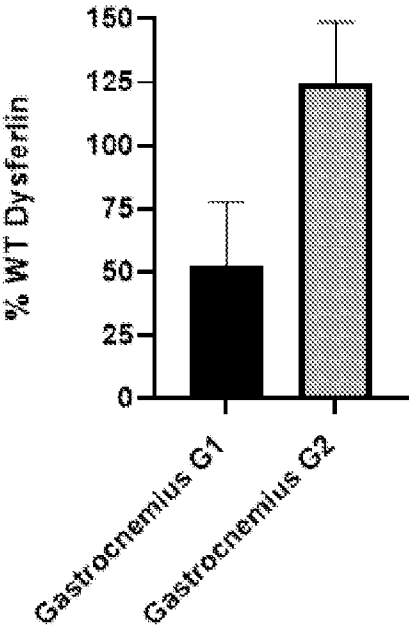


FIG. 13

12/22

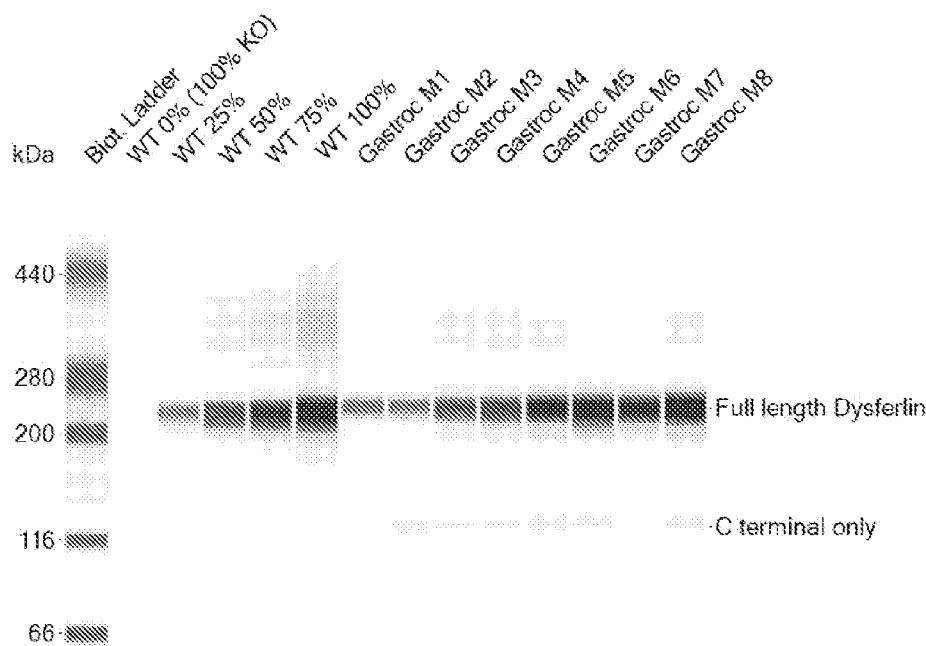


FIG. 14

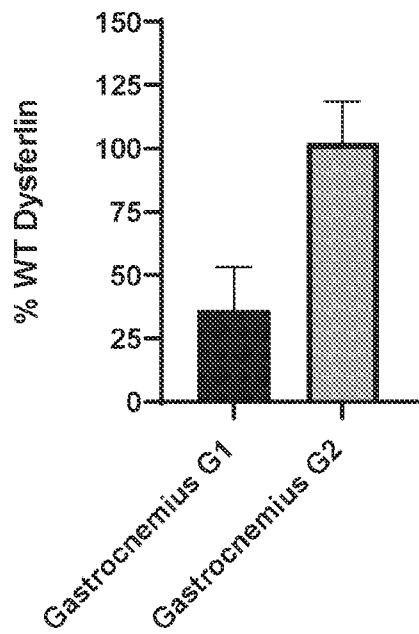


FIG. 15

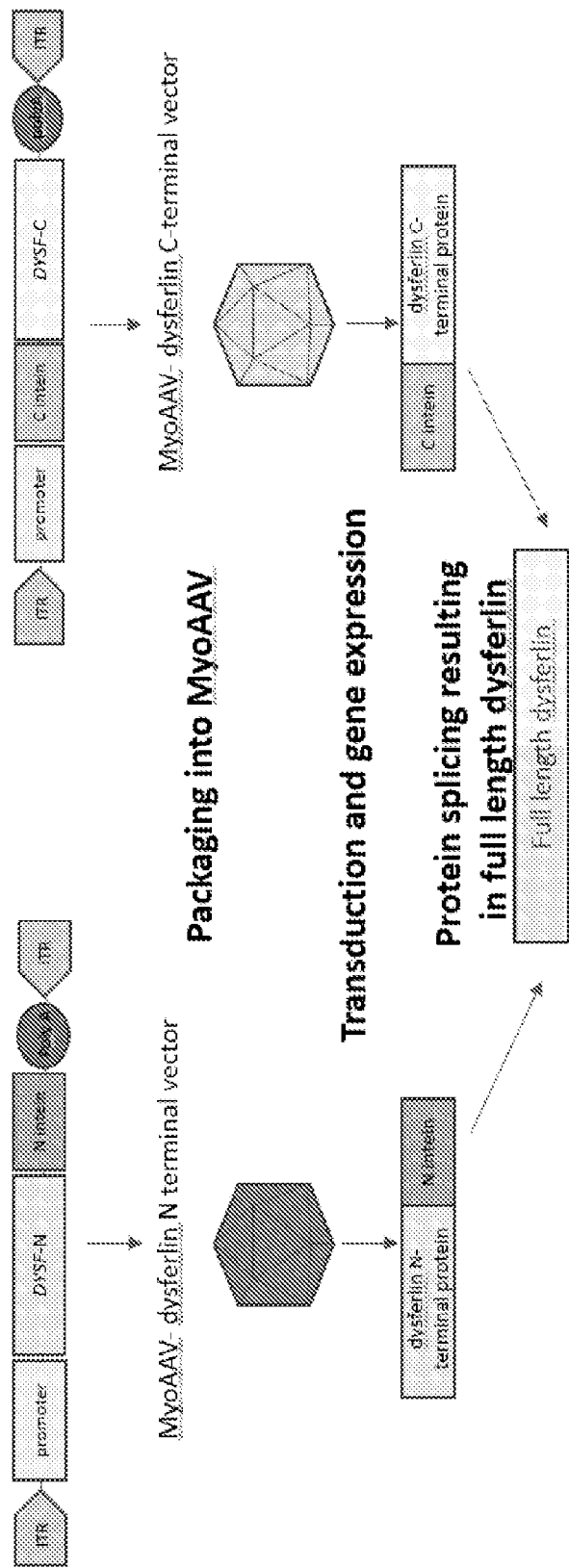


FIG. 16

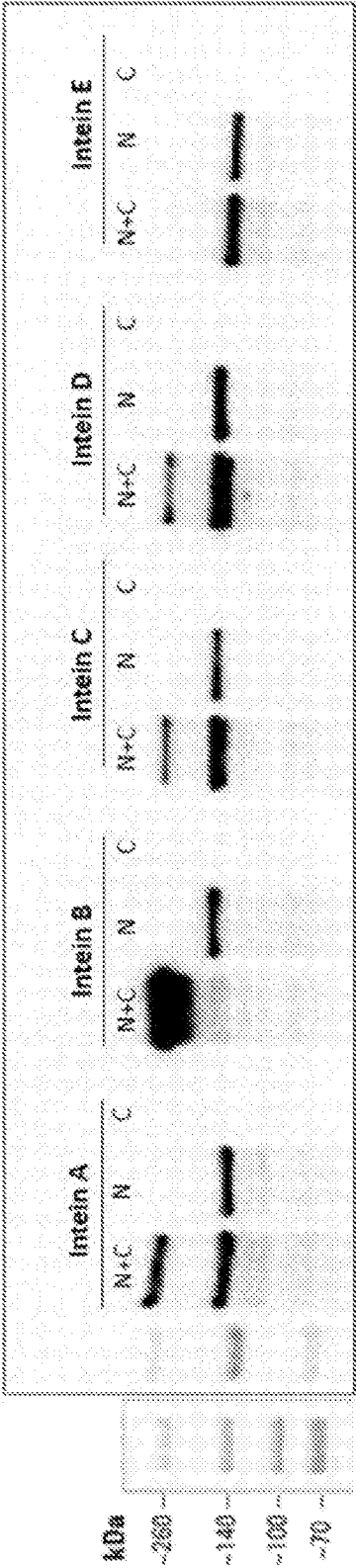


FIG. 17

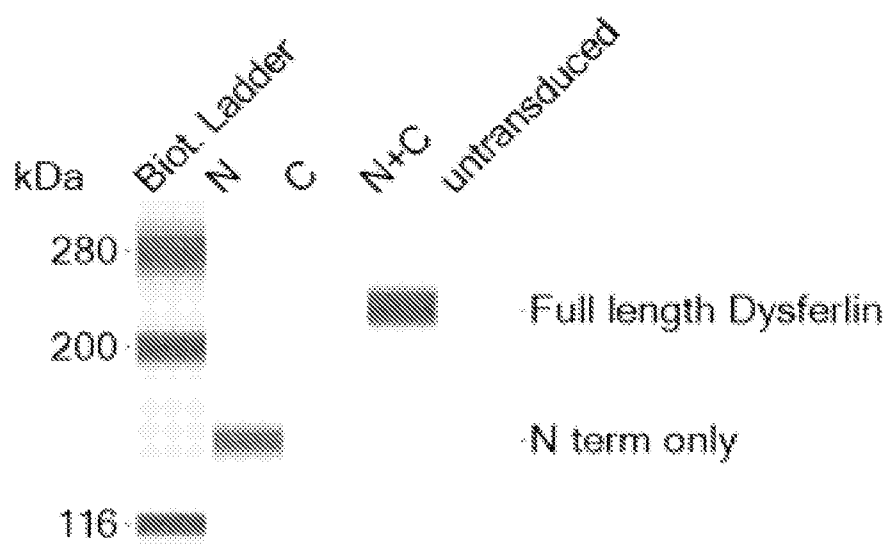


FIG. 18A

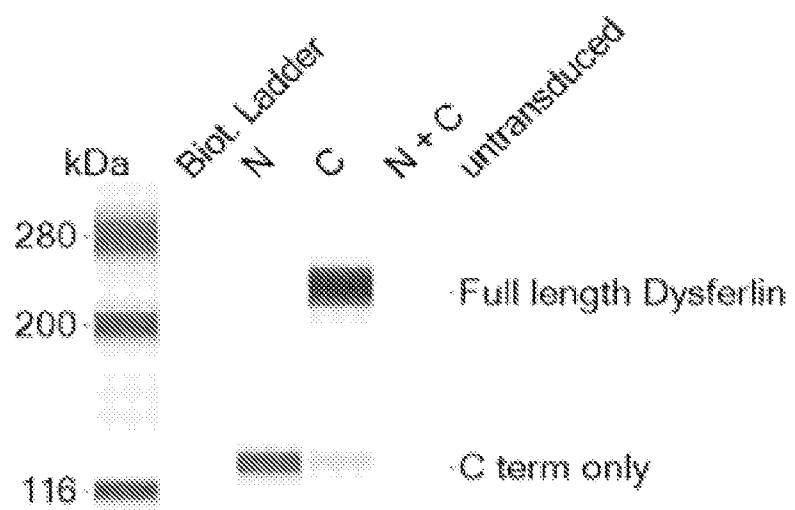


FIG. 18B

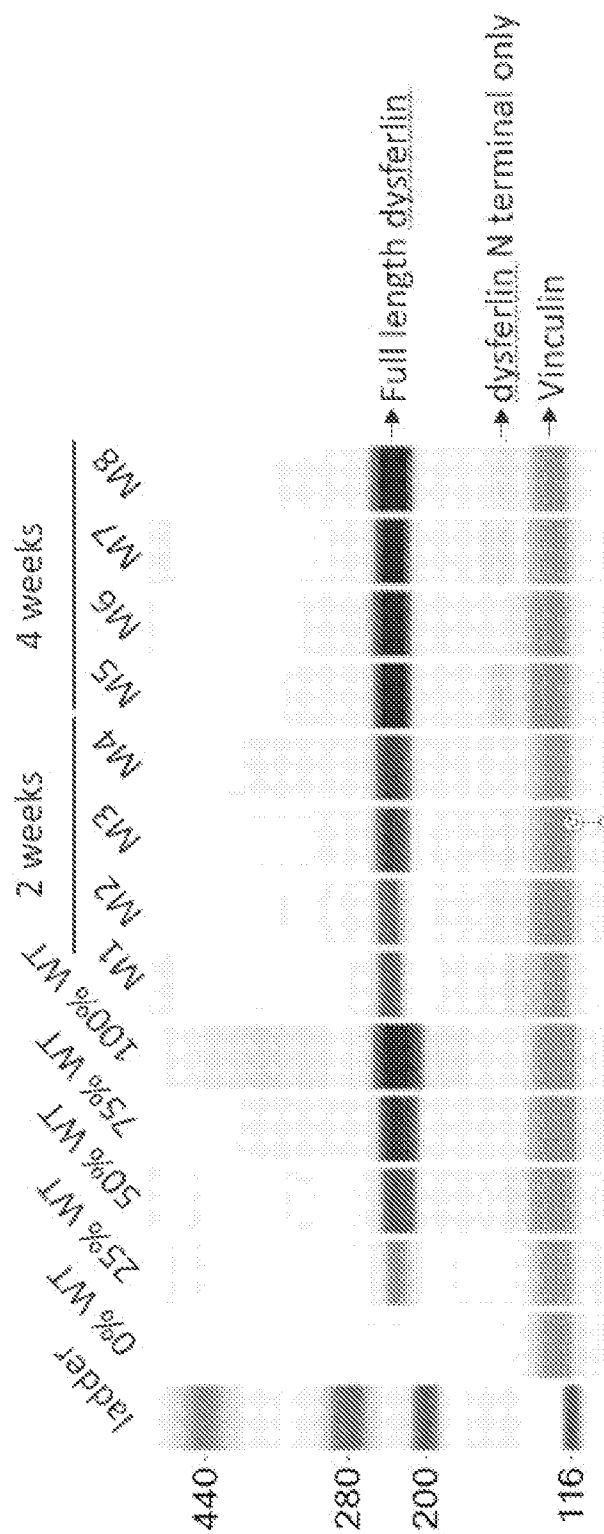


FIG. 19A

17/22

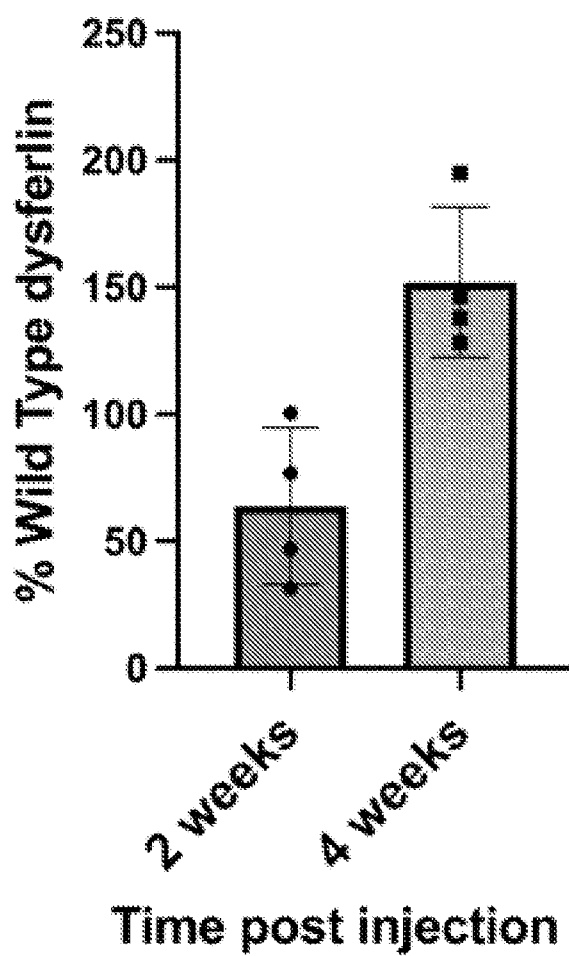


FIG. 19B

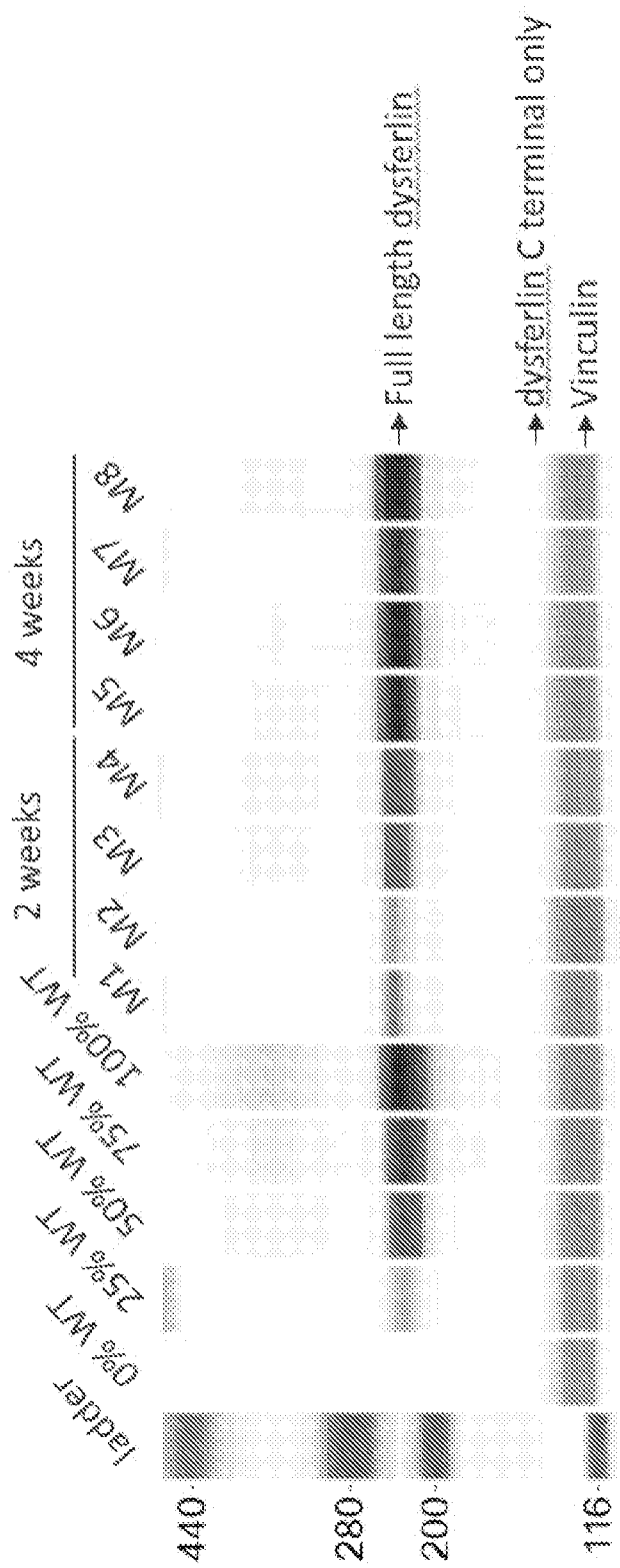


FIG. 19C

19/22

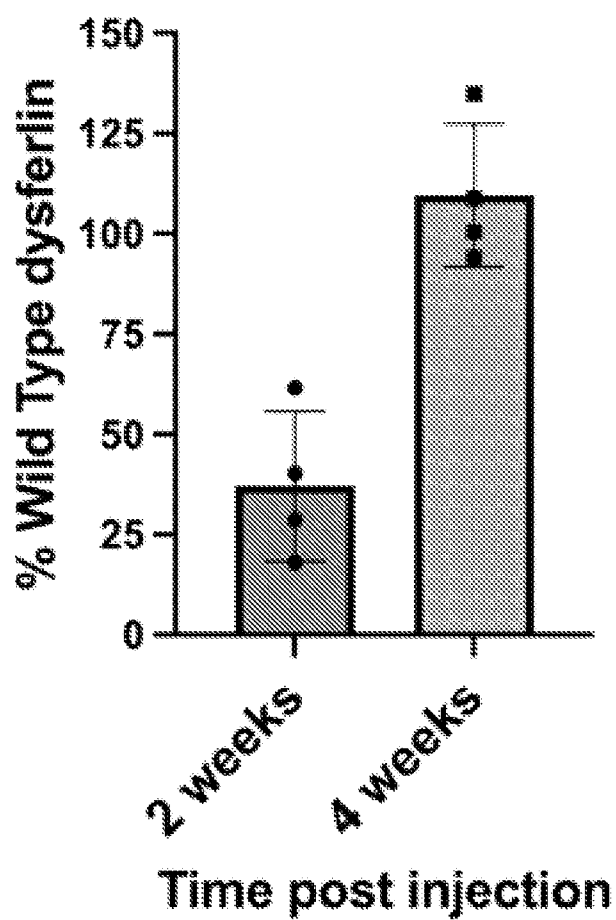


FIG. 19D

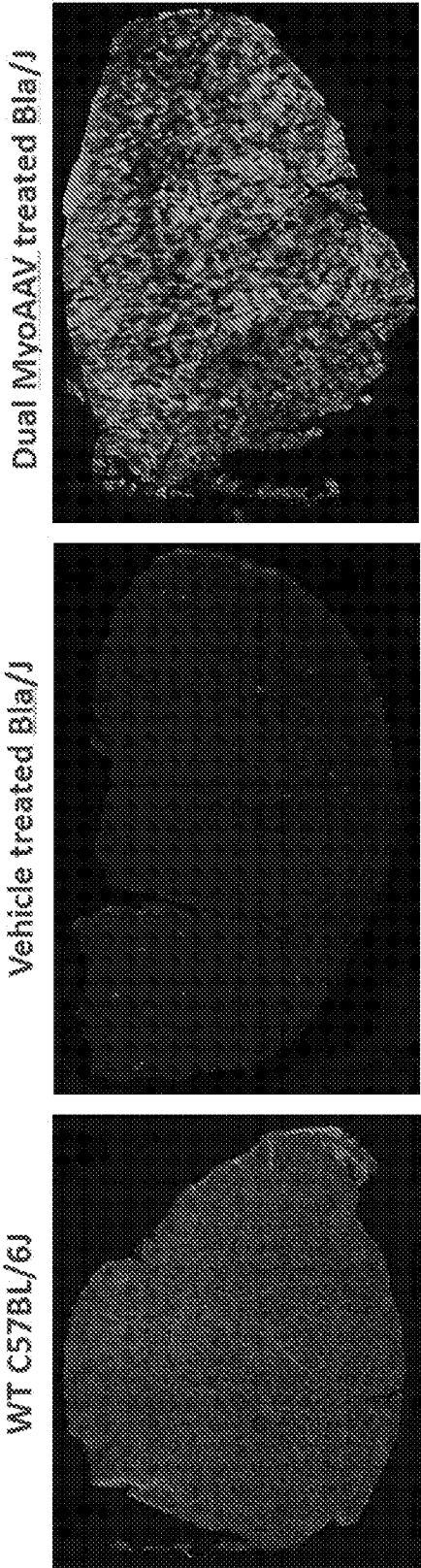


FIG. 20

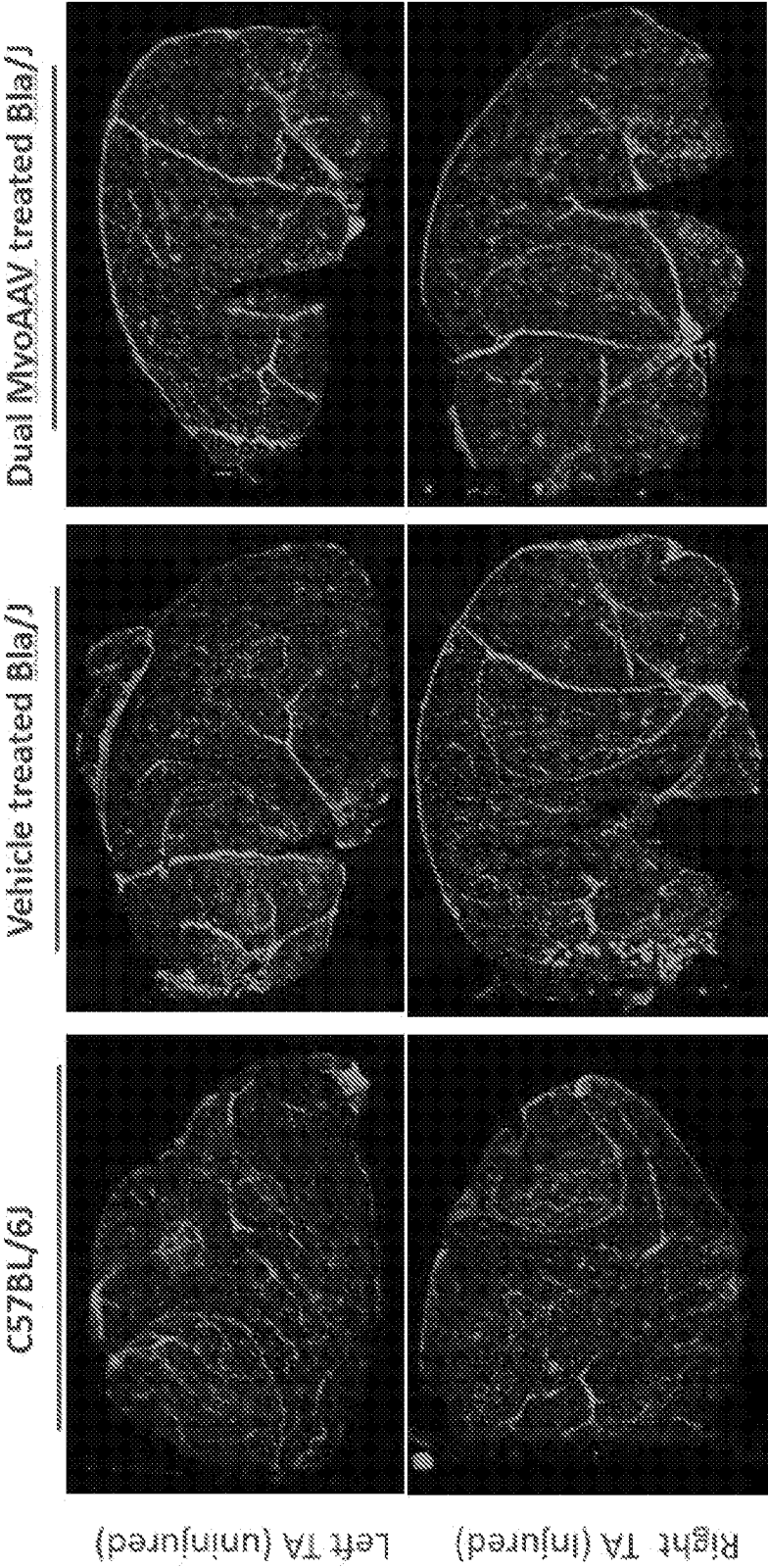


FIG. 21

22/22

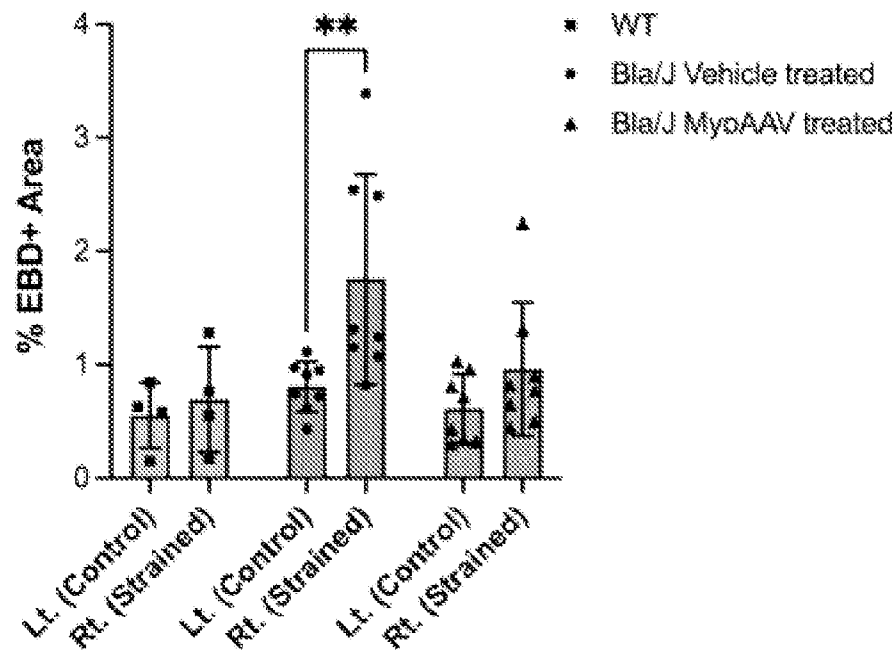


FIG. 22A

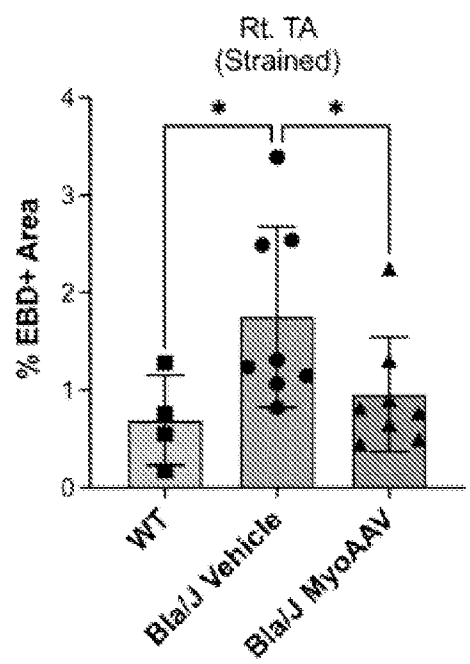


FIG. 22B

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/037366

**A. CLASSIFICATION OF SUBJECT MATTER**IPC: **A61K 38/17** (2024.01); **A61K 38/43** (2024.01); **A61K 48/00** (2024.01); **A61P 21/00** (2024.01); **C07K 14/00** (2024.01); **C12N 15/62** (2024.01); **C12N 15/864** (2024.01); **C12P 21/00** (2024.01); A61K 31/7088 (2024.01)CPC: **A61K 38/17**; **A61K 38/43**; **A61K 48/00**; **A61P 21/00**; **C07K 14/4707**; **C12N 15/62**; **C12N 15/8645**; **C12P 21/00**; C07K 2319/92; A61K 31/7088

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

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

See Search History Document

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

See Search History Document

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                            | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | WO 2023/004125 A2 (UNIVERSITY OF WASHINGTON) 26 January 2023 (26.01.2023)<br>entire document                  | 1-7, 11-16            |
| Y         | entire document                                                                                               | 8-10, 17-20           |
| Y         | US 2021/0107948 A1 (GENETHON et al.) 15 April 2021 (15.04.2021)<br>entire document                            | 8-10, 17-20           |
| Y         | WO 2022/216988 A2 (UNIVERSITY OF WASHINGTON) 13 October 2022 (13.10.2022)<br>entire document                  | 9, 10, 18-20          |
| Y         | US 2020/0407405 A1 (OHIO STATE INNOVATION FOUNDATION et al.) 31 December 2020 (31.12.2020)<br>entire document | 10, 19, 20            |

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

\* Special categories of cited documents:

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

“D” document cited by the applicant in the international application

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

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

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

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

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

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

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

“&amp;” document member of the same patent family

Date of the actual completion of the international search

26 August 2024 (26.08.2024)

Date of mailing of the international search report

09 September 2024 (09.09.2024)

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US  
Commissioner for Patents  
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

MATOS  
TAINA

Telephone No. 571-272-4300