Rational engineering of a functional CpG-free ITR for AAV gene therapy

Xiufang Pan 1, Yongping Yue 1, Maria Boftsi 2,3, Lakmini P. Wasala 1,2, Ngoc Tam Tran 4,5, Keqing Zhang 1, David J. Pintel 1,3, Phillip W. L. Tai 4,5 and Dongsheng Duan 1,6,7,8

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Inverted terminal repeats (ITRs) are the only wild-type components retained in the genome of adeno-associated virus (AAV) vectors. To determine whether ITR modification is a viable approach for AAV vector engineering, we rationally deleted all CpG motifs in the ITR and examined whether CpG elimination compromises AAV-vector production and transduction. Modified ITRs were stable in the plasmid and maintained the CpG-free nature in purified vectors. Replacing the wild-type ITR with the CpG-free ITR did not affect vector genome encapsidation. However, the vector yield was decreased by approximately 3-fold due to reduced vector genome replication. To study the biological potency, we made micro-dystrophin (μDys) AAV vectors carrying either the wild-type ITR or the CpG-free ITR. We delivered the CpG-free μDys vector to one side of the tibialis anterior muscle of dystrophin-null mdx mice and the wild-type μDys vector to the contralateral side. Evaluation at four months after injection showed no difference in the vector genome copy number, microdystrophin expression, and muscle histology and force. Our results suggest that the complete elimination of the CpG motif in the ITR does not affect the biological activity of the AAV vector. CpG-free ITRs could be useful in engineering therapeutic AAV vectors.

Gene Therapy (2022) 29:333–345; https://doi.org/10.1038/s41434-021-00296-0

INTRODUCTION
Adeno-associated virus (AAV) is a helper-dependent parovirus first discovered as contaminating particles in adenovirus stocks [1–3]. AAV contains an ~4.7-kb single-stranded DNA genome [4]. AAV was developed in the late 80s and early 90s as a gene delivery/gene-therapy vector [5]. AAV vectors have been approved by regulatory agencies for treating inherited diseases [8–10]. These include Glybera for treating lipoprotein lipase deficiency, Luxturna (Voretigene neparvovec-rzyl) for treating Leber congenital amaurosis, and Zolgensma (Onasemnogene abeparvovec-xioi) for treating spinal muscular atrophy [11–14]. AAV gene therapy has also resulted in remarkable clinical success in many other genetic diseases such as hemophilia A, hemophilia B, and giant axonal neuropathy [8–10].

Despite the significant advance in AAV gene therapy, there is room for improvement to further enhance efficacy and reduce adverse reactions. A typical AAV vector consists of a protein capsid and a single-stranded vector genome. The vector genome contains an expression cassette and two flanking inverted terminal repeats (ITRs) (Fig. 1A). ITRs are the only genetic components shared by the recombinant vector and wild-type virus. ITRs are essential for AAV genome replication, progeny genome generation, and encapsidation, and conversion of the single-stranded vector genome to the transcription-competent latent form for persistent transgene expression [15–17]. Over the last two decades, most effort in AAV engineering has focused on developing novel capsids and optimizing the expression cassette [9, 18–21]. Few studies have explored ITR engineering because ITR mutagenesis often alters function [22–33]. To study whether synthetic ITRs are suitable for AAV gene therapy, we substituted ~15% of the nucleotides in the AAV2 ITR, the most used ITR in AAV vectors. Specifically, we targeted the CpG motif. The CpG motif refers to a cytosine triphosphate deoxynucleotide followed by a guanine triphosphate deoxynucleotide in a DNA sequence. Unmethylated CpG motifs are immune stimulants [34]. Elimination of the CpG motifs may potentially improve the immunological outcome of AAV gene therapy [35–38]. Our mutagenesis resulted in novel ITRs completely free of CpG motifs. We show that CpG-free ITRs can be used to produce a therapeutic AAV vector. Notably, the biological potency of the vector that has no CpG in the ITR was equivalent to that of the vector carrying the wild-type ITR.

MATERIALS AND METHODS
Cell lines
Fast-growing and highly transfectable human embryonic kidney 293FT cells (ThermoFisher Scientific, Catalog number R70007) were

1Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, MO 65212, USA. 2Pathobiology Area Graduate Program, University of Missouri, Columbia, MO 65212, USA. 3Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65212, USA. 4Horae Gene Therapy Center, UMass Chan Medical School, Worcester, MA 01605, USA. 5Department of Microbiology and Physiological Systems, UMass Chan Medical School, Worcester, MA 01605, USA. 6Department of Biomedical Sciences, College of Veterinary Medicine, University of Missouri, Columbia, MO 65212, USA. 7Department of Neurology, School of Medicine, University of Missouri, Columbia, MO 65212, USA. 8Department of Biomedical, Biological & Chemical Engineering, College of Engineering, University of Missouri, Columbia, MO 65212, USA. 9email: duand@missouri.edu

Received: 30 June 2021 Revised: 10 September 2021 Accepted: 15 September 2021
Published online: 6 October 2021
Fig. 1 Design of the version-1 CpG-free ITR. A Schematic outline of the AAV vector. The expression cassette is composed of a promoter, a transgene, and a polyadenylation (pA) signal. In an AAV vector, the expression cassette is flanked by two ITRs. The 5' and 3' ITRs are highlighted by the dotted boxes. B Two-dimensional drawing of the 5'-ITR in the flip configuration. The sequence is based on Srivastava et al. Journal of Virology 45(2):555-564, 1983. The AAV ITR is divided into four regions, including the A/A' stem (sequence A and its complementary sequence A'), B/B' arm (sequence B, its complementary sequence B', and three intervening thymine nucleotides between sequences B and B'), C/C' arm (sequence C, its complementary sequence C', and three intervening adenine nucleotides between sequences C and C'), and D sequence (underlined). In addition, there is an unpaired thymidine between the B/B' and C/C' arm. Gray letters, nucleotides deleted in the AAV vector. Blue letters, Rep-binding element (RBE). The RBE is a 22-bp sequence (5'-CAGTGACCAGCGAGCACG-3') as reported in Ryan et al. Journal of Virology 70(3):1542-1553, 1996. The core RBE sequence (blue box) consists of a 10-bp sequence (5'-GCAGCGACGC-3'). Green letters, the second Rep-binding element (RBE'). The RBE' is a 10-bp sequence (5'-CCTTGG-3') as reported in Brister and Muzyczka Journal of Virology 74(17): 7762-7771, 2000. Red letters, nucleotides in the version 1 CpG-free ITR. Arrowhead, terminal resolution site (trs). Inset, terminology explanation. C Two-dimensional drawing of the 3'-ITR in the flip configuration. Color coding is the same as in panel B. Purple letters, the core sequence of the trs. The trs core sequence consists of 5'-GT/TGGCC (the forward slash indicates the nicking site). Inset, terminology explanation. RBE, RBE core sequence, and RBE' are as in panel B. Nucleotides modified in the CpG-free ITR are marked in red color. D Alignment of the 3'-ITR of the CpG-free ITR (CpG-free 1) and 3'-ITR of AAV1, 2, 3, 4, 6, and 7. Black dots, nucleotides that are conserved in the 3'-ITR of AAV1, 2, 3, 4, 6, and 7. Asterisks, nucleotides that are conserved in all the listed ITRs. Bold black letters, nucleotides in the AAV ITR that are deleted in the CpG-free ITR. Red letters, nucleotides modified in the CpG-free ITR. Blue box, the GAGC motif in sequence A' and the corresponding GCTC motif in sequence A. Green box, the GAG motif in sequence A' and the corresponding ACTC motif in sequence A. Dash, nucleotides absent in the AAV vector. Underlined black letters, AAV2 ITR Rep-binding element. Underlined italic blue letters, AAV2 ITR terminal resolution site (trs). Underlined italic purple letters, the second Rep-binding element (RBE') of the AAV2 ITR. Arrows, nucleotides in B/B' and C/C' arms that are not detected in the corresponding position of the ITR from AAV1, 2, 3, 4, 6, and 7.

Experimental mice
All animal experiments were approved by the Animal Care and Use Committee of the University of Missouri and were in accordance with NIH guidelines. All animal experiments were conducted at the University of Missouri. Dystrophin-deficient mdx mice (Stock number 001801) were originally purchased from The Jackson Laboratory (Bar Harbor, ME). Experimental mice were generated in-house in a specific-pathogen-free barrier facility at the University of Missouri using founders from The Jackson Laboratory. The genotype of the mice was confirmed using our published protocol [39]. All mice were maintained in a specific-pathogen-free animal facility on a 12-hour light (25 lux):12-hour dark cycle with access to PicoLab rodent diet 20 #5053 (www.labdiet.com) and autoclaved municipal tap water ad libitum. The room temperature and relative humidity were maintained at 68 ± 2°F and 50 ± 20%, respectively. All animals were observed daily for their general condition and well-being. All mice had a unique identification number (ear tag). The sample size was determined by the power analysis and consistent with our previous experience and common practice of the field. No samples or animals were excluded from the analysis.

Generation of the CpG-free ITR
The CpG-free ITRs were designed based on the wild-type ITRs of AAV2. Two different versions of the CpG-free ITR were designed. In the first version (CpG-free 1), the 5'-end CpG-free ITR was designed by replacing guanine in the first CpG motif in the A sequence of the ITR with thymine, guanine in the remaining three CpG motifs in the A sequence of the ITR with adenine, guanine in the first CpG motif and its immediate downstream cytosine in the C sequence of the ITR with adenine and guanine, guanine in the second CpG motif and its immediate downstream guanine in the C sequence of the ITR with cytosine and thymine, guanine in the first CpG motif in the B sequence of the ITR with adenine, and guanine in the second CpG motif in the B sequence of the ITR with cytosine. Corresponding bases in the A', B', and C' sequences of the 5'-end ITR were modified with complementary bases. In the first version (CpG-free 1), the 3'-end CpG-free
ITR was designed by replacing guanine in the first CpG motif in the A sequence of the ITR with thymine, guanine in the remaining three CpG motifs in the A sequence of the ITR with adenine, guanine in the first CpG motif in the B sequence of the ITR with adenine, cytosine and guanine in the second CpG motif of the B arm with guanine and cytosine, respectively, guanine in the first CpG motif and its immediate downstream cytosine in the C sequence of the ITR with adenine and guanine, and guanine in the second CpG motif and its immediate downstream guanine in the C sequence of the ITR with cytosine and thymine. Corresponding bases in the A′, B′, and C′ sequences of the 5′-end ITR are modified with complementary bases.

The second version (CpG-free 2) is different from the first version in B, B′, C, and C′ sequences. Specifically, a guanine-to-thymine substitution in the first CpG motif of the C sequence of the 5′-end ITR and a cytosine-to-guanine substitution in the immediate downstream cytosine in the C sequence. A cytosine-to-adenine substitution in the second CpG motif of the C sequence of the 5′-end ITR and a guanine-to-adenine substitution in the guanine immediately downstream the second CpG motif in the C sequence. A guanine-to-thymine substitution in the first CpG motif of the B sequence of the 5′-end ITR. A cytosine-to-guanine substitution in the second CpG motif in the B sequence of the 5′-end ITR. Corresponding bases in the C′ and B′ sequences of the 5′-end ITR were modified with complementary bases. In the 3′-end ITR of the second version CpG-free ITR, there was a guanine-to-thymine substitution in the first CpG motif of the A sequence of the ITR with adenine and cytosine, respectively, guanine in the second CpG motif of the C sequence and a cytosine-to-guanine substitution in the immediate downstream cytosine in the C sequence. Further, there was a cytosine-to-adenine substitution in the second CpG motif of the C sequence of the 5′-end ITR and a guanine-to-adenine substitution in the guanine downstream of the second CpG motif of the C sequence. Corresponding bases in the B′ and C′ sequences of the 3′-end ITR were modified with complementary bases.

The CpG-free ITRs were synthesized by GenScript (Piscataway, NJ).

**Micro-dystrophin (μDys) expression cassette**

The codon-optimized human μDys gene contains the N-terminal domain, hinge 1, spectrin-like repeats 1, 16, 17, and 24, hinge 4, the cysteine-rich domain, and the syntrophin/dystrobrevin-binding site of human dystrophin [40]. Microdystrophin expression is regulated by the human elongation factor 1-α (EF1-α) core promoter, the mouse cytomegalovirus enhancer, and a synthetic intron from pCpGfree (Invivogen, San Diego, CA, USA), and a synthetic polyadenylation site from pGL3-Basic (Promega, Madison, WI, USA).

**LacZ expression cassette**

The LacZ expression cassette consists of the EF1-α core promoter, the mouse cytomegalovirus enhancer, a synthetic intron, and the LacZ cDNA from the pCpGfree-LacZ (Invivogen, San Diego, CA, USA), and a SV40 late polyadenylation signal from pGL3-Basic (Promega, Madison, WI, USA).

**Vector plasmids**

Two cis-plasmids were used to make the μDys vectors. They carried the exact same μDys expression cassette as described above. One cis-plasmid contained the first version of the CpG-free ITRs and was called pXP15. The other cis-plasmid contained the wild-type ITRs and was called pXP24. To determine the stability of the CpG-free ITR in the vector plasmid, pXP15 was inoculated onto a Luria broth agar plate and a single clone was amplified in the lysogeny broth for plasmid preparation. The resulting plasmid was inoculated onto a Luria broth agar plate to start another round of passage. After six rounds of passage, the ITR of the plasmid was sequenced as described below and used for AAV-vector production.

Two cis-plasmids were used to make the LacZ vectors. They carried the exact same LacZ expression cassette as described above. One cis-plasmid contained the second version of the CpG-free ITRs and was called pXP158. The other cis-plasmid contained the wild-type ITRs and was called pXP159.

**Recombinant AAV production, purification, and titration**

AAV vectors were packaged in Y731F tyrosine mutant AAV9 (a kind gift of Dr. Arun Srivastava, University of Florida) [41, 42], and vector stocks were produced using the transient transfection method according to our published protocol [43, 44]. AAV vectors were purified through two rounds of isopycnic cesium chloride ultracentrifugation followed by three changes of HEPS buffer at 4 °C for 48 h [43]. Viral titer was determined by quantitative PCR using the Fast SYBR Green Master Mix kit (Applied Biosystems, Foster City, CA) in an ABI 7900 HT PCR machine. The pair of primers was designed for the mouse cytomegalovirus enhancer region. The forward primer is 5′-ACAATAAGGTCAATGGGACTTTCGTT AAGCG and the reverse primer is 5′-CAATGTGGGACTTTCGTTAGTCC. Sequencing the ITR in the cis-plasmid

Due to the complicated secondary structure and high GC content, the DNA was first amplified with the GE healthcare Illuira TemplPhi Sequence Resolver Kit (GE healthcare life sciences, Code # 28-9035-29). The amplified product was then subjected to Sanger sequencing using the primer 5′-GATGTCGCTCAAGGGCGATTGA-3′ for the 5′-end ITR and the primer 5′-TTATGCTTCGGCTGCTAGTATG-3′ for the 3′-end ITR.

**Sequencing the ITR in the AAV virus**

Vector genomes were isolated by phenol:chloroform:isoamyl alcohol (Invitrogen) extraction and EtOH precipitation described in Tran et al. [45]. Briefly, SMRT sequencing libraries were built with indexed SMRTbell adapters using the Express Template Prep Kit 2.0 (End-Repair-A-tailing) (PN 100-938-900). Libraries were run on a Sequel II with 15-hour collections. The resultant subreads were processed with recalladapters (https://github.com/PacificBiosciences/recalladapters), and ccs (smrtlink/7.0.1.66975) was run with minimum thresholds, −min-snr = 2.00 and −min-passes = 0.5. Reads were then demultiplexed, mapped to reference-vector genomes, and displayed on IGV [46].

**AAV capsid dot blot**

Purified AAV vectors were loaded on the nitrocellulose membrane. After air-dry, the membrane was blocked with 1X Tris-buffered saline containing 0.1% Tween-20 (1X TBST) and 5% nonfat milk. After a wash in TBST, the membrane was probed with the mouse monoclonal AAV9 antibody (1:100 in 5% nonfat milk 1X TBST; clone ADK9, PROGEN, Heidelberg, Germany, Cat log number 651162) at room temperature for 1 h. The membrane was then washed in 1X TBST three times (five minutes each). After washing, the membrane was probed with the goat anti-mouse IgG antibody (1:1,000 in 1X TBST; Santa Cruz, Dallas, TX) at room temperature for 1 h. After three washes in 1X TBST, the signal was detected using the Clarity Western ECL substrate (Bio-Rad, Hercules, CA) and visualized using the Li-COR Odyssey imaging system (LI-COR Biotechnology). Densitometry quantification was performed using the Li-COR Image Studio version 5.0.21 software (https://www.licor.com).

**Electron microscopy**

AAV particles were examined by transmission electron microscopy as described previously [47]. Specifically, purified and dialyzed AAV virus was diluted to 1–3 × 10^11 vg/ml with ultra-pure water and then placed on a 200-mesh glow-discharge carbon-coated copper grid for five minutes. After four to five rounds of gentle washing in ultrapure water, the virus was stained with 2% NANO-W™ Nanopros, Yaphank, NY, USA) for 5 min. Viral particles were visualized using a JEOL JEM-1400 transmission electron microscope.

**Southern blot evaluation of vector genome replication**

Southern blot was performed blindly. The plasmids used for transfection were assigned a number. The investigator was informed with the plasmid number but was unaware which was the wild-type plasmid and which was the CpG-free plasmid. 293FT cells were grown on 60-mm cell culture plates, until they reached 50–70% confluency. Cells were cotransfected with pHelper, pRepCap, and the cis-plasmid using the Express Template Prep Kit 2.0 (End-Repair-A-tailing) (PN 100-938-900). Libraries were run on a Sequel II with 15-hour collections. The resultant subreads were processed with recalladapters (https://github.com/PacificBiosciences/recalladapters), and ccs (smrtlink/7.0.1.66975) was run with minimum thresholds, −min-snr = 2.00 and −min-passes = 0.5. Reads were then demultiplexed, mapped to reference-vector genomes, and displayed on IGV [46].
pXP15 or pXP24, and it contained the full-length vector genome. The intensity of the replication from AAV genome bands was quantified with the Fujifilm Multi Gauge software (version 3.0).

AAV administration

In all, 2.8 \times 10^{11} \text{ vg/particles/muscle} (in 50 \text{ µl of HEPS buffer}) of the AAV vector was injected into the TA muscle of six 10-m-old female mdx mice using a Hamilton syringe. One side of the TA muscle received the wild-type vector, and the contralateral side of the same mouse received the CpG-free vector. The injection side was randomly assigned by coin tossing.

Muscle AAV vector genome copy-number quantification

Genomic DNA was extracted from OCT-embedded frozen tissue samples. DNA concentration was measured with Nanodrop One \textsuperscript{\textregistered} Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Quantitative TaqMan PCR assays were performed using the PrimerTime Gene Expression Master Mix (Integrated DNA Technologies, IDT, IA) in an ABI 7090 HT qPCR machine (Applied Biosystems, Foster City, CA). The qPCR primers and probe were designed from the human \(\alpha\)EF-1 \textsuperscript{\textregistered} promoter region. The forward primer is \(5'-\text{GGCTTGGTAACTGGGAAA-3'}\), the reverse primer is \(5'-\text{GTTCACAGAGATC-3'}\), and the probe is \(5'-\text{ATGGTGTTACGTCCTACCTT-3'}\). The qPCR reaction was carried out under the following conditions: 10 min at 95 \(^\circ\text{C}\), followed by 40 cycles: 15 s at 95 \(^\circ\text{C}\) and 1 min at 60 \(^\circ\text{C}\). The threshold-cycle (Ct) value of each reaction was converted to the vector DNA concentration was measured with NanoDrop One C Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA).

Injection side was randomly assigned by coin tossing.

Muscle AAV vector genome copy-number quantification

Genomic DNA was extracted from OCT-embedded frozen tissue samples. DNA concentration was measured with NanoDrop One \textsuperscript{\textregistered} Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Quantitative TaqMan PCR assays were performed using the PrimerTime Gene Expression Master Mix (Integrated DNA technologies, IDT, IA) in an ABI 7900 HT qPCR machine (Applied Biosystems, Foster City, CA). The qPCR primers and probe were designed from the human \(\alpha\)EF-1 \textsuperscript{\textregistered} promoter region. The forward primer is \(5'-\text{GGCTTGGTAACTGGGAAA-3'}\), the reverse primer is \(5'-\text{GTTCACAGAGATC-3'}\), and the probe is \(5'-\text{ATGGTGTTACGTCCTACCTT-3'}\). The qPCR reaction was carried out under the following conditions: 10 min at 95 \(^\circ\text{C}\), followed by 40 cycles: 15 s at 95 \(^\circ\text{C}\) and 1 min at 60 \(^\circ\text{C}\). The threshold-cycle (Ct) value of each reaction was converted to the vector DNA concentration was measured with NanoDrop One C Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Quantitative TaqMan PCR assays were performed using the PrimerTime Gene Expression Master Mix (Integrated DNA technologies, IDT, IA) in an ABI 7900 HT qPCR machine (Applied Biosystems, Foster City, CA). The qPCR primers and probe were designed from the human \(\alpha\)EF-1 \textsuperscript{\textregistered} promoter region. The forward primer is \(5'-\text{GGCTTGGTAACTGGGAAA-3'}\), the reverse primer is \(5'-\text{GTTCACAGAGATC-3'}\), and the probe is \(5'-\text{ATGGTGTTACGTCCTACCTT-3'}\). The qPCR reaction was carried out under the following conditions: 10 min at 95 \(^\circ\text{C}\), followed by 40 cycles: 15 s at 95 \(^\circ\text{C}\) and 1 min at 60 \(^\circ\text{C}\). The threshold-cycle (Ct) value of each reaction was converted to the vector DNA concentration was measured with NanoDrop One C Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA).

Morphological analysis

Four months after AAV injection, we performed the terminal TA muscle function assay (see below). After the functional assay, mice were euthanized according to the protocols approved by the University of Missouri Animal Care and Use Committee. The TA muscle was carefully dissected and cut into two pieces. One piece was snap-frozen in liquid nitrogen. The other piece was embedded in liquid nitrogen-cooled isopentane in the optimal cutting-temperature compound (Sakura Finetek Inc., Torrance, CA). Ten-micron cryosections were used for staining. General muscle histopathology was revealed with hematoxylin and eosin (H&E) staining. Dysphosphorylation was evaluated by immunofluorescence staining using a mouse anti-human dystrophin spectrin-like repeat 16 (1:200; the Fiji imaging software (https://fiji.sc) and MyoVision software, respectively.

Western blot and quantification

Whole-muscle lysates were electrophoresed on a 6\% sodium dodecyl sulfate–polyacrylamide gel. Western blot was carried out on an iBind Flex western device (Invitrogen). The \(\mu\)Dys protein was detected with a mouse monoclonal antibody against dystrophin spectrin-like repeat 16 (1:200; a gift from Dr. Glenn Morris, The Roper Jones and Agnes Hunt Orthopedic Hospital, Oswestry, Shropshire, UK) and the Alexa Fluor-594-conjugated goat anti-mouse IgG secondary antibody (#156, 1:100, Invitrogen). Slides were viewed at the identical exposure setting using a Nikon E800 fluorescence microscope. Images were taken with a QImage Retiga 1300 camera. Centrally nuclelated myofibers were determined from digitalized H&E-stained images using the Fiji imaging software (https://fiji.sc) [49]. Percentages of dystrophin-positive cells and the myofiber size were quantified from digitalized dystrophin immunostaining images using the Fiji imaging software (https://fiji.sc) and MyoVision software, respectively.

Skeletal muscle function assay

The muscle-force assay was performed blindly. Specifically, animals were identified by the ear tag rather than the vector injected into the TA muscle. The function of the TA muscle was evaluated in situ according to our published protocols [50, 51]. Specifically, the twitch force, tetanic force, and eccentric contraction profile were measured. Experimental mice were anesthetized via intraperitoneal injection of a cocktail containing 25 mg/ml ketamine, 2.5 mg/ml xylazine, and 0.5 mg/ml acepromazine at 2.5 \text{ µg/g body weight. The TA muscle and the sciatic nerve were carefully exposed. The mouse was then transferred to a custom-designed thermocontrolled footplate platform [52]. Subsequently, forces were measured in situ with a 30SC-LR dual-mode servomotor transducer (Aurora Scientific, Inc., Aurora, ON, Canada). According to our published protocols [50, 52], the absolute twitch force, the optimal maximal isometric tetanic force, and the force drop through 10 repetitive cycles of eccentric contraction were determined. Data acquisition and analysis were performed with the Dynamic Muscle Control and Analysis software (Aurora Scientific Inc.). The specific muscle force was calculated by dividing the absolute muscle force with the muscle cross-sectional area (CSA). Muscle CSA was calculated according to the following equation, CSA = (mass, in g)/(limuscle density, in g/cm\(^3\)) \times (length ratio) \times (optimal muscle length, in cm). Muscle density is 1.06 g/cm\(^3\) [53]. The length ratio refers to the ratio of the optimal fiber length to the optimal muscle length. The length ratio for the TA muscle is 0.6 [54].

RESULTS

Rational design of the CpG-free ITR

The 145-nucleotide AAV2 ITR can be divided into seven segments, the A, A\(^{\prime}\), B, B\(^{\prime}\), C, C\(^{\prime}\), and D sequences (Fig. 1) [4]. Sequences A, B, and C are inversely complementary to sequences A\(^{\prime}\), B\(^{\prime}\), and C\(^{\prime}\), respectively. The pairing of sequences B/B\(^{\prime}\) and C/C\(^{\prime}\) forms the two arms of the T-shaped hairpin structure of the ITR. The pairing of sequences A and A\(^{\prime}\) forms the stem of the T-shaped ITR. The 20 nucleotide-long D sequence is maintained as the single-stranded DNA in an AAV virion (Fig. 1B, C) [15, 55].

The ITR contains three sequence elements that are essential for its function. These include the Rep-binding element (RBE), the second Rep binding element (RBE\(^{\prime}\)), and the terminal resolution site (trs) [22–24]. The RBE is located in the A/A\(^{\prime}\) stem and consists of a 22-bp sequence (5'-CATGAGGGGACGACGCG-3') (Fig. 1B, C) [22]. Within the RBE, there is a 10-bp core sequence (5'-CCTTGGACGAGG-3') (Fig. 1B, C). Dinucleotide transversion mutations in the core sequence reduce the Rep-binding affinity by at least 10-fold [22]. The tetranucleotide repeat GAGY (RCTC in the complementary strand) is the consensus Rep-binding motif in the RBE (Fig. 1D) [15]. The RBE is located at the tip of either the B or C arm. It consists of a 5-nucleotide sequence (5'-CTTGG-3') (Fig. 18–D) [24]. The trs is a 7-nucleotide sequence (5'-GGTGGCC-3') located at the junction of the A/A' stem and the D sequence (Fig. 1D) [23].

To engineer a synthetic ITR that may potentially offer a therapeutic advantage, we opted to target Cpg motifs because it is well-established that Cpg motifs induce the innate immune response (reviewed in [56, 57]). There are 16 Cpg motifs in one ITR (Fig. 1B–D). They are located in the A/A' stem (four in sequence A, four in sequence A'), B/B' arm (two in sequence B, two in sequence B'), and C/C' arm (two in sequence C, two in sequence C'). Of the three essential ITR elements, only the RBE contains Cpg motifs (six in the core sequence and eight in total). There are no Cpg motifs in the RBE' and trs.

The first version of the CpG-free ITR (Cpg-free 1) was engineered based on the following principles. To remove the Cpg motif in the A/A' stem, we first decided to change GAGC (GCTC in the complementary strand) in the core sequence of the RBE to GAGT (ACTC in the complementary strand) (Fig. 18–D). This modification allowed us to remove all 6 Cpg motifs (three in sequence A and three in sequence A') in the core sequence of the RBE (Fig. 1B, C). Our rationales were (1) we opted to introduce a
Fig. 2  Confirmation of the CpG-free ITR in the cis-plasmid. A Nucleotide sequence of the wild-type ITR at the 5′-end of the vector genome (5′-ITR). A, A′, B, B′, C, C′, and D sequences are marked by boxes. The CpG motifs are underlined. B Nucleotide sequence of the CpG-free ITR at the 5′-end of the vector genome. The mutated nucleotides are highlighted in red. C Chromatogram view of the sequencing result of the 5′-ITR of the cis-plasmid that carries the CpG-free ITR. Arrowhead, nucleotides in the CpG-free ITR that are different from those of the wild-type ITR. D Nucleotide sequence of the wild-type ITR at the 3′-end of the vector genome (3′-ITR). A, A′, B, B′, C, C′, and D sequences are marked by boxes. The CpG motifs are underlined. E Nucleotide sequence of the CpG-free ITR at the 3′-end of the vector genome. The mutated nucleotides are highlighted in red. F Chromatogram view of the sequencing result of the 3′-ITR of the cis-plasmid that carries the CpG-free ITR. Arrowhead, nucleotides in the CpG-free ITR that are different from those of the wild-type ITR.
transition (C-T) instead of a transversion mutation because transversion mutation negatively impacts Rep-mediated ITR nicking [22], (2) our modification did not alter the GAGY (RCTC in the complementary strand) consensus motif [15], and (3) the GAGT (ACTC in the complementary strand) motif already exists in the RBE of other AAV serotypes such as AAV3 and AAV4 (Fig. 1D) [58–60]. The last two CpG motifs in the A stem are inside a GC C motif immediately downstream of the last GAGC in sequence A’ and upstream of the first GCTC in sequence A. We changed GGC C in sequence A’ to the more conserved GAGC motif. Accordingly, GC C in complementary sequence A became GCTC. Since the B/B’ arm and C/C’ arm do not contain elements essential for ITR function, the CpG motifs in these segments were arbitrarily mutated as long as the T-shaped tertiary structure was preserved [61, 62] (Fig. 1D).

The GC content is another important consideration in designing the CpG-free ITR. The GC contents of the AAV1, 2, 3, 4, 6, and 7 ITRs are 68.53%, 69.66%, 65.07%, 64.38%, 67.13%, and 68.97%, respectively [15, 63, 64]. In the engineered CpG-free ITR, the GC content was 60% (Fig. 1). Although this is lower than that of the

---

**Fig. 3** SMRT sequencing evaluation of the ITR sequence in the purified AAV vector. A Integrative genomic viewer (IGV) display of SMRT reads from the 5’-ITR of vectors that carry the wild-type ITR. B IGV display of SMRT reads from the 3’-ITR of vectors that carry the wild-type ITR. C IGV display of SMRT reads from the 5’-ITR of vectors that carry the CpG-free ITR. D IGV display of SMRT reads from the 3’-ITR of vectors that carry the CpG-free ITR. Each line represents a single read. Reads mapping on the plus strand are in red, and reads mapping on the minus strand are in blue. The references reflect the flip configurations. Base mismatches are shown as colored bases, inserts are indicated in purple ls, and deletions are shown as dashes. Reads are displayed with soft clipping on to show the A-domain terminus not included in the original reference. Flip-oriented reads are notable by mismatches from the flip reference at positions demarcated by black arrows. Flip reads are indicated in cyan, flip reads in magenta.

**Fig. 4** Quantitative evaluation of AAV production. A Quantification of the vector genome yield. *p < 0.05. B Representative AAV capsid dot blot. The blot was probed with an antibody that only recognizes intact AAV9 particles. C Quantification of the vector capsid yield. **p < 0.01. D Representative transmission electron microscopy images of the CpG-free ITR vector. Arrow, a fully packaged AAV particle. Arrowhead, an empty AAV particle. E Representative transmission electron microscopy images of the wild-type ITR vector. Arrow, a fully packaged AAV particle. Arrowhead, an empty AAV particle. F Quantification of empty particles. Each data point represents the quantification result from one field at the 25,000x magnification. For the wild-type ITR vector, a total of 48 fields were quantified. For the CpG-free ITR vector, a total of 25 fields were quantified.
wild-type ITRs, it is still considerably higher than the 40.9% average GC content of the human genome [65].

Production of the CpG-free vector
To determine whether CpG depletion affects AAV production, we made two AAV μDys vectors that carry the identical expression cassette, but differ in the ITR. One has the wild-type ITR (we named it the wild-type vector in this paper), and the other has the CpG-free ITR (we named it the CpG-free vector in this paper).

To determine the stability of the CpG-free ITR in the plasmid, the ITR was sequenced after the plasmid was passaged for six rounds in bacteria. All mutated nucleotides in the CpG-free ITR were preserved (Fig. 2). Multiple rounds of amplification in bacteria did not alter the ITR sequence. After confirming the intactness of the ITR, we prepared and purified the wild-type vector and the CpG-free vector side-by-side using transient transfection, the most used method for AAV production. The ITR of the purified vectors was sequenced using the single-molecule real-time (SMRT) sequencing approach [45]. In the wild-type vector, all 32 CpG motifs were detected in ITRs. No CpG motif was found in the ITR of the CpG-free vector (Fig. 3).

Two different methods were used to quantify vector production efficiency. Vector genome yield was determined by quantitative PCR. The yields of the wild-type and CpG-free vector were 1.53 ± 0.36 × 10^5 vg/cell and 4.06 ± 1.06 × 10^4 vg/cell, respectively (Fig. 4A). The vector genome yield of the CpG-free vector was reduced by approximately 3-fold. The relative caspid yield of the vector was determined by dot blot with an antibody that recognizes the intact AAV9 particle (Fig. 4B, Supplementary Fig. 1). The caspid yield of the CpG-free vector was about 3-fold lower than that of the wild-type vector (Fig. 4C). The ratio of the genome yield and caspid yield was ~1:1 for both CpG-free and wild-type vectors. This suggests that every vector genome was encapsidated in a capsid in the purified AAV stock. Sufficient empty capsids were produced to package all the single-stranded vector genomes in both CpG-free and wild-type vector. Removal of the CpG motifs did not compromise capsid formation.

Next, we quantified the percentage of empty particles using transmission electron microscopy (Fig. 4D–F). The packaged AAV virions showed homogeneous electron density, while empty particles had a dark center (Fig. 4D, E). On average, there were 11.9 ± 1.2% and 11.8 ± 1.5% of empty particles in the wild-type and CpG-free vector stock, respectively (Fig. 4F).

Reduced vector-genome replication leads to the low yield
To investigate the mechanism underlying the reduced yield of the CpG-free vector, we examined vector plasmid replication. The cis-plasmid used for vector production was cotransfected with the adenoviral helper plasmid and the AAV9 tyrosine mutant Rep–Cap helper plasmid to 293FT cells. The low-molecular-weight Hirt DNA was extracted at different time points post transfection, and vector genome replication was evaluated by Southern blot (Fig. 5). The dimer replicative form (dRF) of the vector genome was readily visible at 48 h post transfection. However, the monomer replicative form (mRF) of the vector genome comigrated with a plasmid DNA band. Hirt DNA was digested with DpnI to remove the plasmid DNA. The disappearance of the larger plasmid DNA bands suggested complete removal of the plasmid DNA. Overall, the mRF and dRF generated from the CpG-free vector genome were ~3-fold lower than those from the wild-type vector genome (Fig. 5, Supplementary Fig. 2).

Optimization of the B/B′ and C/C′ arms in the CpG-free ITR
In the CpG-free vector described above, CpG motifs in B/B′ and C/C′ arms were eliminated by arbitrarily mutating 14 of 44 nucleotides in these regions. As a result, 8 mutated nucleotides were not detected in the corresponding positions of AAV1, 2, 3, 4, 6, and 7 (Fig. 1D). To determine whether these differences contributed to the low yield (Fig. 4A, C), we designed the second version of the CpG-free ITR (CpG-free-2) by modifying the B/B′ and C/C′ arms (Fig. 6). In addition to maintaining the tertiary structure of the T-shaped configuration of the ITR [61], we intentionally selected substitution nucleotides from those that are highly conserved in other AAV serotypes or presented at least once in the corresponding positions of AAV1, 2, 3, 4, 6, and 7 (Fig. 6A). In the

Fig. 5 Evaluation of vector-genome replication during AAV production. Human 293FT cells were cotransfected with the cis-plasmid (pXP15 or pXP24), pHpedia, and pRepCap. Hirt DNA was harvested at the indicated time points and analyzed by Southern blot using homologous genomic probes. A Representative Southern blot using Hirt DNA harvested at 2 and 48 h post-transfection (lanes 2–7). Marker (lane 1), a 5-kb band from Sacl and Sall double digestion of pXP15 or pXP24 to indicate the full-length vector genome. B A biological replicate Southern blot using Hirt DNA harvested at 2, 24, and 48 h post transfection (lanes 2–9). Lane 1 is undigested plasmid pXP24. Arrow, monomer-replication form (mRF), and dimer-replication form (dRF). White arrowhead, the cis-plasmid bands that are larger than the size of mRF. DpnI digestion completely removed these bands. Black arrowhead, the cis-plasmid band that comigrated with mRF.
new version CpG-free ITR (CpG-free-2) (Supplementary Fig. 3), all nucleotides in the B/B′ and C/C′ arms appeared at least once in the corresponding positions of AAV1, 2, 3, 4, 6, and 7 (Fig. 6A). In total, 24 nucleotides in the B/B′ arm and C/C′ arm of the new version CpG-free ITR were identical to those in AAV1, 2, 3, 4, 6, and 7, while only 20 nucleotides were identical in the old version. LacZ AAV vectors were produced side-by-side using the cis-plasmid carrying either the new CpG-free ITR or the wild-type ITR. AAV vectors from the culture medium and cell lysate were separately purified, and the yield was compared by TaqMan PCR (Fig. 6B, C). Despite the improved homology with the wild-type ITR, the yield of the new CpG-free vector remained approximately 3-fold lower than that of the wild-type vector (Fig. 6B, C).

**Evaluation of the biological potency of the CpG-free vector in a mouse model of Duchenne muscular dystrophy (DMD)**

To determine whether CpG depletion affected in vivo transduction of the AAV vector, we generated μDys AAV vectors carrying either the wild-type ITR (wild-type μDys vector) or the CpG-free ITR (CpG-free μDys vector). We then tested these vectors in parallel in the mdx mouse model of DMD, a disease caused by dystrophin deficiency [66]. Specifically, the wild-type and CpG-free μDys vectors were injected into a different side of the tibialis anterior (TA) muscle of the same mdx mouse. Four months after AAV injection, we quantified μDys expression, the vector genome copy number, and the histological and physiologica rescue of the muscle disease by μDys (Figs. 7–9, Supplementary Fig. 4).

Using immunofluorescence staining, we observed robust sarcolemma μDys expression throughout the entire muscle in both wild-type and CpG-free μDys vector-injected TA muscles (Fig. 7A, Supplementary Fig. 4). There was no statistical difference in the percentage of μDys-positive myofibers (Fig. 7B). Using western blot, we detected similar levels of μDys in wild-type and CpG-free μDys vector-injected TA muscles (Fig. 7C, D). Quantification of the vector genome copy number showed no difference between the wild-type and CpG-free μDys vectors (Fig. 7E).

To determine histological rescue, we quantified centronucleation and myofiber-size distribution (Fig. 8). The former reveals degeneration/regeneration, and the latter reveals muscle hypertrophy/atrophy. The wild-type μDys vector-injected muscle contained 54.3 ± 2.1% centrally nucleated myofibers. The CpG-free μDys vector-injected muscle had 59.1 ± 3.1% centrally nucleated myofibers (Fig. 8A). The myofiber size was measured by the minimum Feret diameter (Fig. 8B). Throughout the entire range (from 10 to 56 μm), there was no difference between the wild-type and CpG-free μDys vector-injected muscles.

To determine physiologica rescue, we quantified muscle weight, cross-sectional area, absolute and specific twitch force, absolute and specific tetanic force, force-frequency relationship, and force drop following eccentric-contraction challenge (Fig. 9). In all parameters we examined, there was no statistically significant difference between the wild-type and CpG-free μDys vector-treated muscles (Fig. 9).
Fig. 7 Evaluation of microdystrophin expression. A Representative dystrophin immunofluorescence staining and HE staining micrographs from the tibialis anterior (TA) muscle of dystrophin-null mdx mice that were injected with the CpG-free microdystrophin (µDys) vector (left panels) and the wild-type µDys vector (middle panels). The TA muscle from an age- and sex-matched uninjected mdx mouse was included as the control (right panels). Scale bar applies to all images. B Quantification of dystrophin positive myofibers. C Western blot evaluation of 143-kD µDys from TA muscles of CpG-free µDys vector and wild-type µDys vector-treated mice. Alpha-tubulin (50 kD) is used as the loading control. D Quantification of western blot µDys expression. E Quantification of the vector genome copy number in the TA muscle.
The percentage of myofibers that contained centrally localized nuclei in the mdx muscle treated with the CpG-free μDys vector and the wild-type μDys vector. B The distribution of the percentage of myofibers at different minimal Feret diameters in 616 myofibers from the CpG-free μDys vector-treated muscle (n = 6 muscles, 80–131 myofibers per muscle) and 712 myofibers from the wild-type μDys vector-treated muscle (n = 6 muscles, 87–135 myofibers per muscle).

DISCUSSION

In this study, we developed the CpG-free ITR for AAV gene therapy. We showed that all CpG motifs could be removed from the ITR without compromising the biological potency of the vector. However, removal of all CpG motifs reduced the vector yield.

The only wild-type viral sequence required in cis in the AAV vector is the ITR [9, 10, 67]. AAV vector production depends on the successful rescue of the vector genome from the double-stranded proviral plasmid (cis-plasmid), the subsequent replication of the vector genome through a self-priming mechanism, and the displacement and encapsidation of the single-stranded genome into a preassembled capsid [68–70]. The ITR is essential for all of these processes. Specifically, the large Rep proteins bind to the RBE and RBE’ elements [24]. These interactions position the large Rep proteins to make a sequence- and strand-specific nick at the trs. The free 3’ OH group created by this cleavage serves as the replication primer to synthesize the secondary ITR. Further replication leads to a new complementary strand and the displacement of the original complementary strand. The displaced strand (vector genome) is pumped through a replication leads to a new complementary strand and the replication primer to synthesize the secondary ITR. Further

Single-nucleotide transversion mutation of the trs nearly abolishes ITR nicking by the large Rep proteins [23]. Truncation of the B and C arm leads to an 8-fold decrease in AAV replication [27]. Deleting the trs in one ITR completely prevents AAV genome replication from the mutated ITR [31]. Deletion and/or substitution of the D sequence renders AAV to package only the plus- or the minus-strand genome, instead of both [28, 29, 32]. Defective ITR has also been associated with the packaging of nonvector sequences [25, 33, 72, 77]. Collectively, these studies reveal the challenges in ITR engineering.

To determine whether ITR modification is a viable approach for AAV vector engineering, we made targeted mutations at the CpG sites of the ITR. Our rationales were (1) CpG motifs are dispersed throughout the palindromic hairpin structure of the ITR (A/A’, B/B’, and C/C’). CpG mutagenesis will give us hints about the editability of most regions of the ITR; and (2) CpG reduction might offer therapeutic benefits because this has been shown to improve transgene expression in both preclinical studies and clinical trials [35–38].

To engineer a CpG-free ITR that can support AAV production and transduction, we carefully designed the mutagenesis strategy. To avoid altering the tertiary structure, we made matching changes in the corresponding locations of the palindromic complementary strand. To maximally preserve the Rep-binding activity, we only conducted transition mutations in the core sequence of the RBE, because transversion mutations have been shown to compromise Rep binding and nicking [22–24]. In addition, we guided our design with the ITR sequences from other AAV variants. The ITR sequence from more than 12 AAV variants (AAV1, 2, 3, 3b, 4, 5, 6, 7, and avian-VR865, avian-DA1, bovine, and serpentine AAV) has been published [4, 58, 59, 63, 78–84]. However, only a subset of these ITRs (AAV1, 2, 3, 4, 6, and 7) have been used to make the recombinant vector using AAV2 Rep proteins [63, 64, 85, 86]. The alignment of these ITR sequences provided a road map for us to make an educated decision on nucleotide selection for mutagenesis (Figs. 1D and 6A).

Despite the precautions, we found an approximately 3-fold reduction in the yield of the CpG-free vectors (Figs. 4A, C, and 6B, C). We hypothesized that our modification might have somehow weakened the binding of the large Rep proteins. To test this hypothesis, we examined vector-genome replication by Southern blot (Fig. 5). Interestingly, we detected a roughly 3-fold reduction of the replication from vector genome (Fig. 5; Supplementary Fig. 2). This piece of evidence suggests that reduction in large Rep protein-mediated AAV genome replication is likely the primary reason for the low yield.

To package the single-stranded AAV genome into the preassembled capsid may have also reduced vector yield. We believe this likely contributed minimally in our case. First, the reduction of the vector-genome yield was proportional to the reduction of the vector-capsid yield (Fig. 4A, C). Second, we detected similar amounts of empty particles (~12%) in the wild-type and CpG-free vectors (Fig. 4F). The large Rep proteins are important for vector-genome rescue and replication 67. However, the small Rep proteins play a more critical role in encapsidation [71, 87]. The current model of AAV packaging suggests that AAV encapsidation is mainly accomplished by small Rep proteins. Specifically, small Rep proteins unwind and import the AAV genome through the 5-fold channel in a 3′–5′ direction. In other words, the 3′-end of the genome is imported first through the pore at the fivefold axis [69, 70]. It is currently unclear how small Rep proteins recognize and interact with the 3′-ITR to initiate the encapsidation process. Our data suggest that CpG elimination likely does not interfere with this process.

Biological competence is an important parameter in vector characterization. To study this, we performed a study in parallel within the same animal. On one side of the muscle, we injected the wild-type vector, and on the contralateral side, we injected the

Fig. 8 Evaluation of centronucleation and myofiber size distribution. A The percentage of myofibers that contained centrally localized nuclei in the mdx muscle treated with the CpG-free μDys vector and the wild-type μDys vector. B The distribution of the percentage of myofibers at different minimal Feret diameters in 616 myofibers from the CpG-free μDys vector-treated muscle (n = 6 muscles, 80–131 myofibers per muscle) and 712 myofibers from the wild-type μDys vector-treated muscle (n = 6 muscles, 87–135 myofibers per muscle).

Gene Therapy (2022) 29:333 – 345

X. Pan et al.
Fig. 9  Quantitative evaluation of the tibialis anterior muscle contractility from mdx mice that were treated with CpG-free and the wild-type μDys vectors. A TA muscle weight. B TA muscle cross-sectional area (CSA). C Absolute twitch force (Pt). D Specific twitch force (sPt). E Absolute tetanic force (Po). F Specific tetanic force (sPo). G Force-frequency relationship. H Eccentric contraction profiles.
CpG-free vector. Surprisingly, we did not detect any difference in the vector-genome copy number and transgene expression (Fig. 6, Supplementary Fig. 4). Notably, both vectors were equally effective in attenuating histological and pathological defects in diseased mice (Figs. 8 and 9).

In summary, our study suggests that the rationally engineered CpG-free ITR can support recombinant AAV-vector production and transduction. Although not directly tested in this study, we speculate that the vector made with the CpG-free ITR is less immunogenic because of the well-established relationship between the CpG motif and the immune response [35, 36, 56, 57, 88, 89]. Future in vitro and in vivo studies will help to establish the immunological advantages of the CpG-free ITR. The low yield of the current version of the CpG-free ITR may limit its use for high-dose systemic gene therapy (such as μDyst treatment of DMD) [90]. However, such a concern may not be an issue for AAV gene therapy that requires a low amount of the vector (such as treating retinal diseases and hearing loss).

REFERENCES

1. Atchison RW, Casto BC, Hammon WM. Adenovirus-associated defective virus particles. Science. 1965;149:754–6.
2. Carter BJ. Adeno-associated virus and the development of adeno-associated virus vectors: a historical perspective. Mol Ther. 2004;10:981–9.
3. Flotte TR, Berns KJ. Adeno-associated virus: a ubiquitous commensal of mammals. Hum Gene Ther. 2005;16:101–7.
4. Srivastava A, Lusby EW, Berns KJ. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol. 1983;45:555–64.
5. Muzyczka N. Use of adeno-associated virus as a general transduction vector for mammalian cells. Curr Top Microbiol Immunol. 1992;158:97–129.
6. Muzyczka N, Berns KJ. AAV’s golden jubilee. Mol Ther. 2015;23:807–8.
7. Hastie E, Samulski RJ. Adeno-associated virus at 50: a golden anniversary of vectors: a historical perspective. Mol Ther. 2005;10:981–8.
8. High KA, Roncarolo MG. Adeno-associated virus at 50: a golden jubilee. Mol Ther. 2015;23:807–8.
9. Wang XS, Ponnazhagan S, Srivastava A. Rescue and replication signals of the adeno-associated virus 2 genome. J Mol Biol. 1995;250:573–80.
10. Lai Y, Duan D. Design of muscle gene therapy expression cassette. In: Duan D, editor. Muscle Gene Therapy, Second edn. Springer Nature Switzerland. 2019:30200–13.
11. Tai PWL. ITRs: The terminal frontier. Hum Gene Ther. 2020;31:143–4.
12. Duan D, Yan Z, Yue Y, Engelhardt JF. Structural analysis of adeno-associated virus transduction intermediates. Virology. 1999;261:8–14.
13. Kottrem MA, Schaffer DV. Engineering adeno-associated viruses for clinical gene therapy. Nat Rev Genet. 2014;15:445–51.
14. Wagner HU, Weber W, Fussenegger M. Synthetic biology: emerging concepts to design and advance adeno-associated viral vectors for gene therapy. Adv Sci. 2021;8:2004018.
15. Lai Y, Duan D. Design of muscle gene therapy expression cassette. In: Duan D, Mendell JR, editors. Muscle Gene Therapy, Second edn. Springer Nature Switzerland. 2019.
16. Choi HY, Yu NK, Baek GC, Bakes J, Seo D, Nam HJ, et al. Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. Mol Brain. 2014;7:17.
17. Ryan JH, Zolotukhin S, Muzyczka N. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. J Virol. 1996;70:1542–53.
18. Brister JR, Muzyczka N. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J Virol. 1999;73:9325–36.
49. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nature methods. 2012;9:676–82.

50. Hakim CH, Li D, Duan D. Monitoring murine skeletal muscle function for muscle gene therapy. Methods Mol Biol. 2011;709:75–89.

51. Hakim CH, Wasala NB, Pan X, Kodippili K, Yue Y, Zhang K, et al. A five-repeat microtubulin alpha gene ameliorated dystrophic phenotype in the severe DBA/2-mdx model of Duchenne muscular dystrophy. Molecular therapy. Meth Clin Dev. 2017;6:216–30.

52. Hakim CH, Wasala NB, Duan D. Evaluation of muscle function of the extensor digitorum longus muscle ex vivo and tibialis anterior muscle in situ in mice. J Vis Exp. 2013;72:e50183.

53. Mendiola J, Keys A. Density and composition of mammalian muscle. Metabolism. 1960;9:184–8.

54. Burkholder TJ, Fingado B, Baron S, Lieber RL. Relationship between muscle fiber types and sizes and muscle architectural properties in the mouse hindlimb. J Morphol. 1994;221:177–90.

55. Berns KI. The unusual properties of the AAV inverted terminal repeat. Hum Gene Ther. 2020;31:518–23.

56. Krieg AM. CpG motifs in bacterial DNA and their immune effects. Annu Rev Immunol. 2002;20:709–60.

57. Ishii KJ, Akira S. Innate immune recognition of, and regulation by, DNA. Trends Immunol. 2006;27:525–32.

58. Muramatsu S, Mizukami H, Young NS, Brown KE. Nucleotide sequencing and length, weight and GC content of the human genome. BMC Res Notes. 2008;1:60.

59. Chejanovsky N, Carter BJ. Mutagenesis of an AUG codon in the adeno-associated virus genome can be excised in vitro and its use to study viral replication. J Virol. 1999;73:161–9.

60. Amiss TJ, McCarty DM, Skulimowski A, Samulski RJ. Identification of adeno-associated virus serotype-specific inverted terminal repeat hairpin DNA. J Virol. 1997;71:8233–33.

61. Amoss TJ, McCarty DM, Skulimowski A, Samulski RJ. Identification and characterization of an adeno-associated virus integration site in CV-1 cells from the African green monkey. J Virol. 2003;77:1904–15.

62. Im DS, Muzycka N. Factors that bind to adeno-associated virus terminal repeats. J Virol. 1989;63:3095–104.

63. Askhtorab H, Srivastava A. Identification of nuclear proteins that specifically interact with adeno-associated virus type 2 inverted terminal repeat hairpin DNA. J Virol. 1989;63:3034–9.

64. Early LF, Conatser LM, Lue VM, Dobbins AL, Li C, Hirsch ML, et al. Adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. J Virol. 1997;71:16823–33.

65. Grimm D, Pandey K, Nakai H, Storm TA, Kay MA. Liver transduction with recombinant adeno-associated virus is primarily restricted by capsid serotype not vector genotype. J Virol. 2006;80:426–39.

66. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. BMC Res Notes. 2012;5:1206.

67. Duan D, Goemans N, Takeda S, Mercuri E, Aartsma-Rus A. Duchenne muscular dystrophy. Nat Rev Dis Primers. 2021;7:131.

68. Samulski RJ, Muzycka N. AAV-mediated gene therapy for research and therapeutic purposes. Annu Rev Virol. 2014;1:427–51.

69. Ling N, Wang Y, Lu Y, Wang L, Jayandharan GR, Aslanid G, et al. The adeno-associated virus genome packaging puzzle. J Mol Genet Med. 2015;9:175.

70. Bennett A, Mietzsch M, Agbandje-Mckenna M. Understanding capsid assembly and genome packaging for adeno-associated viruses. Fut Virol. 2017;12:233–84.

71. King JA, Dubielzig R, Grimm D, Kleinschmidt JA. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. Embryo. 2001;203:282–91.

72. Savvy A, Dickie Y, Nawynck L, Bonnin D, Merton OW, Gilbrett L. Impact of inverted terminal repeat integration on AAV8 production using the Baculovirus/SF9 cells system. Hum Gene Ther. 2017;28:277–89.

73. Duan D, Sharma P, Dudas L, Zhang Y, Sanillogu S, Yan Z, et al. Formation of adeno-associated virus circular genomes is differentially regulated by adenovirus E4 ORF6 and E2a gene expression. J Virol. 1999;73:161–9.

74. Yan Z, Zak R, Zhang Y, Engelhardt JF. Inverted terminal repeat sequences are important for intermolecular recombination and circularization of adeno-associated virus genomes. J Virol. 2005;79:364–79.

75. Samulski RJ, Chang LS, Shenk T. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. J Virol. 1987;61:3096–101.

76. Tai PWL, Xie J, Fong K, Seetin M, Heiner C, Su Q, et al. Adeno-associated virus genome population sequencing achieves full vector genome resolution and reveals human-vector chimeras. Molecular therapy. Meth Clin Dev. 2018;5:310–41.

77. Xiaowei Ch, Chirumuru B, Berta SC, McCullough B, Gao G, Wilson JM. Gene therapy vectors based on adeno-associated virus type 1. J Virol. 1999;73:3994–4003.

78. Lusby E, Fife KH, Berns KI. Nucleotide sequence of the inverted terminal repetition in adeno- associated virus DNA. J Virol. 1980;34:402–9.

79. Rutledge EA, Halbert CL, Russell DW. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. J Virol. 1998;72:309–19.

80. Chiorini JA, Kim F, Yang L, Kotin RM. Cloning and characterization of adeno-associated virus type 5. J Virol. 1999;73:1309–19.

81. Estvez C, Villagras P. Sequence analysis, viral rescue from infectious clones and generation of recombinant viruses of the avian adeno-associated virus. Virus Res. 2004;105:195–208.

82. Bossis I, Chiorini JA. Cloning of an avian adeno-associated virus (AAV) and generation of recombinant AAV particles. J Virol. 2003;77:6799–810.

83. Schmidt M, Katoano H, Bossis I, Chiorini JA. Cloning and characterization of a bovine adeno-associated virus. J Virol. 2004;78:6509–16.

84. Grimm D, Kay MA, Kleinschmidt JA. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. Mol Ther. 2003;7:839–50.

85. Halbert CL, Allen JM, Miller AD. Adeno-associated virus type 6 (aa6v) vectors mediate efficient transduction of airway epithelial cells in mouse lungs compared to that of aa2v vectors. J Virol. 2001;75:6615–24.

86. Chejanovsky N, Carter BJ. Mutagenesis of an AUG codon in the adeno-associated virus rep gene: effects on viral DNA replication. Virology. 1989;173:120–8.

87. Yew NS, Cheng SH. Reducing the immunostimulatory activity of CpG-containing plasmid DNA vectors for non-viral gene therapy. Expert Opinion Drug Deliv. 2004;1:115–25.

88. Hyde SC, Pringle IA, Abdullah S, Lawton AE, Davies LA, Varathalingam A, et al. CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. Nat Biotechnol. 2008;26:491–51.

89. Duan D. Systemic AAV. Micro-dystrophin gene therapy for duchenne muscular dystrophy. Mol Ther. 2018;26:2337–56.

ACKNOWLEDGEMENTS
This work was supported in part by grants from National Institutes of Health (NS-090634 and AR-70517 to DD, AI-116595 to DJP), Department of Defense (MD130014 to DD), Jesse’s Journey—The Foundation for Gene and Cell Therapy (to DD), and Jackson Freell DMD Research Fund (to DD). The authors thank Dr. Arun Srivastava (University of Florida) for providing the AAV9 tyrosine mutant Rep–Cap packaging plasmid and Dr. Glenn Morris (The Robert Jone and Agnes Hunt Orthopedic Hospital, UK) for providing the MANNEX 44 A antidystrophin antibody. The authors thank Duan lab members for helpful discussion. The authors thank Lisa Burger for expert technical assistance.

AUTHOR CONTRIBUTIONS
Conceived and designed experiments: XP and DD. Performed the experiments: XP, YY, MB, LPW, NTT, KZ, and PWLT. Analyzed the data: XW, YY, MB, LPW, DJP, PWLT, and DD. Wrote the paper: XP and DD. All authors edited the paper and approved submission.

COMPETING INTERESTS
DD is a member of the scientific advisory board for Solid Bosciences and equity holders of Solid Biosciences. DD and YY are inventors on patents that were licensed to Solid Biosciences. The Duan lab has received research support (unrelated to this project) from Solid Biosciences and Edgewise Therapeutics in the last three years.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41434-021-00296-0.

Correspondence and requests for materials should be addressed to Dongsheng Duan.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.