Correction of a splicing defect in a mouse model of congenital muscular dystrophy type 1A using a homology-directed-repair-independent mechanism

Dwi U Kemaladewi[1], Eleonora Maino[1,2], Elzbieta Hyatt[1], Huayun Hou[1,2], Maylynn Ding[1], Kara M Place[1], Xinyi Zhu[1], Prabhpreet Bassi[1,2], Zahra Baghestani[1], Amit G Deshwar[3], Daniele Merico[1,3,4], Hui Y Xiong[3], Brendan J Frey[3,5], Michael D Wilson[1,2,6] , Evgueni A Ivakine[1] & Ronald D Cohn[1,2,7]

Splice-site defects account for about 10% of pathogenic mutations that cause Mendelian diseases[1]. Prevalence is higher in neuromuscular disorders (NMDs)[2], owing to the unusually large size and multi-exonic nature of genes encoding muscle structural proteins. Therapeutic genome editing to correct disease-causing splice-site mutations has been accomplished only through the homology-directed repair pathway[3–5], which is extremely inefficient in postmitotic tissues such as skeletal muscle[6]. Here we describe a strategy using nonhomologous end-joining (NHEJ) to correct a pathogenic splice-site mutation. As a proof of principle, we focus on congenital muscular dystrophy type 1A (MDC1A), which is characterized by severe muscle wasting and paralysis[7]. Specifically, we correct a splice-site mutation that causes the exclusion of exon 2 from *Lama2* mRNA and the truncation of Lama2 protein in the *dy[2J]/dy[2J]* mouse model of MDC1A[8]. Through systemic delivery of adeno-associated virus (AAV) carrying clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 genome-editing components, we simultaneously excise an intronic region containing the mutation and create a functional donor splice site through NHEJ. This strategy leads to the inclusion of exon 2 in the *Lama2* transcript and restoration of full-length Lama2 protein. Treated *dy[2J]/dy[2J]* mice display substantial improvement in muscle histopathology and function without signs of paralysis.

Congenital muscular dystrophy type 1A (MDC1A) is an autosomal recessive neuromuscular disease characterized by neonatal onset of muscle weakness, hypotonia, dysmyelinating neuropathy, and white-matter abnormalities. MDC1A is caused by a diverse range of mutations in the *LAMA2* gene, which encodes the α2 chain (Lama2) of the laminin-211 protein complex expressed in the basement membrane of muscle and Schwann cells. In MDC1A, laminin 211 loses its proper interactions with receptors such as integrin $\alpha_7\beta_1$ and dystroglycan, which results in apoptosis and degeneration of muscle and Schwann cells and leads, in turn, to fibrosis and loss of muscle function. So far, therapeutic developments for MDC1A have focused mainly on preventing fibrosis and apoptosis (recently summarized in ref. 7), but targeting of the primary genetic defect has not been attempted.

The degree of LAMA2 deficiency correlates with the severity of the disease phenotype in individuals with MDC1A and in mouse models of the disease[9]. Therefore, restoration of LAMA2 expression holds tremendous potential for the treatment of MDC1A. However, a standard gene-therapy approach for reintroducing the *LAMA2* coding region of about 9.5 kb into the genome is impeded by the limited capacity of AAV vectors. Moreover, a therapeutic strategy designed to miniaturize the gene, as in the case of microdystrophin or exon skipping in Duchenne muscular dystrophy (recently reviewed in[10]), is unlikely to be successful, owing to the lack of redundant regions in the protein.

Approximately 40% of individuals with MDC1A harbor mutations in *Lama2* splice sites[9,11]. The MDC1A mouse model, *dy[2J]/dy[2J]* has a splice-site mutation in intron 2 of *Lama2*, annotated as c.417+1 g→a, resulting in skipping of exon 2 and truncation of the N-terminal protein domain (**Fig. 1a**)[8,12]. The lack of functional Lama2 causes severe muscle atrophy and hind-limb paralysis in these mice.

The ability to correct disease-causing mutations has been improved greatly by the discovery of CRISPR–Cas9 genome-editing technology[13,14]. Cas9 endonucleases generate double-strand breaks (DSBs) in a specific genomic region, which is located adjacent to a protospacer-adjacent motif (PAM) and targeted by a complementary single guide RNA (sgRNA). In the presence of a DNA repair template, precise genomic modification can take place through homology-directed repair (HDR), whereas NHEJ repairs Cas9-induced breaks in the absence of the exogenous template.

[1]Program in Genetics and Genome Biology, The Hospital for Sick Children Research Institute, Toronto, Canada. [2]Department of Molecular Genetics, University of Toronto, Canada. [3]Deep Genomics, Toronto, Canada. [4]The Centre for Applied Genomics (TCAG), The Hospital for Sick Children Research Institute, Toronto, Canada. [5]Department of Electrical and Computer Engineering and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada. [6]Heart & Stroke Richard Lewar Centre of Excellence in Cardiovascular Research, Toronto, Canada. [7]Department of Pediatrics, University of Toronto and The Hospital for Sick Children, Toronto, Canada. Correspondence should be addressed to E.A.I. (zhenya.ivakine@sickkids.ca) or R.D.C. (ronald.cohn@sickkids.ca).
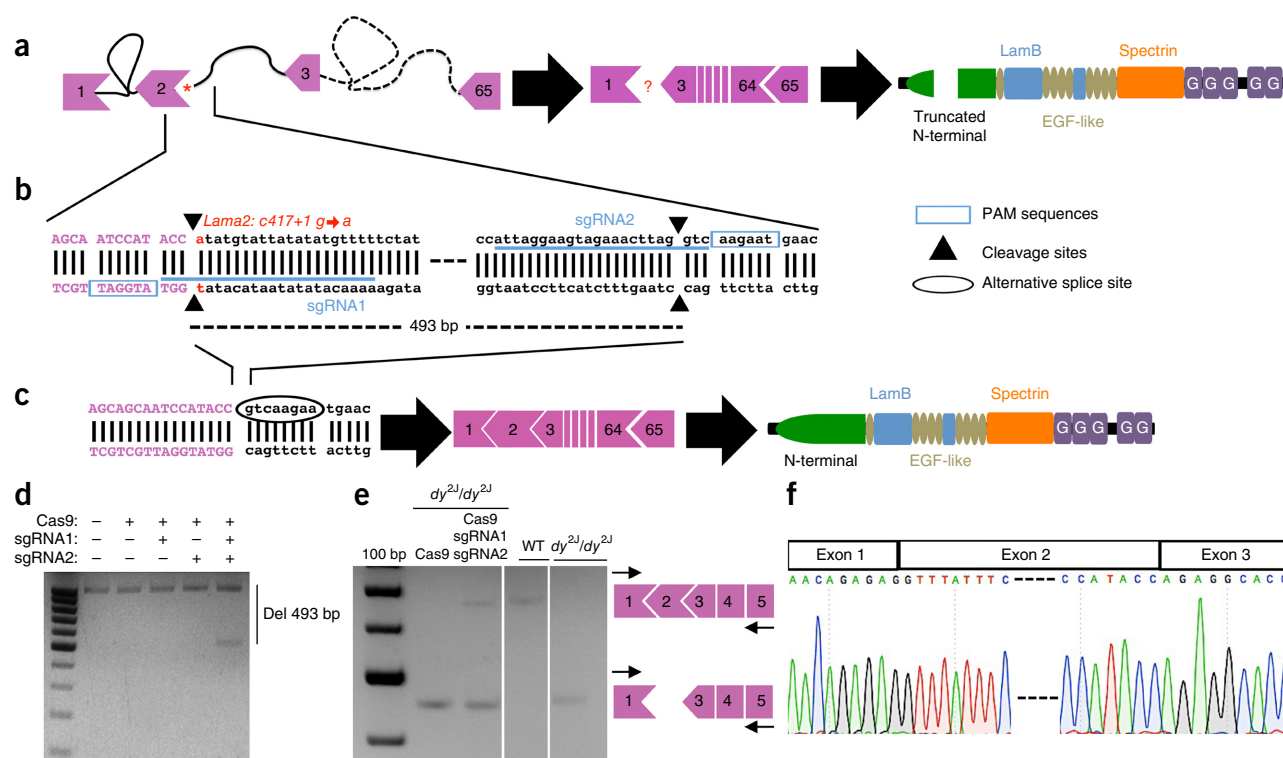
Figure 1 An HDR-independent strategy for correcting a *Lama2* splicing defect. (**a**) In *dy*[2J]/*dy*[2J] mice, a point mutation (red asterisk) in *Lama2* intron 2 disrupts the splice donor site, causing spontaneous skipping of exon 2 (red question mark) and truncation of the N-terminal domain of Lama2 protein. LamB, laminin B domain; EGF-like, epidermal-growth-factor-like domain; spectrin, spectrin repeats; G, globular repeats. (**b**) Two *Sa*Cas9 genomic targets were identified in the proximal region of the intron 2 (lowercase). The sgRNA1 target overlaps the splice-site mutation, annotated as *Lama2: c417+1* g→a (red), whereas the sgRNA2 target is located 493 bp downstream of the exon (uppercase)–intron boundary. PAM sequences (blue boxes), *Sa*Cas9 cleavage sites (black triangles), and two sgRNAs (blue, underlined) are indicated. (**c**) A combination of *Sa*Cas9, sgRNA1, and sgRNA2 creates simultaneous DSBs in the genomic region flanking the two targets, which is then repaired through NHEJ. A functional splice donor site in *Lama2* is created (oval), which leads to exon 2 inclusion and restoration of the expression of Lama2 protein. (**d**,**e**) PCR analysis on genomic DNA (**d**) and RT–PCR on RNA (**e**) isolated from *dy*[2J]/*dy*[2J]-derived myoblasts transfected with *Sa*Cas9 and the corresponding sgRNAs. Arrows correspond to primers in exons 1 and 5 used in RT–PCR. PCR and RT–PCR data from one of out of three independent experiments is shown. (**f**) Verification of exon 2 inclusion using Sanger sequencing.

The correction of point mutations, including splice-site mutations, mainly utilizes the HDR pathway. In metabolic liver diseases, the injection of CRISPR–Cas9 and a DNA repair template corrected mutations in fumarylacetoacetate hydrolase (*Fah*) and ornithine transcarbamylase (*Otc*) in hepatocytes, which resulted in amelioration of disease symptoms in mice[3–5]. However, HDR in postmitotic tissues, such as skeletal muscle and/or neurons, is extremely inefficient, necessitating the development of alternative strategies to correct mutations.

Using the *dy*[2J]/*dy*[2J] mice as a model, we designed an HDR-independent strategy to correct a splice-site mutation in the *Lama2* gene. Most mammalian donor splice sites have a nine-nucleotide consensus sequence of A/C-A-G-g-t-a/g-a-g-u and are recognized by canonical base pairing to the 5′ end of U1 small nuclear RNA[15]. The first two intronic nucleotides, 'gt,' are essentially invariant, whereas certain flexibility is allowed at other positions. We hypothesized that targeted CRISPR–Cas9-mediated removal of an intronic region containing a pathogenic mutation would enable the *Lama2* exon 2 to be joined with a 'gt' dinucleotide, and thus reconstitute a functional donor splice site.

We searched for *Staphylococcus aureus* Cas9 (*Sa*Cas9) PAM sequences (NNGRR(T)) in proximity to the g→a mutation in *Lama2*, and we detected one at the end of exon 2 (**Fig. 1b**). *Sa*Cas9 is predicted to generate a DSB three nucleotides upstream of the PAM sequence[14],

immediately upstream of the mutation (**Fig. 1b**). sgRNA targeting of this region was designed and annotated as sgRNA1, serving as 'mutation-eliminating sgRNA.'

We evaluated ~1,400 nucleotides in a proximal region of intron 2 for *Sa*Cas9 targets to generate genomic deletions in combination with sgRNA1 and, after NHEJ, to create potential donor splice sites (a 'splice-site-generating cut') (**Fig. 1c**). Ten out of 13 possible splice-site-generating sgRNAs (sgRNA2–sgRNA11) were tested experimentally in *dy*[2J]/*dy*[2J] myoblasts through co-transfection with sgRNA1 (**Fig. 1d**, **Supplementary Fig. 1a** and **Supplementary Table 1**). RT–PCR analysis revealed the presence of an amplicon similar to the wild-type control for sgRNA2, sgRNA3, and sgRNA5, indicating an inclusion of exon 2 of *Lama2* after the generation of intronic deletions (**Fig. 1e** and **Supplementary Fig. 1b**). Among these three functional splice-site-generating sgRNAs, the sgRNA2 target is located closest to the exon–intron boundary (**Fig. 1b**), and Sanger sequencing confirmed that full-length exon 2 was incorporated successfully into the *Lama2* mRNA in sgRNA1 + sgRNA2–treated *dy*[2J]/*dy*[2J] myoblasts (**Fig. 1f**).

To investigate the impact of the full-length Lama2 restoration in skeletal muscle, we treated 3-week-old (early-disease onset) *dy*[2J]/*dy*[2J] mice with AAV serotype 9 (AAV9) encoding hemagglutinin (HA)-tagged *Sa*Cas9 and either sgRNA1 or sgRNA2. Each mouse received a single
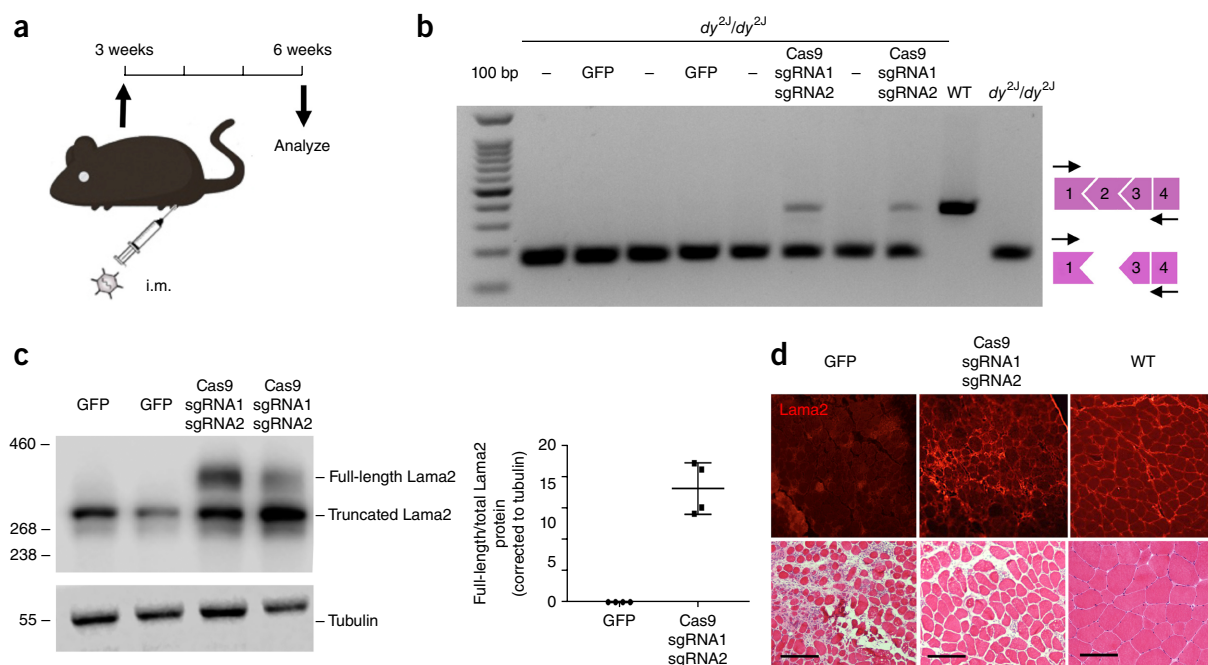
Figure 2 Restoration of full-length Lama2 and improvement of muscle pathology following intramuscular AAV9–CRISPR administration in adult dy^{2J}/dy^{2J} mice. (**a**) AAV9-Cas9–sgRNA1 and AAV9-Cas9–sgRNA2 (7.5×10^{11} viral genome each) were injected into the left TA muscle of 3-week-old dy^{2J}/dy^{2J} mice ($n = 4$). Mice receiving AAV9-GFP ($n = 4$) at equivalent doses serve as a control. All animals were euthanized 3 weeks after treatment. (**b**) The TA muscles were isolated and analyzed for exon 2 inclusions by RT–PCR using primers depicted by the arrows. Minus (–) sign indicates noninjected, contralateral TA muscles. (**c**) Protein lysates isolated from the injected TA muscles were probed for Lama2 expression by western blot. Tubulin serves as a loading control. Representative image of two animals per group is shown on gel. Densitometry analysis on the amount of full-length form over total Lama2 proteins, corrected to tubulin expression, was done on four mice, and the average was plotted. Error bar represents s.d. (**d**) 8-μm cross-sections were analyzed for localization of Lama2 and general muscle architecture by immunofluorescence (top) and H&E (bottom) staining, respectively. A representative sample from four mice per treatment group is shown. Scale bars, 50 μm.

intramuscular injection of AAV9 cocktail of 7.5×10^{11} vector genomes of AAV9-Cas9–sgRNA1 and 7.5×10^{11} vector genomes of AAV9-Cas9–sgRNA2 in the left tibialis anterior (TA) (**Fig. 2a**). Mice injected with titer-matched AAV9 particles encoding GFP served as controls. All animals were killed 3 weeks after time of injection.

RT–PCR analysis detected the inclusion of exon 2 in transcripts in all left-TA muscles treated with AAV9-Cas9–sgRNA1 and AAV9-Cas9–sgRNA2, but not in contralateral muscles; no exon-2-containing *Lama2* transcripts were detected in AAV9-GFP-treated mice (**Fig. 2b**).

We subsequently analyzed the expression of Lama2 by western blot and observed that all dy^{2J}/dy^{2J}-isolated TA muscles injected with the AAV9-Cas9–sgRNA1 or the AAV9-Cas9–sgRNA2 combination expressed the full-length protein (**Fig. 2c**), demonstrating successful translation of the exon-2-containing transcript.

Immunofluorescence staining of restored Lama2 revealed sarcolemmal expression in the treated TA, similar to in the wild-type controls (**Fig. 2d**). H&E staining exhibited considerably improved muscle architecture, as compared to the contralateral TA, with reduced areas of fibrosis and inflammatory cells, a finding that supports the therapeutic potential of our strategy in the dy^{2J}/dy^{2J} mice (**Fig. 2d**).

Subsequently, we tested whether the dystrophic pathology and paralysis could be prevented by systemic restoration of Lama2 in neonatal mice. 2-day-old (P2) dy^{2J}/dy^{2J} mice were injected with AAV9-Cas9–sgRNA1 and AAV9-Cas9–sgRNA2 (7.5×10^{11} vector genomes each) through the intraperitoneal route and compared to AAV9-GFP-treated controls (**Fig. 3a**). Western blot analysis of gastrocnemius muscles isolated 10 weeks after injection—a time

point that corresponds to an intermediate stage of disease progression—showed expression of the HA-tagged Cas9 only in the injected mice (**Supplementary Fig. 2**). RT–PCR and quantitative RT–PCR indicated the inclusion of exon 2 in ~20% of *Lama2* transcripts (**Fig. 3b,c**). Western blot of muscle lysates demonstrated approximately 20% recovery of full-length Lama2 protein (**Fig. 3d,e**), and sarcolemmal expression in muscle by immunofluorescence staining (**Fig. 3f**, left). Restoration of Lama2 expression improved muscle morphology and reduced the extent of collagen deposition, as demonstrated by H&E and picrosirius red staining, respectively (**Fig. 3f**, middle and right). Morphometric analyses showed ~50% reduction of fibrosis (**Fig. 3g**) and a 12% and 9% increase, respectively, in cross-sectional area and minimum Feret diameter (**Supplementary Fig. 3a,b**). We also observed a mild shift of distribution toward larger muscle fibers (**Supplementary Fig. 3c**) and a decrease in the number of centrally located nuclei after treatment (**Fig. 3h**).

Importantly, RT–PCR and immunofluorescence analyses of sciatic nerves showed no restoration of the *Lama2* transcript or Lama2 protein (**Fig. 3i,j**). In line with this finding, general locomotion (**Fig. 3k**) and extent of paralysis showed no substantial improvement, which is consistent with previous studies demonstrating that muscle-specific overexpression of laminin 211 in Lama2-deficient mice was not sufficient to restore paralysis[16].

To achieve efficient targeting of peripheral nerves in addition to skeletal muscle[17–19], we injected AAV9-Cas9–sgRNA1 and AAV9-Cas9-sgRNA2 (7.5×10^{11} vector genomes each) through the temporal vein of P2 neonatal dy^{2J}/dy^{2J} mice (**Fig. 4a**). At 30 weeks after injection (progressive disease stage), the expected ~493-bp genomic deletion
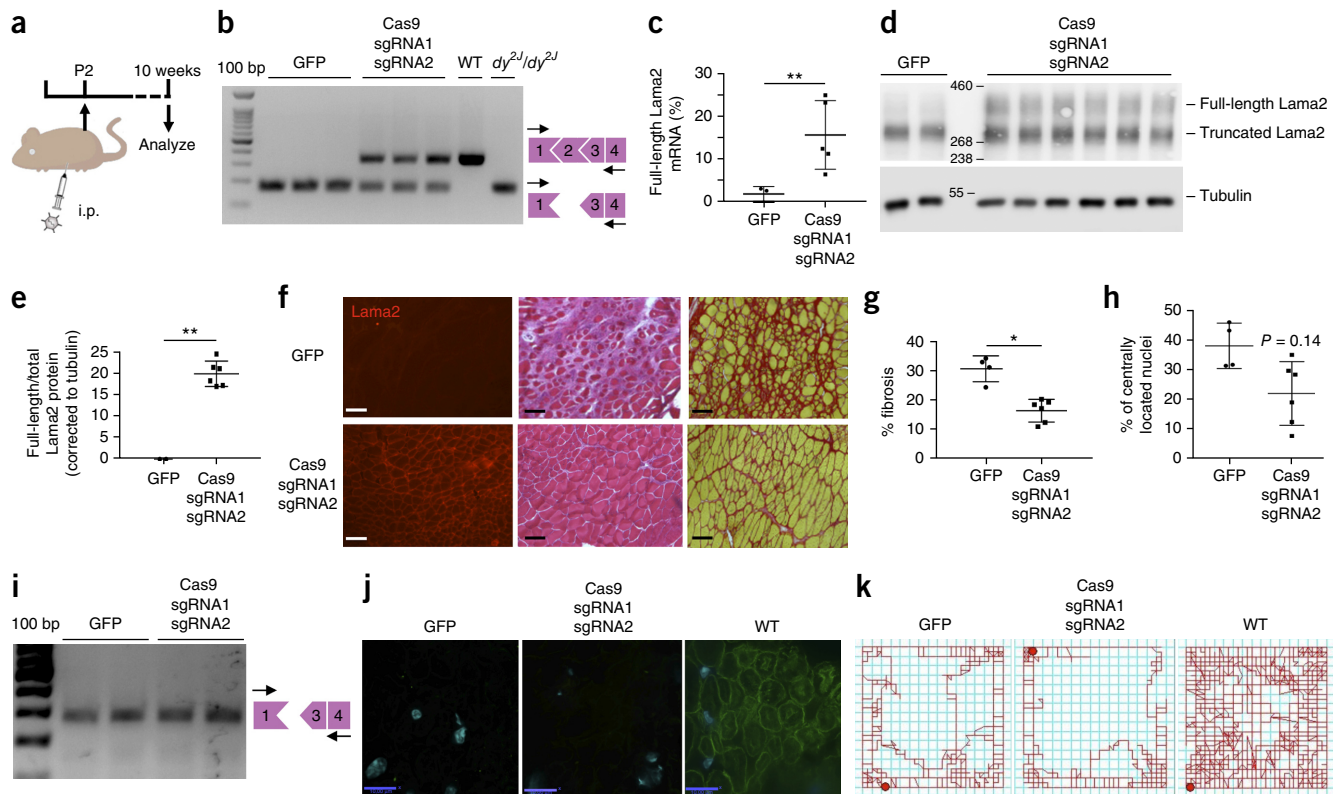
Figure 3 Intraperitoneal administration of AAV9-CRISPR in neonatal dy^{2J}/dy^{2J}. (**a**) 2-day-old dy^{2J}/dy^{2J} pups were injected with AAV9-GFP ($n = 4$) or a combination of AAV9-Cas9–sgRNA1 and AAV9-Cas9–sgRNA2 ($n = 6$) (7.5×10^{11} viral genome each) through the intraperitoneal route, and then killed 10 weeks from time of injection. (**b**) Gastrocnemius muscle was analyzed for exon 2 inclusion by RT–PCR. A representative sample of three mice/group is shown. (**c**) Quantification of transcripts containing exon 2 by qPCR. (**d**) Western blot analysis to detect the restoration of full-length Lama2 expression. Tubulin serves as a loading control. (**e**) Densitometry analysis to quantify the amount of full-length form over total Lama2 proteins, corrected to tubulin expression. (**f**) Immunofluorescence staining with antibodies against Lama2 (left), H&E (middle), and picrosirius red (right) staining on 8-μm gastrocnemius muscle cross-section. A representative sample from each treatment group is shown. Scale bars, 50 μm. (**g**) Quantification of fibrosis indicated by Sirius red staining in **f**. (**h**) Percentage of myofibers with centrally located nuclei. Data in **c** and **e** are presented as means ± s.d., whereas **g** and **h** are presented as means ± s.e.m. Statistical analyses were performed using Student's t-test. *$P < 0.05$, **$P < 0.01$. (**i**) RT–PCR analysis on RNA isolated from sciatic nerves. (**j**) Immunofluorescence staining to detect Lama2 expression (green) on sciatic nerves. (**k**) Example of mouse trajectories in 20-min open-field test.

was detected in muscle and sciatic nerve by PCR (**Supplementary Fig. 4a**), and quantification by droplet digital PCR (ddPCR) revealed ~3% editing at the DNA level (**Supplementary Fig. 4b**). Deep sequencing showed that the majority of DSBs occurred at the intended sites and were properly joined by the NHEJ mechanism, leading to 56.29%-perfect ligation (**Supplementary Table 2**), resulting in the reconstitution of a splice donor site that allowed for the inclusion of exon 2. RT–PCR and western blot analyses of the TA muscles indicated expression of the full-length *Lama2* transcript and protein (**Fig. 4b–e**) of approximately 20% and 25%, respectively. Histological analysis showed expression of the Lama2 protein at the sarcolemma (**Fig. 4f**), significant reduction of fibrosis (**Fig. 4g**) and centrally located nuclei (**Fig. 4h**), and an increase in minimum Feret diameter and cross-sectional area (**Fig. 4i** and **Supplementary Fig. 4c,d**). Similar improvement in dystrophic pathology was achieved in the gastrocnemius muscle (**Supplementary Fig. 5**). In the diaphragm, Lama2 restoration significantly reduced fibrosis ($P < 0.05$), minimum Feret diameter ($P < 0.01$) and cross-sectional area ($P < 0.01$) (**Supplementary Fig. 6**). This is consistent with previously published results in 30-week-old dy^{2J}/dy^{2J} mice treated with the anti-apoptotic drug omigapil[20], which is currently undergoing testing in a clinical trial for treatment of individuals with MDC1A. There was no change

in the number of inflammatory cells expressing CD11b after treatment (**Supplementary Fig. 7**).

Importantly, full-length *Lama2* mRNA and protein expression was also detected in the sciatic nerve (**Supplementary Fig. 8a,b** and **Fig. 4j**). To assess the impact of Lama2 restoration on myelination, we stained the sciatic nerves with antibody against myelin binding protein (MBP). In control-treated dy^{2J}/dy^{2J} mice, we observed neurofilament-H-expressing axons lacking MBP expression (**Fig. 4j**). In treated mice, axons are encapsulated by MBP, similarly to in the wild-type sciatic nerve, indicating improvement of the myelination defect. Consequently, the degree of paralysis in the hind limbs and mobility of the treated mice were markedly improved (**Supplementary Videos 1** and **2**). An open-field test demonstrated significantly increased total distance traveled in treated mice as compared to the control dy^{2J}/dy^{2J} group (**Fig. 4k**). Vertical activity, indicating rearing behavior that requires hind-limb strength[21], was detected in treated dy^{2J}/dy^{2J} mice at a near-wild-type levels (**Fig. 4l**). Furthermore, specific tetanic force, measuring the aggregate torque produced by the dorsi flexor muscles, was also restored toward wild-type levels (**Fig. 4m**). This assay requires proper innervation of the muscles[22,23], and thus demonstrates that early restoration of Lama2 expression in both peripheral nerves and skeletal muscles results in near-complete amelioration of both phenotypes.
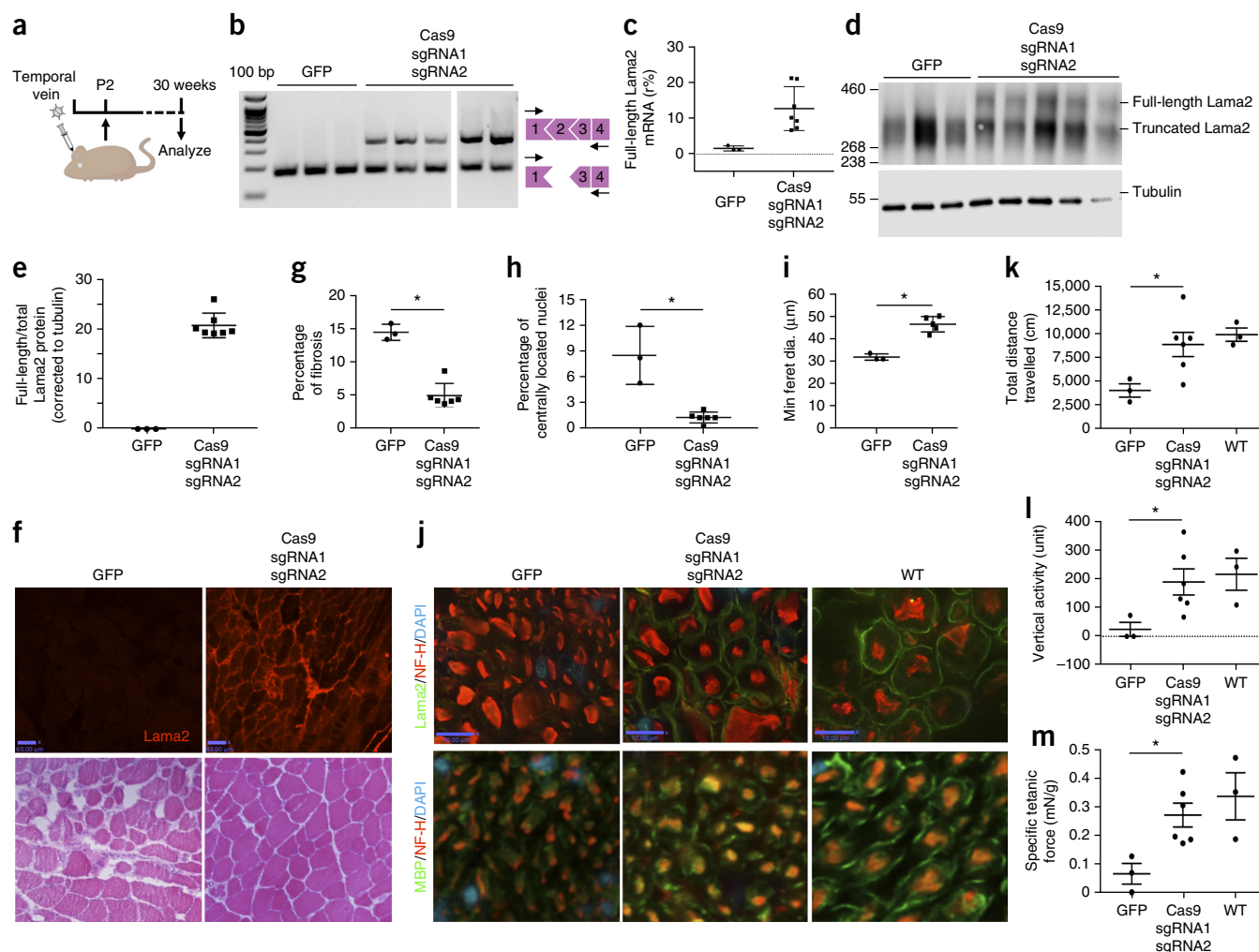
Figure 4 Temporal-vein administration of AAV9-CRISPR in neonatal *dy2J/dy2J* mice. (**a**) 2-day-old *dy2J/dy2J* pups were injected with AAV9-GFP (*n* = 3) or a combination of AAV9-Cas9–sgRNA1 and AAV9-Cas9–sgRNA2 (*n* = 7) (7.5 × 10[11] viral genome/vector) through the temporal-vein route, and were killed 30 weeks after injection. (**b**) TA muscle was analyzed for exon 2 inclusion by RT–PCR. (**c**) Quantification of transcripts containing exon 2 by qPCR. (**d**) Western blot analysis to detect restoration of full-length Lama2 expression. Tubulin serves as a loading control. (**e**) Densitometry analysis to quantify the amount of full-length form over total Lama2 proteins, corrected to tubulin expression. (**f**) Immunofluorescence staining with antibodies against Lama2 (top) and H&E (bottom) staining on 8-μm TA muscle cross-section. A representative sample from each treatment group is shown. Scale bar, 50 μm. (**g**) Quantification of fibrosis indicated by H&E staining in **f**. (**h**) Percentage of myofibers with centrally located nuclei in TA muscles. (**i**) Mean minimum Feret diameter of TA muscles. (**j**) Immunofluorescence staining on sciatic nerves with antibodies against Lama2 (green, top) or MB, (green, bottom) and neurofilament H (red). DAPI (blue) stains the nuclei. A representative sample from each treatment group is shown. (**k,l**) 10 weeks after injection, the mice were tested in an open-field chamber in which total distance (**k**) and vertical activity (**l**) were assessed. (**m**) Specific tetanic force was measured using an *in vivo* muscle-function analyzer. Data in **c,e,k** and **m** are presented as means ± s.d., whereas **g**–**i** are presented as means ± s.e.m. Statistical analyses were performed using a Student's *t*-test. **P* < 0.05, ***P* < 0.01.

Cas9 endonucleases can sometimes generate unintended DSBs at off-target sites throughout the genome[24], which presents a potential safety concern. We did not observe any off-target mutations at 20 potential off-target sites (ten for each sgRNA) in the mouse genome (**Supplementary Fig. 9**). Off-target sites may present at locations beyond those predicted *in silico*; hence, a comprehensive and unbiased analysis using whole-genome sequencing might be necessary to establish the safety of this approach[25].

The strategy described here is particularly attractive for the correction of underlying defects of genetic diseases caused by splice-site mutations, especially those affecting tissues in which HDR is inefficient. The discovery of Cas9 proteins derived from various species that differ in their PAM-sequence requirements[13,14,26,27], as well as

engineered Cas9 derivatives with purposefully altered PAM specificities[28,29], offers unprecedented flexibility in genome-editing endeavors. About 10% of ~80,000 mutations reported in the human-gene-mutation database (HGMD) affect splice sites[1], and, specifically, there are 3,390 disease-causing mutations that occur at the +1 donor splice site. These mutations affect 2,754 exons in 901 genes. We predict that 57.7% of the splice-site mutations can be cut at the exon–intron junction by using nine available Cas9 endonucleases. The majority of those (~96%) can be cleaved concurrently in the intron to generate putative splice sites (**Supplementary Fig. 10a**) (**Supplementary Table 3**). For these targetable mutations, we detected a median of 22 intronic 'gt's (**Supplementary Fig. 10b**), which corresponds to approximately 47% of total 'gt's in the intron (**Fig. 10c**).

With future discoveries and directed engineering of additional Cas9 species, or by using ZFNs or TALENs fused with a type IIS restriction enzyme that generates blunt-ended DSB, nearly every exon–intron junction and intronic 'gt' could be precisely targeted.

Using current state-of-the-art, *in silico* splice-site predictors (MaxEntScan[30] and DeepScan), a survey of GENCODE v19 transcripts revealed that, at a stringent splice-site detection threshold, 54% of exons have at least one intronic 'gt' site within 1 kb that can be used to constitute a new splice site by deleting the intervening intronic sequence, with an average of about three 'gt' sites per exon (**Supplementary Table 4**). Given that ~47% of all intronic 'gt's are targetable (**Supplementary Fig. 10c**), there are, on average, one or two 'gt' sites per exon that are predicted to generate a functional splice site after juxtaposition to a corresponding exon. It is important to emphasize that this average is most likely an underestimate, because our experimentally validated sgRNA2-, sgRNA3-, and sgRNA5-generated junctions produced meaningful rankings but very low absolute scores (**Supplementary Table 5**). Furthermore, by manually inspecting our experimentally validated 'gt' sites, we noticed that two of them present an intronic consensus bulge[31], which is one of the noncanonical splice sites and not modeled optimally by either predictor. Future refinements of the models, while taking into consideration existing noncanonical registers for U1 snRNA binding[31,32], are needed to improve *in silico* predictions for novel genetically engineered exon–intron junctions. Importantly, although our current analysis focused only on +1 splice-donor-site mutations, similar approaches are applicable for correcting disease-causing 5′-splice-site variants at other positions (+2, …, +6), which substantially broadens the applicability of this technology for therapeutic interventions.

As a whole, our strategy for correcting noncoding mutations by modifying an intronic region to create a functional splice donor site provides a novel therapeutic approach that may be applicable not only to MDC1A, but also to various other inherited diseases.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

D.U.K., E.A.I., and R.D.C. conceived the study, designed the experiments, analyzed data, and wrote and edited the manuscript. D.U.K., E.M., E.H., M.D., X.Z., K.M.P., P.B., and Z.B. performed the experiments. A.G.D., D.M., H.Y.X., B.J.F., H.H., and M.D.W. performed bioinformatics analyses. All authors provided feedback and agreed on the final manuscript.

1. Scotti, M.M. & Swanson, M.S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
2. Bladen, C.L. *et al.* The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. *Hum. Mutat.* **36**, 395–402 (2015).
3. Yang, Y. *et al.* A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).
4. Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32**, 551–553 (2014).
5. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components *in vivo. Nat. Biotechnol.* **34**, 328–333 (2016).
6. Vahidi Ferdousi, L. *et al.* More efficient repair of DNA double-strand breaks in skeletal muscle stem cells compared to their committed progeny. *Stem Cell Res.* **13** 3 Pt A, 492–507 (2014).
7. Durbeej, M. Laminin-α2 chain-deficient congenital muscular dystrophy: pathophysiology and development of treatment. *Curr. Top. Membr.* **76**, 31–60 (2015).
8. Sunada, Y., Bernier, S.M., Utani, A., Yamada, Y. & Campbell, K.P. Identification of a novel mutant transcript of laminin alpha 2 chain gene responsible for muscular dystrophy and dysmyelination in dy2J mice. *Hum. Mol. Genet.* **4**, 1055–1061 (1995).
9. Geranmayeh, F. *et al.* Genotype-phenotype correlation in a large population of muscular dystrophy patients with LAMA2 mutations. *Neuromuscul. Disord.* **20**, 241–250 (2010).
10. Guiraud, S. *et al.* The pathogenesis and therapy of muscular dystrophies. *Annu. Rev. Genomics Hum. Genet.* **16**, 281–308 (2015).
11. Stenson, P.D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* **133**, 1–9 (2014).
12. Xu, H., Wu, X.R., Wewer, U.M. & Engvall, E. Murine muscular dystrophy caused by a mutation in the laminin alpha 2 (Lama2) gene. *Nat. Genet.* **8**, 297–302 (1994).
13. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
14. Ran, F.A. *et al. In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
15. Roca, X., Krainer, A.R. & Eperon, I.C. Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes Dev.* **27**, 129–144 (2013).
16. Kuang, W. *et al.* Merosin-deficient congenital muscular dystrophy. Partial genetic correction in two mouse models. *J. Clin. Invest.* **102**, 844–852 (1998).
17. Tanguy, Y. *et al.* Systemic AAVrh10 provides higher transgene expression than AAV9 in the brain and the spinal cord of neonatal mice. *Front. Mol. Neurosci.* **8**, 36 (2015).
18. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J.E. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16**, 1073–1080 (2008).
19. Gombash Lampe, S.E., Kaspar, B.K. & Foust, K.D. Intravenous injections in neonatal mice. *J. Vis. Exp.* **93**, e52037 (2014).
20. Yu, Q. *et al.* Omigapil treatment decreases fibrosis and improves respiratory rate in dy(2J) mouse model of congenital muscular dystrophy. *PLoS One* **8**, e65468 (2013).
21. Tatem, K.S. *et al.* Behavioral and locomotor measurements using an open field activity monitoring system for skeletal muscle diseases. *J. Vis. Exp.* **91**, 51785 (2014).
22. Iyer, S.R., Valencia, A.P., Hernández-Ochoa, E.O. & Lovering, R.M. *In vivo* assessment of muscle contractility in animal studies. *Methods Mol. Biol.* **1460**, 293–307 (2016).
23. Call, J.A. & Lowe, D.A. Eccentric contraction-induced muscle injury: reproducible, quantitative, physiological models to impair skeletal muscle's capacity to generate force. *Methods Mol. Biol.* **1460**, 3–18 (2016).
24. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).
25. Tsai, S.Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
26. Hirano, H. *et al.* Structure and engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).
27. Kim, E. *et al. In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni. Nat. Commun.* **8**, 14500 (2017).
28. Kleinstiver, B.P. *et al.* Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
29. Kleinstiver, B.P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
30. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
31. Roca, X. *et al.* Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.* **26**, 1098–1109 (2012).
32. Tan, J. *et al.* Noncanonical registers and base pairs in human 5′ splice-site selection. *Nucleic Acids Res.* **44**, 3908–3921 (2016).

ONLINE METHODS

Cloning and virus production. Each sgRNA (**Supplementary Table 1**) was cloned into a pX601 plasmid (Addgene 61591) containing codon-optimized *S. aureus* Cas9, as previously described[14]. As a control, a GFP plasmid was constructed by replacing the *Sa*Cas9 cassette in pX601 with an eGFP codon from pEGFP-C1 plasmid (Clontech Takara) using Age1 and EcoR1 directional cloning. A 2,100-bp stuffer sequence was incorporated into the Xba1 site of the vector to maintain AAV genome-packaging capacity. pX601-GFP, pX601-sgRNA1, or pX601-sgRNA2 plasmids were packaged as AAV9 by Vigene Biosciences. Briefly, the transfer plasmid above was co-transfected with Ad-Helper vector and a vector encoding the *rep* and serotype-specific *cap* into HEK293 cells, and viral particles were purified by Iodixanol gradient ultracentrifugation and concentrated to 500 µl. An injection-ready virus solution containing 7.5×10^{11} viral genome/AAV9 was prepared and brought to 30 µl with 1× PBS (Gibco).

Cell culture and guide screening. Primary myoblasts were isolated from Extensor Digitorum Longus (EDL) muscle of C57BL/6J wild-type and *dy2J/dy2J* mice, as previously described[33], and maintained in DMEM supplemented with 1% chicken embryo extract (Gemini Biosciences), 10% horse serum, 1% penicillin/streptomycin, and 1% L-glutamine (all from Gibco, unless indicated otherwise) at 37 °C with 5% CO_2. Transfection was performed using Amaxa Basic Nucleofector Kit VPI-1002 (Lonza) in Nucleofector device (Lonza) under program number U-023. Cells were grown for 72 h, and genomic DNA was isolated using a blood and tissue DNA isolation kit according to the manufacturer's protocol (Qiagen). The ability of the guide combination to remove the intervening region was assessed by PCR using primers in *Lama2* exon 2 (RDC 551) and intron 2 (RDC 552) flanking the genomic deletion. The sequences are listed in **Supplementary Table 2**.

Animal studies. The *dy2J/dy2J* and C57BL/6J mice were purchased from the Jackson Laboratory and maintained in the Toronto Centre for Phenogenomics (TCP). All animal experiments were performed according to Animal Use Protocol number 16-0234H. No statistical methods were used to predetermine sample size.

Intramuscular injection was performed on 3-week-old *dy2J/dy2J* mice under isofluorane anesthesia. Treatments were randomized among littermates. No blinding was done in the intramuscular treatment.

For systemic-administration experiments involving neonatal pups, breeding cages were monitored daily for birth, and the entire litter of P2 newborns were injected with either GFP- or guide-containing vectors. Mice were genotyped[34] and weaned at the age of 3 weeks old. Randomization occurred, and all investigators were subsequently blinded for the different treatment groups throughout the study until the time of necropsy.

Functional tests. Open-field activity test and assessment of *in vivo* muscle force were performed on 10-week-old mice (for both intraperitoneal and temporal vein cohorts) at the Lunenfeld–Tanenbaum Research Institute's Centre for Modeling Human Disease Mouse Phenotyping Facility. For the open-field test, mice were placed in the frontal center of a transparent Plexiglas open field (41.25 cm × 41.25 cm × 31.25 cm) illuminated by 200 lx. A trained operator, who is unaware of the nature of the projects and treatments, performed the experiments. The VersaMax Animal Activity Monitoring System recorded vertical activities and total distance traveled for 20 min per animal.

An *in vivo* muscle-contraction test was performed using 1300A: 3-in-1 Whole Animal System and analyzed using dynamic muscle control/analysis (DMC/DMA) high-throughput software suite (Aurora Scientific). The mice were anesthetized by an intraperitoneal injection of a ketamine/xylazine cocktail at 100 mg/kg and 10 mg/kg of body weight, respectively. Contractile output was measured through percutaneous electrodes that stimulate specific nerves innervating the plantar flexors. Specific tetanic force (200 ms of 0.5-ms pulses at 125 Hz) was recorded, and corrected according to body weight.

RNA isolation, RT–PCR, and quantitative PCR (qPCR). RNA from cultured cells was isolated using Trizol (Invitrogen) according to the manufacturer's instructions. For mouse tissues, muscles were sectioned into 10-µm thin slices, collected in 1.4-mm Zirconium Beads pre-filled tubes (OPS Diagnostics), and homogenized using a MagNA Lyser (Roche Diagnostics) using Trizol (Invitrogen). N6-primed cDNA was synthesized from 1,000 ng of RNA using SuperScript III (Invitrogen). PCR amplification to detect exon inclusion was performed using multiplex PCR mix (Qiagen) and primers in exons 1 and 4 (RDC 449 and 450; **Supplementary Table 2**).

qPCR was performed using Fast SYBR Green Master Mix on Step One Plus Real Time PCR (Applied Biosystems). Control PUC-Lama2 plasmid was generated by directional cloning in our laboratory. Briefly, exons 1–7 *Lama2* cDNA was amplified from wild-type muscles using primers RDC 1411 and RDC 1479 and cloned into a PUC19 plasmid at BamH1 restriction sites. The expression of products with exon 2 inclusion was calculated using primers in the exon 2–exon 3 junctions and normalized to the expression of products amplified with exon 4. The copy numbers of all samples were calculated relative to a PUC-*Lama2* plasmid control.

Protein isolation and western blot. Sectioned muscles collected in the Zirconium beads tubes were homogenized in 500 µl of RIPA homogenizing buffer (50-mM Tris HCl pH 7.4, 150-nM NaCl, 1-mM EDTA, supplemented with protease-inhibitor cocktails (Roche)) and lysed with a MagNA Lyser. Subsequently, 500 µl of RIPA double-detergents buffer (2% deoxycholate, 2% NP40, 2% Triton X-100 in RIPA homogenizing buffer) was added to the lysates, which were then incubated for 45 min at 4 °C with gentle agitation and centrifuged for 10 min at 13,000 r.p.m. Protein concentration was measured using a BCA Assay (Thermo Scientific). Protein was separated on a 3–8% Tris-Acetate gel and transferred using the Novex system (Invitrogen). The primary antibodies used were rat monoclonal anti-Lama2 (Abcam 11576 4H8-2, 1:500), rabbit polyclonal anti-tubulin (Abcam 4074, 1:5,000), and mouse monoclonal anti-HA (Abcam 18181, 1:2,500). The original gel images are provided in **Supplementary Figure 11**.

Immunofluorescence, H&E, and picrosirius red staining. The antibodies used for immunofluorescence staining were rat monoclonal anti-Lama2 (Abcam 11576 4H8-2, 1:500), mouse monoclonal anti-neurofilament H (NF-H, BioLogend 801601 SMI 31, 1:1,000), and FITC-conjugated rat anti-mouse CD11b clone M1/70 (BD Pharmingen 553310, 1:50), and were imaged using either Zeiss Axiovert 200M or Nikon TE-2000 epifluorescence microscopes, or an Olympus 1X81 quorum-spinning disk confocal microscope. PerkinElmer Volocity software was used for image acquisition.

H&E staining was performed using a standard protocol, and picrosirius red staining was performed at the Pathology Laboratory, the Hospital for Sick Children, Toronto. 300 fibers obtained from four different field of views per animal were analyzed using ImageJ 1.51h software for a minimum Feret diameter and cross-sectional area. Fibrosis quantification was performed from picrosirius-red-stained (for the intraperitoneal cohort) and H&E-stained (for temporal-vein cohort) images by image segmentation using ImageJ 1.51h software.

Off-target analysis by T7 endonuclease assay. The top ten predicted off-target sites for each sgRNA were computed using a Benchling CRISPR Design tool[35]. Genomic DNA isolated from treated and control mice was PCR amplified using primers targeting loci corresponding to each hit (**Supplementary Tables 6** and **7**). Purified PCR products were then digested using T7 Endonuclease I (NEB) according to the manufacturer's instructions.

Deep sequencing. Genomic DNA extracted from TA muscles from the temporal-vein cohorts (Qiagen Blood and Tissue kit) was amplified using oligos RDC 1474 and RDC 1461 (**Supplementary Tables 1** and **6**). A second round of PCR was performed to add 5′- and 3′-end barcodes, and the sample was sequenced with 125-bp paired-end reads on Illumina HiSeq 2500 at the Toronto Center for Applied Genomics (TCAG). Bioinformatic analyses were done similarly, as previously described[36].

Quantitative droplet digital PCR (ddPCR). ddPCR was performed on temporal-vein samples using a QX200 Droplet Digital PCR System with QX200 ddPCR EvaGreen Supermix (BioRad). Three-step ddPCR was performed

under the following conditions: 95 °C (10 min), 94 °C (30 s), 62 °C (30 s), 72 °C (1 min) for 45 cycles, and 98 °C (10 min) using a *Lama2* guide screen 1F (RDC 1474) and 1R (RDC 1475). The data were analyzed using QuantaSoft Pro (Bio-Rad).

Bioinformatic analysis of the applicability of the exon inclusion strategy. Positions of all exon–intron junctions were obtained from GENCODE gene annotation gtf file (version 19, https://www.gencode-genes.org/releases/19.html) using customized scripts (https://github.com/hy09/splice_PAM_search). 5′-splice-site (+1) mutations were extracted from the Human Gene Mutation Data database (HGMD, version 20140404). Note that all of these are disease-causing mutations (DMs). By our definition, the first nucleotide of the intron is defined as the '+1' position. For each transcript, introns were extracted. For introns that are longer than 1,000 base pairs, only the first 1,000 base pairs were retained for analysis. For each gene, overlapping introns from different transcripts were merged. If a shorter intron overlaps with a larger one, then the shorter intron was used. For the introns containing a 5′ splice site (+1) mutation, we first examined whether a 'mutation-eliminating cut' site could be found. A mutation-eliminating cut is defined as when any of the available Cas9s cleavage site occurred after the following PAM sequences: NGG, NAG (*Streptococcus pyogenes*), NGA (*S. pyogenes* EQR variant), NGCG (*S. pyogenes* VRER variant), NNGRRT (*S. aureus*), NNNRRT (*S. aureus* KKH variant), NNAGAA (*Streptococcus thermophilus*), NNNNRYAC (*Campylobacter jejuni*), YG (*Francisella novicida*) could lead to a cut that results in the elimination of the mutation[14,27,29,35,37,38]. For mutations occurring at the +1 position, the cut has to be made immediately upstream of the +1 position ('cut at +1'). The search for PAM sequences was done on both strands. For the introns in which a mutation-eliminating cut could be found, we continued our search into the introns and examined whether a 'splice-site-generating cut' site could be found. A splice-site-generating cut site is defined as a cut that, upon re-ligation with the corresponding exon, can restore the 5′ donor site. We required the first two nucleotides ('gt' at +1 and +2 sites) to be restored. Finally, we filtered for splice-site-generating cut sites that were at least 50 bp away from the next exon to avoid interference with the splicing machinery.

Splice-site prediction. Splice sites were scored using (i) the MaxEntScan donor model, receiving in input 3 exonic bp and 6 intronic bp[30] centered at the splice site; (ii) the DeepScan donor model, a proprietary predictor developed at Deep Genomics, receiving in input 40 exonic bp and 40 intronic bp centered at the splice site. Both predictors were trained on the basis of human annotated splice sites and genome sequences and then were applied to the mouse sequence without further adjustments. MaxEntScan was previously shown to be one of the best-performing publicly available methods for evaluating the deleteriousness of splice-site-altering variants[39]. DeepScan is based on deep

neural network architecture[40], with convolutional layers and hidden layers consisting of rectified linear units trained using stochastic-gradient descent to distinguish sequences of annotated splice sites from decoy splice sites found in introns; it achieved 98% area under the receiver operating characteristic (ROC) curve for distinguishing held-out donor splice sites from held-out intronic sequences, with the core donor dinucleotide GT achieving a better performance than MaxEntScan (**Supplementary Table 4**).

To derive genome-wide statistics on donor splice-site prediction, all intronic GT sites were extracted for GENCODE Basic version 19 transcripts, using the reverse complement of the hg19/b37 human reference genome sequence for negative-strand transcripts; only sites for which the intronic categorization was not ambiguous were retained (i.e., intronic in all transcripts); sites were then categorized as annotated splice site donors or not. Finally, prediction models (MaxEntScan donor model, DeepScan donor model) were used to score these sites and derive genome-wide statistics at different prediction score thresholds. Exon statistics for reconstituted splice-site donors (percentage of exons with at least one GT site within 1 kb predicted as donor, when deleting the intervening intronic sequence and retaining the exon sequence; average number of such sites per exon) were calculated with respect to the longest transcript isoform of each gene in GENCODE basic version 19.

Statistical analysis. All statistical analyses were performed using GraphPad Prism (GraphPad software). Two-tailed Student's *t*-tests were performed to evaluate significant differences between two groups. The exact numbers of animals used in each analysis are presented as individual values in dot-plot graphs, and data are presented as average ± s.d., unless indicated otherwise. $P < 0.05$ was considered to be significant.

Data-availability statement. All data generated or analyzed during this study are included in this published article (and its supplementary-information files).

33. Kemaladewi, D.U. *et al*. Cell-type specific regulation of myostatin signaling. *FASEB J.* **26**, 1462–1472 (2012).
34. Vilquin, J.T. *et al*. Identification of homozygous and heterozygous dy2J mice by PCR. *Neuromuscul. Disord.* **10**, 59–62 (2000).
35. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
36. Wojtal, D. *et al*. Spell checking nature: versatility of CRISPR/Cas9 for developing treatments for inherited disorders. *Am. J. Hum. Genet.* **98**, 90–101 (2016).
37. Cong, L. *et al*. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
38. Mali, P., Esvelt, K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
39. Jian, X., Boerwinkle, E. & Liu, X. *In silico* prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).
40. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).