

# Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy

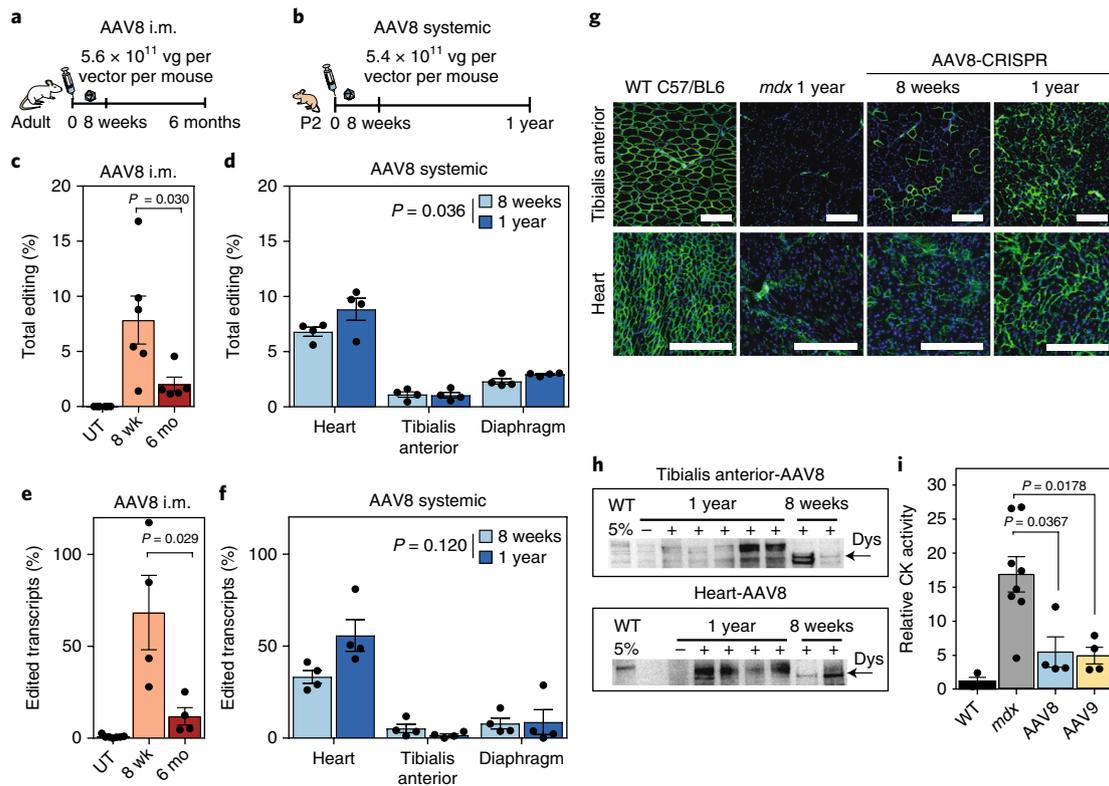
Christopher E. Nelson<sup>1,2</sup>, Yaoying Wu<sup>1</sup>, Matthew P. Gemberling<sup>1,2</sup>, Matthew L. Oliver<sup>1</sup>, Matthew A. Waller<sup>1,2</sup>, Joel D. Bohning<sup>1,2</sup>, Jacqueline N. Robinson-Hamm<sup>1,2</sup>, Karen Bulaklak<sup>1,2</sup>, Ruth M. Castellanos Rivera<sup>3</sup>, Joel H. Collier<sup>1</sup>, Aravind Asokan<sup>4,5</sup> and Charles A. Gersbach<sup>1,2,6\*</sup>

**Duchenne muscular dystrophy (DMD) is a monogenic disorder and a candidate for therapeutic genome editing. There have been several recent reports of genome editing in pre-clinical models of Duchenne muscular dystrophy<sup>1–6</sup>, however, the long-term persistence and safety of these genome editing approaches have not been addressed. Here we show that genome editing and dystrophin protein restoration is sustained in the *mdx* mouse model of Duchenne muscular dystrophy for 1 year after a single intravenous administration of an adeno-associated virus that encodes CRISPR (AAV-CRISPR). We also show that AAV-CRISPR is immunogenic when administered to adult mice<sup>7</sup>; however, humoral and cellular immune responses can be avoided by treating neonatal mice. Additionally, we describe unintended genome and transcript alterations induced by AAV-CRISPR that should be considered for the development of AAV-CRISPR as a therapeutic approach. This study shows the potential of AAV-CRISPR for permanent genome corrections and highlights aspects of host response and alternative genome editing outcomes that require further study.**

Duchenne muscular dystrophy (DMD) is a debilitating and prematurely fatal genetic disease caused by mutations in the *DMD* gene leading to the absence of dystrophin<sup>8,9</sup>. Despite recent clinical advancements<sup>10,11</sup>, a curative approach remains to be found. Adeno-associated viruses (AAVs) are being used as a gene delivery vector for recently initiated DMD clinical trials and for two approved gene therapy products and have been tested in more than 100 clinical trials<sup>12</sup>. Multiple groups are using AAVs to deliver genome-editing technologies to make permanent genetic modifications to treat disease, including the first human genome editing clinical trial using AAV that is currently underway using zinc finger nuclease technology<sup>13,14</sup>. Genome editing has been used to repair the *DMD* gene by exon deletion<sup>1–6</sup>, splice-site targeting<sup>15</sup> or homology directed repair<sup>6</sup> in mouse models of DMD and most recently in a canine model of DMD<sup>16</sup>. These studies show that genome editing restores dystrophin expression in mouse models of DMD, leading to an improvement in skeletal muscle function. The enthusiasm for a genome-editing strategy is based on the potential that a single administration can have life-long therapeutic benefits. However, published studies have focused on short-term restoration of dystrophin, typically assessed at 4–8 weeks after treatment. In this study, we treated mice with a dual-AAV system, consisting of one AAV encoding Cas9 and the other AAV encoding two guide

RNAs (gRNAs) that were designed to excise exon 23 from the *Dmd* gene in *mdx* mice. For viral packaging, we used the smaller 3.2-kb Cas9 derived from *Staphylococcus aureus* (SaCas9)<sup>17</sup>. We examined both AAV serotype 8 (AAV8) and 9 (AAV9) (Fig. 1a,b), which have differential tissue tropism for the heart, skeletal muscle, and liver in animal models that are not perfectly predictive of human tropism<sup>18</sup>. We examined adult and postnatal day 2 (P2) neonatal mice treated locally by intramuscular (i.m.) injection and systemically by intravenous facial-vein injection (FVI), respectively, for restoration of dystrophin expression (Fig. 1a,b). We adapted an unbiased deep-sequencing method for precise quantification of gene-editing efficiencies. Mice injected by i.m. injection as adults had a significant decrease in genome-editing levels over time (Fig. 1c and Extended Data Fig. 1). By contrast, systemically treated mice had a modest statistically significant increase in genome-editing levels over 1 year (Fig. 1d and Extended Data Fig. 1). The SaCas9 expression cassette was driven by a constitutive CMV promoter that is expressed in multiple muscle cell types, including striated muscle and muscle progenitors<sup>5</sup>. However, genome-editing events were also detected in other tissues, including liver, spleen, kidney and brain, as well as the testis at levels barely above the limit of detection (approximately 0.1%; Extended Data Fig. 2). Use of a myocyte-specific promoter could restrict editing to striated muscle nuclei<sup>6</sup>, but potentially at the cost of editing muscle progenitor cells. Analysis of *Dmd* mRNA transcripts by droplet digital PCR (ddPCR) showed the same trend as the genomic deletions with significant increases over time noted in cardiac muscles obtained from systemically treated mice (Fig. 1e,f). Sustained restoration of dystrophin was detected by immunofluorescence staining and western blot of cardiac and skeletal muscle from systemically treated mice for at least 1 year after a single administration (Fig. 1g,h and Extended Data Fig. 3). The restored dystrophin was slightly smaller in younger mice than in older mice, potentially owing to a smaller protein isoform produced at the early time point while nearly full-length dystrophin was detected at 1 year. Serum creatine kinase levels were reduced at 8 weeks after treatment in mice treated systemically as neonates, demonstrating protection from muscle damage by the restored dystrophin protein (Fig. 1i). Deep sequencing of the top ten predicted off-target sites showed no significant increase in off-target cutting after 1 year with slight activity above background noted for gRNA1 at off-target site 8 (gRNA1-OT8), as was previously identified following local administration<sup>4</sup> (Supplementary Tables 1 and 2).

<sup>1</sup>Department of Biomedical Engineering, Duke University, Durham, NC, USA. <sup>2</sup>Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. <sup>3</sup>Gene Therapy Center, University of North Carolina Chapel Hill, Chapel Hill, NC, USA. <sup>4</sup>Department of Surgery, Duke University Medical Center, Durham, NC, USA. <sup>5</sup>Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC, USA. <sup>6</sup>Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC, USA. \*e-mail: [charles.gersbach@duke.edu](mailto:charles.gersbach@duke.edu)

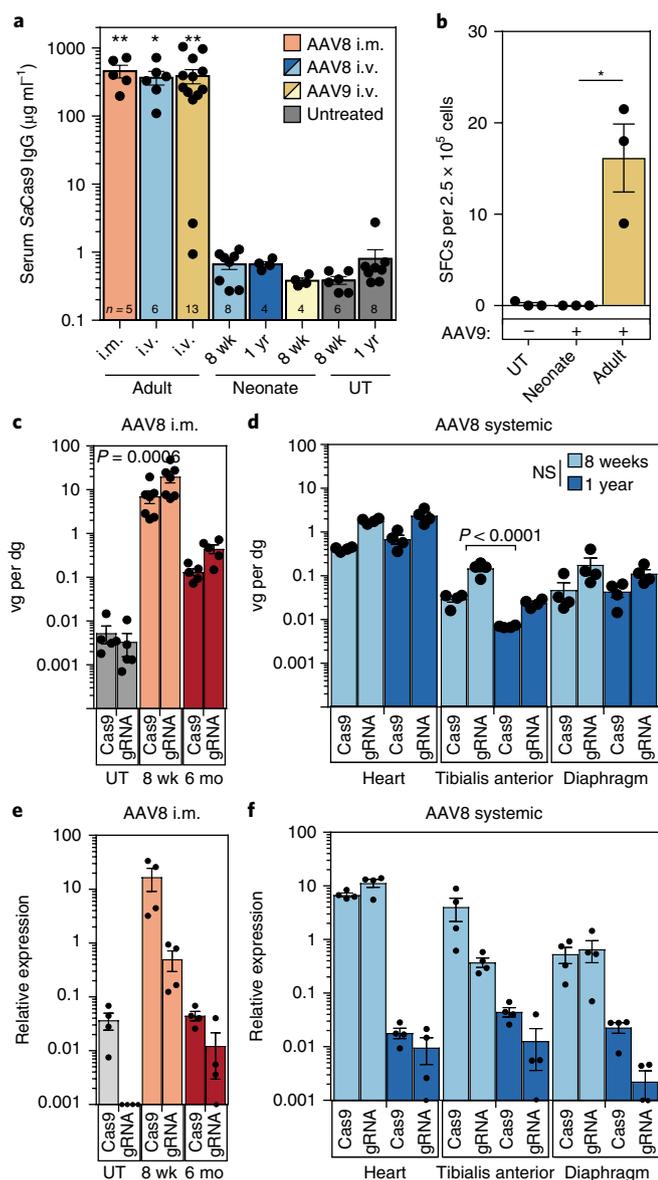


**Fig. 1 | Genome editing is sustained for 1 year in neonatal mice treated by intravenous administration.** **a,b**, Mice were treated as adults by AAV injection into the tibialis anterior (**a**) or systemically by FVI as neonates (**b**). vg, vector genomes. **c,d**, Quantification of total gene modification shows a significant decrease over 6 months following local administration (**c**;  $n = 6$ , 8 weeks;  $n = 5$ , 6 months; one-sided  $t$ -test) and a significant increase in neonates treated systemically (**d**;  $n = 4$ ; two-way ANOVA). UT, untreated. 8 wk, 8 weeks; 6 mo, 6 months. **e,f**, ddPCR shows the same trend for deletion of exon 23 from the transcript for local injections (**e**;  $n = 4$ ; one-sided  $t$ -test) and systemic injections (**f**;  $n = 4$ ; two-way ANOVA). **g**, Dystrophin expression is sustained in cardiac muscle and skeletal muscle 1 year after systemic administration into neonates. Scale bars, 200 μm. WT, wild-type mice. Histological images are available as Source data. **h**, Western blot confirms the presence of dystrophin (Dys) in skeletal and cardiac muscle. Full uncropped blots and are available as Source data. **i**, After 8 weeks, systemically treated neonatal mice show a significant decrease in creatine kinase (CK) ( $n = 3$ , wild-type mice;  $n = 8$ , untreated mice;  $n = 4$ , treated mice; one-way ANOVA with multiple comparisons correction). Data are mean  $\pm$  s.e.m.

An important consideration to long-term therapeutic benefit of *in vivo* genome editing is the host response to bacteria-derived Cas9 proteins. In our study, a humoral immune response was detected against the SaCas9 protein in nearly all mice injected as adults ( $n = 31$  out of 32 mice; Fig. 2a and Extended Data Fig. 4). By contrast, no humoral response against SaCas9 was detected in mice treated as neonates by FVI or intraperitoneal injections ( $n = 0$  out of 19 mice; Fig. 2a). A cellular response was detected by restimulation with SaCas9 to produce interferon (IFN) $\gamma$ -secreting T cells in mice treated as adults but not neonates, regardless of administration route (Fig. 2b and Extended Data Fig. 4). The *mdx* mouse model has an increased baseline number of infiltrating macrophages and neutrophils owing to muscle degeneration and inflammation<sup>19</sup>, which treatment with AAV-CRISPR has been shown to decrease<sup>4</sup>. Expression of FOXP3, granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin (IL)-1 $\beta$ , and IL-12 $\beta$  decreased relative to untreated *mdx* mice after i.m. injection. However, IFN $\gamma$  significantly increased approximately seven-fold after local injection relative to untreated *mdx* mice (Extended Data Fig. 4). By contrast, systemically treated adults and neonates showed no significant changes in these markers of inflammatory cell infiltration. Most AAV vector genomes remain episomal after cell entry and are stably maintained in non-dividing cells<sup>20</sup>. In this study, AAV vectors persisted between 8 weeks and 1 year in cardiac muscle but were significantly lost in skeletal muscle after i.m. or FVI injection (Fig. 2c,d).

Regardless, expression of SaCas9 mRNA and both gRNAs is almost absent after 6 months or 1 year by either route of administration (Fig. 2e,f), which may be the result of promoter silencing<sup>21</sup>. The host response to AAV-CRISPR will need to be carefully considered for future clinical development, including pre-existing immunity in humans<sup>22</sup>. We have previously shown CRISPR-based gene silencing elicits a Cas9-dependent host response that resolves without intervention *in vivo*<sup>23</sup>. Our data here indicates that a significant host response is avoided if AAV-CRISPR is administered at the neonatal stage. Although the P2 mice have an undeveloped immune system that can be exploited for antigen-specific tolerance including Cas9<sup>24–26</sup>, it is not yet clear to what extent this approach applies in newborn humans. Other methods that could be explored to avoid anti-Cas9 immune response include transient immunosuppression for the length of vector expression, induction of immune tolerance<sup>27</sup>, removal of T cell epitopes<sup>28</sup>, the use of self-limiting/cleaving vectors, or other transient delivery vehicles including non-viral vectors<sup>29</sup>.

The methods used to assess *in vivo* genome-editing efficiencies have typically been designed to quantify the frequency of expected genome-editing outcomes. Additionally, different methods often must be used to quantify the various possible editing outcomes. For example, PCR-based methods for deep sequencing can detect the formation of insertions and deletions (indel) after genome editing but cannot quantify gene deletions and do not capture larger structural changes that remove one or both primer sites. Previously, we used

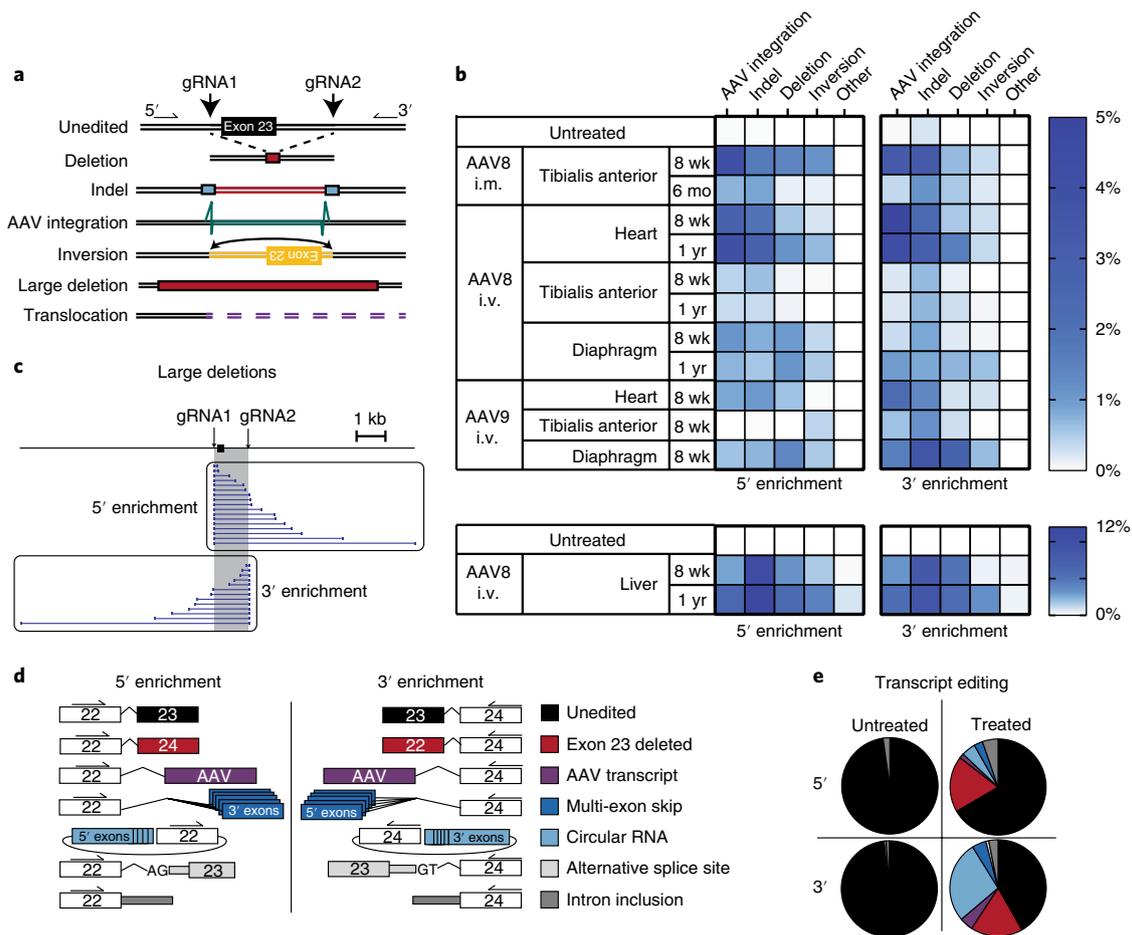


**Fig. 2 | Host response to AAV-CRISPR for DMD.** **a**, Antibodies against SaCas9 are detected in mice treated as adults but not in mice treated as neonates after 8 weeks or 1 year (one-way ANOVA with multiple comparisons  $*P < 0.05$ ,  $**P < 0.01$  compared with untreated mice at 8 weeks,  $n$  numbers as indicated). **b**, Mice injected as adults with an AAV encoding Cas9 have T cells that are stimulated by exposure to SaCas9 to produce IFN $\gamma$  as shown by enzyme-linked immune absorbent spot (ELISpot, one-sided  $t$ -test,  $*P = 0.0246$ ,  $n = 3$  mice per condition). SFCs, spot-forming colonies. **c,d**, A significant loss in total AAV vector genomes (vg) per diploid genome (dg) is detected in skeletal muscle following intramuscular injection and intravenous injection but not in cardiac muscle and diaphragm (two-way ANOVA with Tukey's multiple comparisons test,  $n = 7$ , i.m., 8 weeks;  $n = 5$ , i.m., 6 months,  $n = 4$ , all systemic groups). NS, not significant. **e,f**, Expression of both the Cas9 mRNA and gRNAs dissipates between the early and late time points. Skeletal muscle also shows lower gRNA expression than cardiac muscle (two-way ANOVA with Tukey's multiple comparisons test,  $n = 4$ , all groups). Data are mean  $\pm$  s.e.m.

ddPCR to quantify genetic changes including deletions<sup>4</sup>; however, ddPCR requires separate priming strategies to amplify each gene-editing outcome, including unedited alleles and different editing events, and cannot detect unexcised events. To comprehensively

map all possible genome-editing outcomes with an unbiased approach, we adapted Illumina's Nextera-transposon-based library preparation method for unbiased sequencing<sup>30</sup>. This method used a single genome-specific forward primer for target enrichment and a reverse primer specific for the DNA tag integrated by the transposon. In addition to genomic deletions, this method is sensitive to indel formation, inversion of exon 23 and surrounding introns and integration of the AAV genome (Fig. 3a). Using this method, we show that quantifiable and heterogeneous genome-editing events at the on-target *Dmd* locus occur, including deletions, inversions, indels, and AAV integrations in all treated mice (Fig. 3b and Extended Data Fig. 5). Importantly, no chromosomal translocations driven by off-target DNA cutting were detected in this experiment (estimated limit of detection of approximately 0.01%; Extended Data Fig. 5). The majority of deletion events were perfect deletions consistent with previous observations<sup>4,31</sup>. We detected a low prevalence ( $<0.5\%$ ) of large asymmetrical deletions (Fig. 3c and Supplementary Table 3), consistent with a previous report using long-read sequencing to monitor genome-editing outcomes in pluripotent cells in vitro<sup>32</sup>. However, our method cannot detect large deletions that remove both primer sites. The sequencing method used here is reproducible and matches indel quantification collected through a more standard next-generation sequencing method (Extended Data Fig. 6). We also applied this Nextera-based sequencing approach to cDNA of treated mice. This approach is sensitive to exon 23 removal and unexpected transcript changes including aberrant splicing (Fig. 3d and Extended Data Fig. 7). We detected removal of exon 23, changes in splicing including multi-exon skipping, putative circular RNA formation, and AAV splicing events (Fig. 3e and Supplementary Tables 4 and 5). Splicing events with the AAV vector genome contained canonical splice acceptors or donors (Extended Data Fig. 8 and Supplementary Table 6). Multi-exon skipping may lead to a partially functional or dysfunctional protein, depending on the change to the reading frame. The circular RNAs do not resemble a functional mRNA and will not be translated into protein and therefore are expected to have little biological importance. The relative enrichment of circular RNAs seen here may be caused by the stability of circular RNAs against exonuclease activity<sup>33</sup>. Transcript isoforms that contain partial AAV genomes have an unknown biological effect. The levels of exon 23 excision determined by this sequencing method are comparable to the results obtained by ddPCR analysis for quantification of exon 23 removal (Extended Data Fig. 8). Sequencing of the cDNA isolated at different time points indicated that the transcript isoform levels are sustained over 1 year (Extended Data Fig. 7).

AAVs are being used extensively as a delivery vector for CRISPR-Cas9 in preclinical studies to treat inherited diseases including DMD<sup>1,34</sup>. Although the safety of AAVs as a gene-delivery vehicle has been shown preclinically and through over 100 clinical trials, the potential genotoxicity of the combination of AAV and CRISPR requires further characterization. Here we adapted next-generation sequencing modalities to characterize unintended genome-editing events and AAV genome integrations. In this study, AAV typically integrated within the viral inverted terminal repeats (ITRs, 62%), resembling canonical integration<sup>35</sup>; however, insertions within the viral genome were also detected (38%) (Fig. 4a). Insertions that occur internally within the vector genome may be the result of vector truncations from AAV packaging or from AAV genome insertion during DNA repair. Separately, a primer specific to the AAV vector genome was used in conjunction with the same transposon-specific primer to map genome-wide AAV vector episome integration into the mouse genome (Fig. 4b). This showed that the targeted site within the *Dmd* gene was the preferential location for integration in both neonatal liver and cardiac muscle. In tissues that were analyzed 8 weeks after systemic delivery in neonatal mice, 94 AAV integration sites were identified in the liver and 72 sites in cardiac muscle, with the majority of integration events



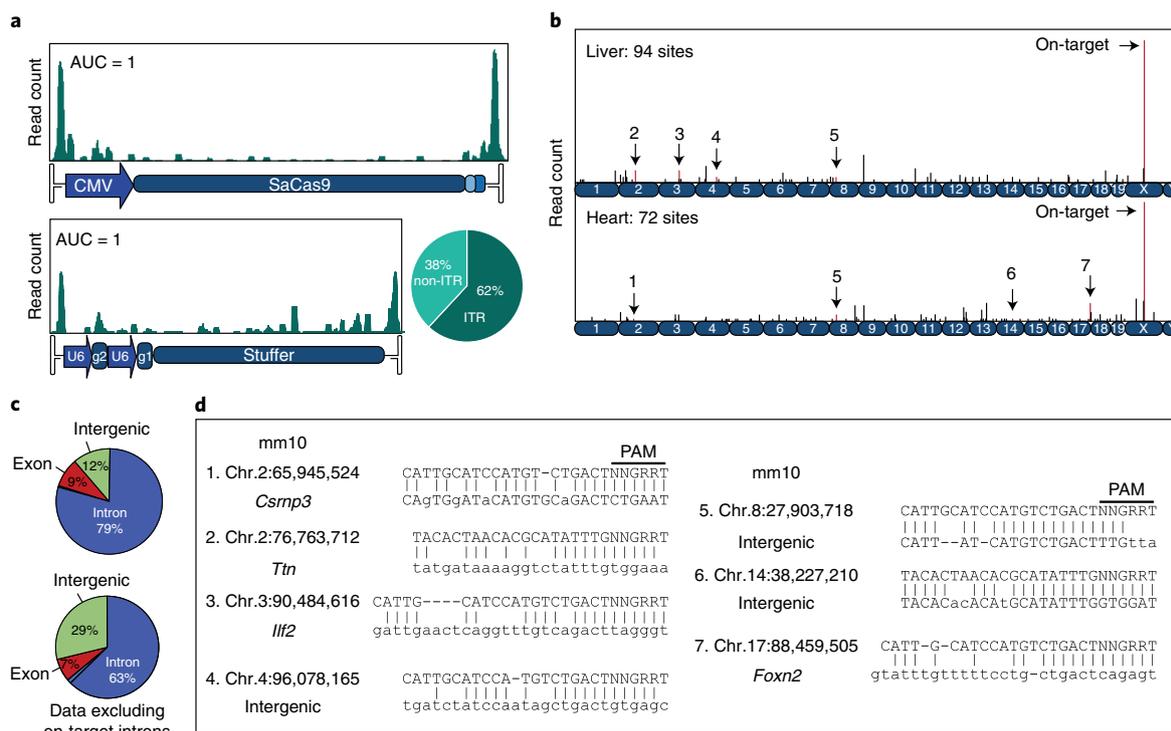
**Fig. 3 | In vivo genome editing generates diverse on-target genome modifications including AAV integrations and aberrant splicing.** **a**, Potential on-target genomic changes that resulted from targeted DNA cleavage are shown schematically. **b**, Unbiased analysis of the target site in multiple tissues after administration of AAV-CRISPR quantifies the level of each category of editing events. **c**, Large deletions around the target site were detected in the liver of mice treated with AAV8. **d**, Unbiased sequencing applied to the cDNA showed diverse transcript outcomes including aberrant splicing as shown schematically. **e**, Data from extracted cardiac muscle 8 weeks after administration show the proportion of transcript-editing events. **b,d**, Data were obtained from  $n = 4$  individual mice for treated and untreated groups.

occurring within introns of genes consistent with previous observations, including several previously identified integration sites<sup>36</sup> (Fig. 4c and Supplementary Table 7). Several putative gRNA off-target sites were also identified by AAV integration (Fig. 4b,d), including a previously predicted off-target site in an intergenic region of chromosome 14 for which there was no detectable activity by conventional targeted deep sequencing in the same samples<sup>4</sup> (Fig. 4c). This suggests that unbiased mapping of AAV integrations may be a more sensitive approach to determine the specificity of genome-editing reagents than typical methods.

In this study, the frequency of AAV integrations into the CRISPR-induced double-stranded break was higher than the intended deletion (Fig. 3b). AAV integration into targeted double-stranded breaks was reported more than a decade ago<sup>35</sup> and has also been applied as a therapeutic gene therapy approach<sup>14</sup>. AAVs can integrate into random breaks across the genome by non-homologous end-joining and can also be copied into target loci by homologous recombination without expression of nucleases<sup>37</sup>. Preclinical reports of hepatocellular carcinoma caused by genotoxicity of the vector have been controversial and risks can be managed by vector design<sup>36,38–40</sup>. AAVs are currently the gene delivery vehicle for more than 100 clinical trials targeting liver, skeletal muscle, cardiac muscle, central nervous system and other tissues with no reported adverse events

caused by genotoxicity of the vector. However, the induction of a novel DNA break by any genome-editing construct could potentially change the integration landscape and genotoxicity profile of the AAV (Fig. 4b–d). Additionally, each genome-engineering construct will have different genome-wide insertional mutagenesis profiles and should be carefully considered when developing vectors for therapeutic genome editing. Preclinical work can monitor *cis* activation of oncogenes and clonal expansion of AAV integration sites to reduce potential genotoxicity risks of genome-editing technologies delivered by AAVs<sup>36</sup>, analogous to efforts to characterize lentiviral vector integration, which also has an excellent safety profile in human clinical trials.

Important future preclinical developments will be focused on increasing the overall editing efficiency and increasing the proportion of the intended gene modification by optimizing delivery and the gene-editing strategy. This study further establishes the feasibility of permanent gene correction as a therapeutic approach for DMD and potentially other diseases. Despite the presence of a host response to Cas9 and persistent unintended genome modifications, AAV-CRISPR was well-tolerated for 1 year with no sign of toxicity, although much larger studies are required to confirm the absence of genotoxicity risk. Moreover, the restoration of dystrophin expression was sustained over this period. New developments



**Fig. 4 | AAV integrations into the *Dmd* locus and genome-wide. a**, AAV integrations were detected in the gRNA target sites. In total, 62% of integrations occurred within the ITRs and the remaining 38% occurred within the viral genome. AUC, area under the curve; g1, gRNA1; g2, gRNA2; stuffer, promoterless sequence included for AAV packaging. **b**, A similar approach was used to map AAV integrations genome-wide in both the liver and heart, 8 weeks after systemic administration in neonatal mice. The top hits include the two gRNA target sites flanking exon 23 of the *Dmd* gene, but numerous other AAV integration sites were detected. **c**, Genome-wide AAV integration sites were primarily located within the introns of genes. **d**, Possible sites of CRISPR off-target activity were identified by searching for sequences similar to the intended gRNA on-target site nearby the recovered genomic AAV integrations. PAM, protospacer adjacence motif.

to characterize safety and efficiency in larger animal models and mitigate the potential immune response will be crucial to translate this technology to treat genetic disease.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41591-019-0344-3>.

Received: 7 August 2018; Accepted: 21 December 2018;

Published online: 18 February 2019

### References

- Nelson, C. E., Robinson-Hamm, J. N. & Gersbach, C. A. Genome engineering: a new approach to gene therapy for neuromuscular disorders. *Nat. Rev. Neurol.* **13**, 647–661 (2017).
- Xu, L. et al. CRISPR-mediated genome editing restores dystrophin expression and function in *mdx* mice. *Mol. Ther.* **24**, 564–569 (2016).
- Long, C., Amoasii, L., Bassel-Duby, R. & Olson, E. N. Genome editing of monogenic neuromuscular diseases: a systematic review. *JAMA Neurol.* **73**, 1349–1355 (2016).
- Nelson, C. E. et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403–407 (2016).
- Tabebordbar, M. et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407–411 (2016).
- Bengtsson, N. E. et al. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nat. Commun.* **8**, 14454 (2017).
- Chew, W. L. et al. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).
- Flanigan, K. M. Duchenne and Becker muscular dystrophies. *Neurol. Clin.* **32**, 671–688 (2014).
- Hoffman, E. P., Brown, R. H. Jr. & Kunkel, L. M. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* **51**, 919–928 (1987).
- Chamberlain, J. R. & Chamberlain, J. S. Progress toward gene therapy for Duchenne muscular dystrophy. *Mol. Ther.* **25**, 1125–1131 (2017).
- Robinson-Hamm, J. N. & Gersbach, C. A. Gene therapies that restore dystrophin expression for the treatment of Duchenne muscular dystrophy. *Hum. Genet.* **135**, 1029–1040 (2016).
- Dunbar, C. E. et al. Gene therapy comes of age. *Science* **359**, eaan4672 (2018).
- Sharma, R. et al. In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777–1784 (2015).
- Laoharawee, K. et al. Dose-dependent prevention of metabolic and neurologic disease in murine MPS II by ZFN-mediated in vivo genome editing. *Mol. Ther.* **26**, 1127–1136 (2018).
- Amoasii, L. et al. Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Sci. Transl. Med.* **9**, eaan8081 (2017).
- Amoasii, L. et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* **362**, 86–91 (2018).
- Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
- Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nat. Rev. Genet.* **15**, 445–451 (2014).
- Spencer, M. J., Montecino-Rodriguez, E., Dorshkind, K. & Tidball, J. G. Helper (CD4<sup>+</sup>) and cytotoxic (CD8<sup>+</sup>) T cells promote the pathology of dystrophin-deficient muscle. *Clin. Immunol.* **98**, 235–243 (2001).
- Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral vectors for gene therapy: translational and clinical outlook. *Annu. Rev. Biomed. Eng.* **17**, 63–89 (2015).
- Brooks, A. R. et al. Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. *J. Gene Med.* **6**, 395–404 (2004).
- Charlesworth, C. T. et al. Identification of pre-existing adaptive immunity to Cas9 Proteins in humans. *Nat. Med.* <https://doi.org/10.1038/s41591-018-0326-x> (2019).
- Thakore, P. I. et al. RNA-guided transcriptional silencing in vivo with *S. aureus* CRISPR-Cas9 repressors. *Nat. Commun.* **9**, 1674 (2018).

24. Hu, C. & Lipshutz, G. S. AAV-based neonatal gene therapy for hemophilia A: long-term correction and avoidance of immune responses in mice. *Gene Ther.* **19**, 1166–1176 (2012).
25. Lee, E. K. et al. Long-term survival of the juvenile lethal arginase-deficient mouse with AAV gene therapy. *Mol. Ther.* **20**, 1844–1851 (2012).
26. Singh, K. et al. Efficient in vivo liver-directed gene editing using CRISPR/Cas9. *Mol. Ther.* **26**, 1241–1254 (2018).
27. Zhang, P. et al. Immunodominant liver-specific expression suppresses transgene-directed immune responses in murine Pompe disease. *Hum. Gene Ther.* **23**, 460–472 (2012).
28. Ferdosi, S. R. et al. Multifunctional CRISPR/Cas9 with engineered immunosilenced human T cell epitopes. Preprint at <https://www.biorxiv.org/content/early/2018/07/02/360198> (2018).
29. Nelson, C. E. & Gersbach, C. A. Engineering delivery vehicles for genome editing. *Annu. Rev. Chem. Biomol. Eng.* **7**, 637–662 (2016).
30. Giannoukos, G. et al. UDiTa<sup>SM</sup>, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* **19**, 212 (2018).
31. Iyombe-Engembe, J. P. et al. Efficient restoration of the dystrophin gene reading frame and protein structure in DMD myoblasts using the CinDel method. *Mol. Ther. Nucleic Acids* **5**, e283 (2016).
32. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
33. Memczak, S. et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
34. Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
35. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **36**, 767–773 (2004).
36. Chandler, R. J. et al. Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J. Clin. Invest.* **125**, 870–880 (2015).
37. Barzel, A. et al. Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* **517**, 360–364 (2015).
38. Bell, P. et al. No evidence for tumorigenesis of AAV vectors in a large-scale study in mice. *Mol. Ther.* **12**, 299–306 (2005).
39. Donsante, A. et al. AAV vector integration sites in mouse hepatocellular carcinoma. *Science* **317**, 477 (2007).
40. Gil-Farina, I. et al. Recombinant AAV integration is not associated with hepatic genotoxicity in nonhuman primates and patients. *Mol. Ther.* **24**, 1100–1105 (2016).

## Acknowledgements

This work has been supported by Sarepta Therapeutics, the Allen Distinguished Investigator Program through The Paul G. Allen Frontiers Group, the Muscular Dystrophy Association (grant MDA277360), a Duke–Coulter Translational Partnership Grant, a Duke/UNC-Chapel Hill CTSA Consortium Collaborative Translational Research Award, NIH grant R01AR069085, an NIH Director's New Innovator Award (DP2OD008586) and the Office of the Assistant Secretary of Defense for Health Affairs, through the Duchenne Muscular Dystrophy Research Program under awards W81XWH-15-1-0469 and W81XWH-16-1-0221. C.E.N. was supported by a Hartwell Foundation Postdoctoral Fellowship and the NIH Pathway to Independence Award (K99EB023979). J.N.R.-H. was supported by a National Science Foundation Graduate Research Fellowship and American Heart Association Predoctoral Fellowship (17PRE33350013).

## Author contributions

C.E.N. designed and conducted experiments, Y.W. provided expertise on assays to measure immune responses including ELISpot, M.P.G. conducted experiments and mouse procedures, M.L.O. purified proteins and performed histological analyses, M.A.W. completed qPCR experiments, J.D.B. performed histological analyses, J.N.R.-H. performed creatine kinase assays, K.B. performed mouse procedures, R.M.C.R. produced recombinant AAV, A.A., J.H.C. and C.A.G. designed experiments, C.E.N. and C.A.G. wrote and revised the manuscript.

## Competing interests

J.N.R.-H., C.E.N. and C.A.G. have filed patent applications related to genome editing for Duchenne muscular dystrophy. C.A.G. is an advisor for Sarepta Therapeutics, and a co-founder of and advisor for Element Genomics and Locus Biosciences. A.A. is a co-founder of and advisor for StrideBio.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-019-0344-3>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41591-019-0344-3>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to C.A.G.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

## Methods

**AAV preparation.** SaCas9- and gRNA-containing AAV constructs were generated as previously described<sup>3</sup>. In brief, a SaCas9 expression plasmid with a CMV promoter containing ITRs and a plasmid containing two gRNA expression cassettes driven by the human U6 polIII promoter were used to prepare recombinant AAV8 and AAV9. ITRs were confirmed by SmaI digestion before AAV production. Multiple batches of AAV8 and AAV9 were produced and titers measured by qPCR as previously described.

**In vivo administration of AAV-CRISPR.** All experiments involving animals were conducted with strict adherence to the guidelines for the care and use of laboratory animals of the National Institute of Health (NIH). All experiments were approved by the Institutional Animal Care and Use Committee (IACUC) at Duke University. The mouse strain C57BL/10ScSn-Dmdmdx/J (*mdx*) was obtained from Jackson laboratories. C57BL/10 mice were used as a wild-type control. Adult 8-week-old male mice were administered locally into the tibialis anterior muscle with 30  $\mu$ l of  $5.6 \times 10^{11}$  vector genomes per vector per mouse. Adult 8-week-old mice were administered intravenously with 200- $\mu$ l injections of  $2.7 \times 10^{12}$  vector genomes per vector per mouse. Two-day-old (P2) neonatal mice were administered intravenously through the facial vein<sup>41</sup> with  $5.4 \times 10^{11}$  vector genomes per vector per mouse. At set time points (Fig. 1a), mice were euthanized and the organs collected for experiments involving multiple skeletal muscles, cardiac muscles, other organs and serum.

**Genomic DNA extraction and next-generation sequencing.** Genomic DNA was extracted from various mouse tissues at defined time points by digestion in proteinase K and ALT buffer at 56 °C overnight while shaking. DNA was further extracted with the DNeasy kit (Qiagen). Endpoint PCR to confirm deletion was performed with AccuPrime Polymerase (Invitrogen) using primers DMDin22F and DMDin23R (Supplementary Table 8). Indel formation was detected by next-generation sequencing. Amplicons were produced by PCR using AccuPrime Polymerase (Invitrogen) and a series of primers for each locus (Supplementary Table 8). A second short-cycle PCR was used to add Illumina flowcell-binding sequences and experiment-specific barcodes (Supplementary Table 8). The resulting PCR products were sequenced with 150-bp paired-end reads on a Miseq instrument (Illumina). Indel analysis was performed using a local distribution of CRISPResso Pooled<sup>42</sup> using a 5-bp window and standard settings.

**Transposon-mediated target enrichment and sequencing.** Using a Nextera Tn5 transposon (Illumina), 100 ng to 1  $\mu$ g genomic DNA was tagged following the manufacturer's instructions, except the transposon was diluted 1:8 from specifications to encourage large fragment size (Extended Data Fig. 9). To enrich the targeted sequence, a single PCR reaction using a genome specific primer (DMDin22-Nextera-F or DMDin23-Nextera-R) was used paired with a reverse primer specific for the tag sequence inserted by the transposon (Nextera-R) for 25 cycles. Amplicons were purified with Ampure beads (Beckman Coulter) at 1.8 $\times$ . A short 10-cycle PCR was used to add experimental barcodes and Illumina adapter sequences. Amplicons were gel-purified selecting the fragment size shown in Extended Data Fig. 9. Sequencing was conducted on an Illumina Miseq using v.2 chemistry and 2  $\times$  150-cycle paired-end reads. Analysis was performed by aligning amplicons to the targeted locus and discarding misprimed sequences. Targeted amplicons ranged from 5 to 40% depending on the primer used (Extended Data Fig. 9). Reads were then aligned to the expected products, including deletions, inversions, AAV integrations and genome-wide translocations. Alignments to the AAV genome (Fig. 4a) used the NeedlemanWunsch algorithm with a GapOpenValue of 10. Some reads within the AAV ITRs were ambiguous and were randomly assigned to one of the two ITRs for alignment. This Nextera-based method is expected to reduce PCR-related bias from amplicon size; however, some bias may remain for the transposon selectivity<sup>30</sup>.

**Transcript evaluation and sequencing.** RNA was extracted using the Qiagen Universal kit. Subsequently, 1  $\mu$ g was used to perform First-strand cDNA synthesis using the Vilo kit (ThermoFisher) according to the manufacturer's instructions. cDNA was diluted 1:4 in ultrapure water, aliquoted and stored for further analysis. For transposon-based sequencing, second-strand cDNA synthesis was performed using Klenow fragment DNA polymerase (NEB). cDNA was treated with the Nextera Tn5 transposon at 1:8 the recommended concentration. Enrichment of the target transcript was performed by PCR using transcript-specific primers (Ex22-Nextera-F or Ex24-Nextera-R) and a constant reverse primer specific to the tag inserted by the transposon (Nextera-R). Amplicons were purified by Ampure beads at 1.8 $\times$  and a second 10-cycle PCR was used to add adapters and barcodes. Reads were aligned to predicted amplicons and mis-aligned reads were discarded. Reads were then aligned to expected products and unexpected products were identified and quantified. qPCR was conducted using QuantaBio PerfeCTa SYBR Green SuperMix using the primers listed in Supplementary Table 8.

**Western blot.** Frozen muscle biopsies were disrupted with mortar and pestle and suspended in RIPA buffer (Sigma) with a proteinase inhibitor cocktail (Roche) and incubated for 30 min on ice with intermittent vortexing. Samples were

centrifuged at 16,000g for 30 min at 4 °C and the supernatant was isolated and quantified with a bicinchoninic acid assay (Pierce). Protein isolates were mixed with NuPAGE loading buffer (Invitrogen) with 5%  $\beta$ -mercaptoethanol and boiled at 100 °C for 10 min. Samples were flash-frozen in liquid nitrogen and stored at -80 °C. Subsequently, 25  $\mu$ g of total protein per lane was loaded into a 10 well 4-12% NuPAGE Bis-Tris gel (Invitrogen) with MES buffer (Invitrogen) and electrophoresed for 30 min at 200 V. Proteins were transferred to a nitrocellulose membrane for 1 h at 400 mA at 4 °C in transfer buffer containing 1 $\times$  Tris-glycine, 10% methanol and 0.01% SDS. The blot was blocked overnight in tris-buffered saline with 0.1% Tween 20 (TBST) with 5% milk at 4 °C. The blot was probed with MANDYS8 (1:200, Sigma D8168) and rabbit anti-GAPDH (1:5,000, Cell Signaling 2118S). The blot was washed with TBST and probed with mouse or rabbit horseradish peroxidase-conjugated secondary antibodies (Sigma) for 30 min in 5% milk-TBST. Blots were visualized using Western-C ECL substrate (Biorad) on a ChemiDoc chemiluminescent system (Biorad). The full blots are shown in the Source data associated with this paper.

**Immunofluorescence staining.** Skeletal and cardiac muscles were dissected and embedded in OCT using liquid nitrogen-cooled isopentane and stored at -80 °C. Subsequently, 10  $\mu$ m sections were cut onto pretreated histological slides using a cryostat (Leica). Slides were washed in PBS and blocked in PBS supplemented with 5% FBS and 5% goat serum with 0.5% Triton X-100. Dystrophin was detected with MANDYS8 (1:200, Sigma D8168) in blocking buffer overnight at 4 °C. Slides were washed 3 $\times$  with PBS for 10 min and the secondary antibody was applied with DAPI (1:5000) for 30 min at room temperature. Slides were washed and mounted with ProLong Gold Antifade Mountant (Invitrogen) and imaged with an inverted microscope (Leica).

**Creatine kinase assay.** Serum creatine kinase was measured using a Liquid Creatine Kinase Reagent set (Pointe Scientific) following the manufacturer's instructions. In brief, 5  $\mu$ l of serum was diluted into 20  $\mu$ l of sterile PBS and incubated with reagent for 2 min and measured by absorbance every minute for three readings at 37 °C using a nanodrop spectrophotometer set for 340 nm readings. Calculations for total creatine kinase in U l<sup>-1</sup> were made according to the manufacturer's instructions and plotted relative to levels in serum samples from wild-type mice.

**Recombinant SaCas9 production.** A plasmid containing an IPTG-inducible bacterial SaCas9 expression cassette was transformed into Rosetta 2 cells (Millipore EMD) and plated on LB plates with 50  $\mu$ gml<sup>-1</sup> kanamycin and 30  $\mu$ gml<sup>-1</sup> chloramphenicol. Colonies were selected and grown in a starter culture overnight then in a 1 l culture for 4-6 h until the optical density at 600 nm reached 0.6-0.8. The temperature was reduced to 18 °C and induced with 0.2 mM IPTG and incubated for 12-16 h overnight. Cells were isolated and resuspended to 6 mlg<sup>-1</sup> in lysis buffer containing EDTA-free protease inhibitor cocktail (Roche), 1 mM PMSE and 1 mgml<sup>-1</sup> lysozyme (Sigma) and kept at 4 °C for the remainder of the protocol. Cells were sonicated 10 s on and 10 s off for 15 min. Cell debris was pelleted by centrifugation for 30 min at 16,000g at 4 °C. Protein was isolated with Ni-NTA agarose beads (Qiagen) following the manufacturer's instructions. DNA was removed with Sepharose (Sigma) collecting the protein flow-through. The protein was dialyzed overnight at 4 °C with 10 kDa MWCO dialysis tubing. Proteins were concentrated with Vivaspin 20 50 kDa MWCO spin filters (GE Healthcare). Samples were aliquoted, flash-frozen and stored at -80 °C. Proteins were analyzed by protein gel and western blot (see Source data).

**Antibody ELISA.** Antibodies against SaCas9 were detected by adapting previously published protocols<sup>7,43</sup>. In brief, recombinant SaCas9 protein was diluted in 1 $\times$  coating buffer (KPL) and used to coat a 96-well Nunc MaxiSorp plate with 0.5  $\mu$ g of protein per well. Proteins were incubated overnight at 4 °C to adsorb to the plate. Plates were washed three times 5 min each with 1 $\times$  wash buffer (KPL). Plates were blocked with 1% BSA blocking solution (KPL) for 1 h at room temperature. Standard curves for IgG were generated using an anti-SaCas9 antibody (Diagenode C15200230). Serum samples were added in dilutions ranging from 1:40 to 1:20,000 and plates were incubated for 5 h at 4 °C with shaking. Plates were washed three times 5 min each and 100  $\mu$ l of blocking solution containing goat anti-mouse IgG (Sigma 1:4,000) was added to each well and incubated at 1 h at room temperature. Plates were washed four times 5 min each and 100  $\mu$ l of ABTS ELISA HRP substrate (KPL) was added to each well. Optical density at 410 nm was measured with a plate reader.

**T cell ELISpot.** T cell ELISpots were performed as previously described<sup>44</sup>. In brief, splenocytes from AAV9-SaCas9-injected mice were isolated and purified using Lympholyte M (Cederlane). Then, 250,000 cells were mixed with either Cas9 protein (Applied Biological Materials) or cell medium only as a negative control, and subsequently plated in a 96-well ELISpot plate (Millipore, MSIPS4510) in 100  $\mu$ l per well. Stimulation was performed with 0.02  $\mu$ gml<sup>-1</sup> of Cas9 protein at 37 °C in a humidified incubator, 7% (v/v) CO<sub>2</sub>, for 40 h. Mouse IFN $\gamma$  ELISpot pairs (3321-3-250 and 3321-6-250) and streptavidin-alkaline phosphatase (3310-10) were purchased from Mabtech. Spots were developed using substrate Sigmafast

BCIP/NBT (Sigma, B5655). Plates were shipped to Zellnet Consulting and enumerated using a Zeiss KS ELISpot system. Full plates images are shown in Extended Data Fig. 4.

**Statistical methods and reproducibility.** Single comparisons were carried out using a MannWhitney *U*-test with *P* values reported in the figures (Figs. 1c,e and 2c,e). Multiple comparisons were made using Kruskal–Wallis test (Fig. 1i). Two-way ANOVA with post-hoc Tukey test was used to evaluate systemically treated groups with *P* values reported in the figures (Figs. 1d,f and 2d,f). qPCR for gene expression was evaluated by *t*-test with Holm–Bonferroni multiple comparison correction (Extended Data Fig. 4). All plotted dots are independent biological replicates (individual mice).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

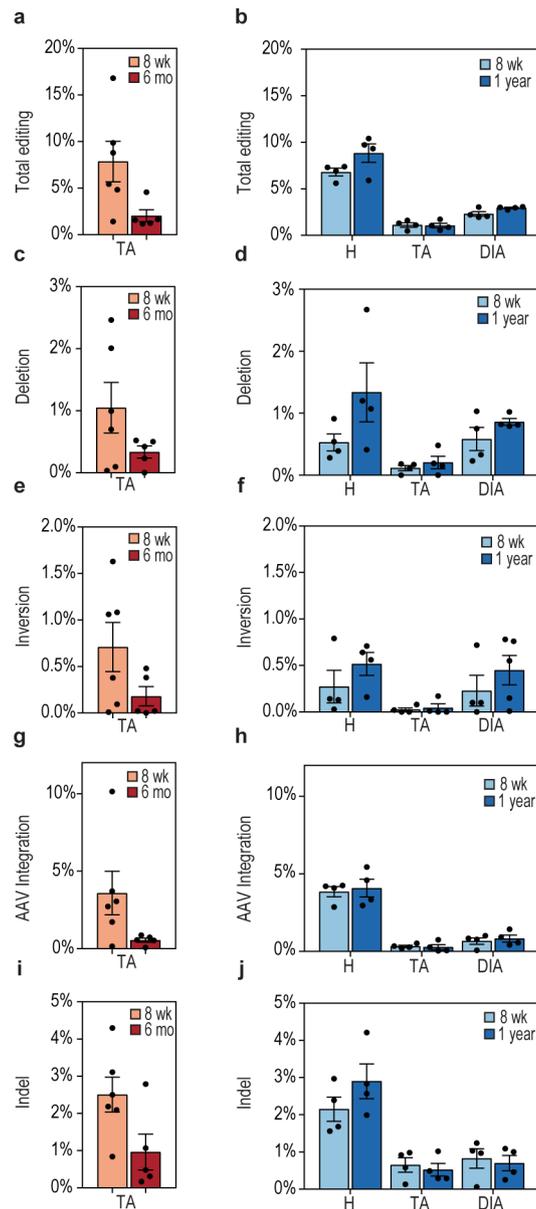
### Data availability

All custom code for reproducing Figs. 1c,d, 3b,d, and 4a–d have been made available online (<https://github.com/chrisnelsonlab/CRISPR-Tn5/>). All sequencing data used in this study have been deposited in the National Center for

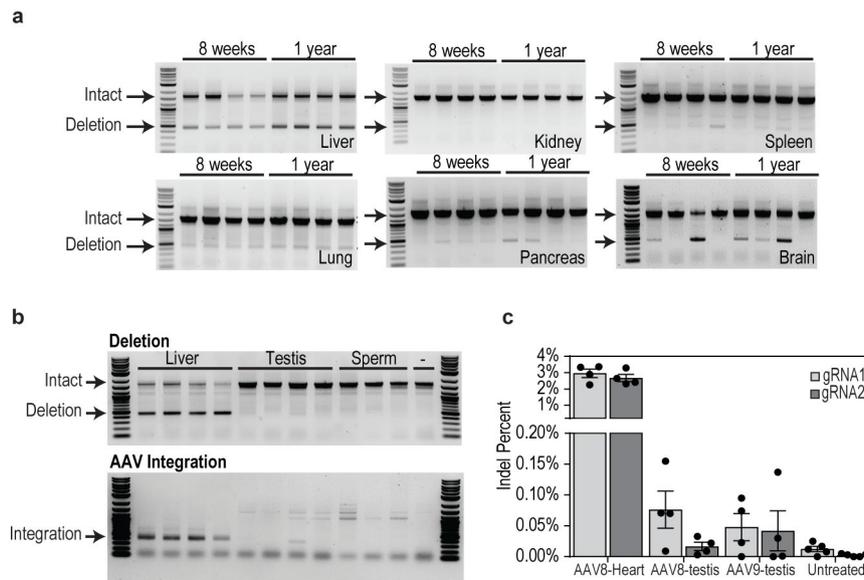
Biotechnology Information Sequence Read Archive (SRA) database (SRP157083). Full uncropped gels are included as Source data. All other relevant raw data are available from the corresponding author on request.

### References

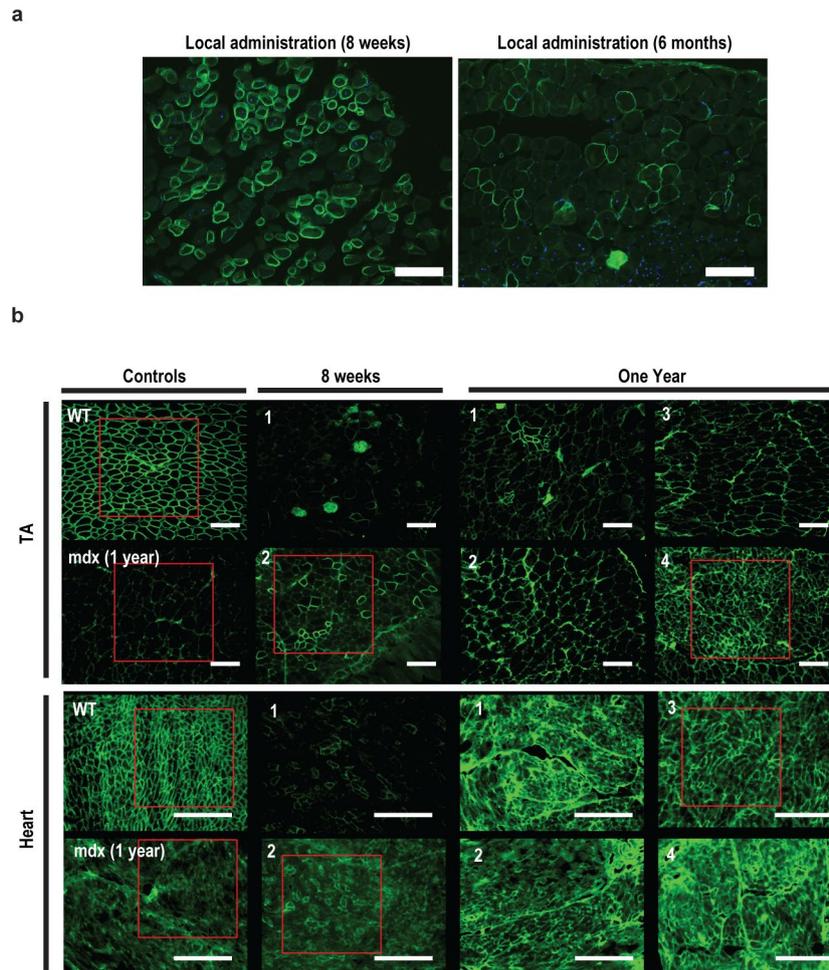
- Gombash Lampe, S. E., Kaspar, B. K. & Foust, K. D. Intravenous injections in neonatal mice. *J. Vis. Exp.* **93**, e52037 (2014).
- Pinello, L. et al. Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).
- Wang, D. et al. Adenovirus-mediated somatic genome editing of *Pten* by CRISPR/Cas9 in mouse liver in spite of Cas9-specific immune responses. *Hum. Gene Ther.* **26**, 432–442 (2015).
- Chen, J. et al. The use of self-adjuvanting nanofiber vaccines to elicit high-affinity B cell responses to peptide antigens without inflammation. *Biomaterials* **34**, 8776–8785 (2013).
- Zincarelli, C. et al. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16**, 1073–1080 (2008).
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).



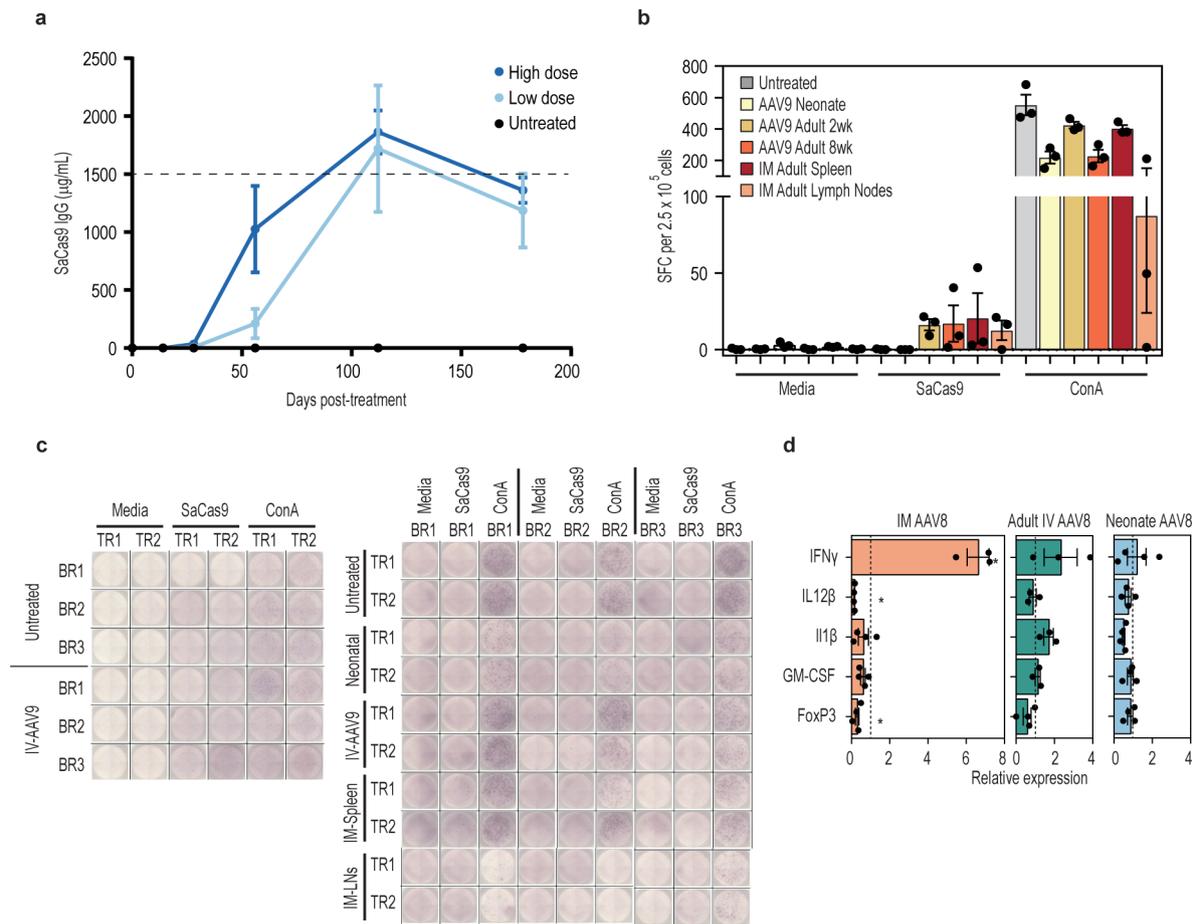
**Extended Data Fig. 1 | Illumina Nextera-based unidirectional sequencing shows diverse genome changes including deletions, AAV integration, inversions and indel formation. a**, Total editing for local administration. **b**, Total editing for systemic administration in neonates. **c,d**, Deletion frequency. **e,f**, Inversion frequency. **g,h**, AAV integration frequency. **i,j**, Indel frequency. Data are mean  $\pm$  s.e.m. Locally injected mice,  $n=6$ ; systemically injected mice,  $n=4$ .



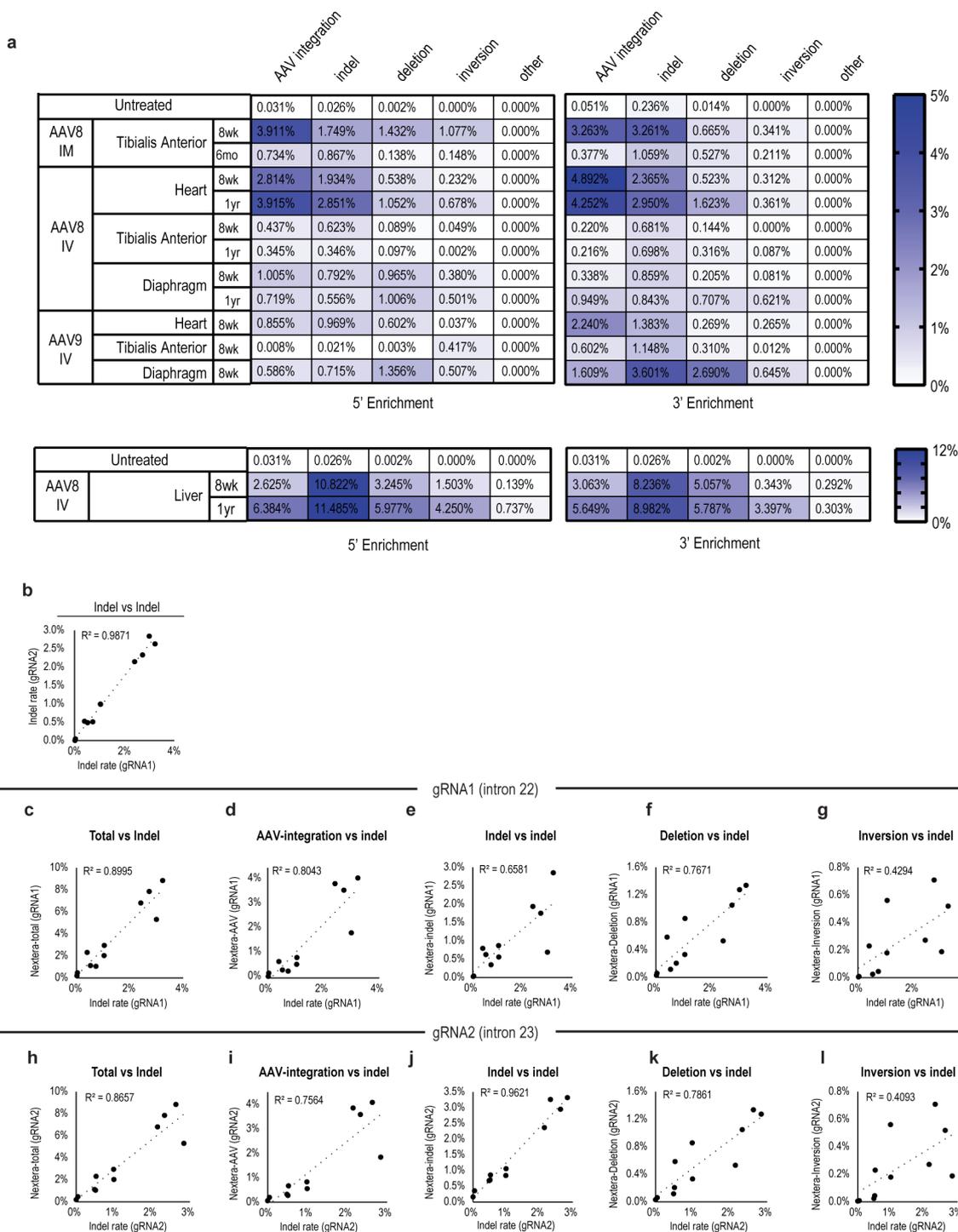
**Extended Data Fig. 2 | On-target genome editing activity in somatic and germline tissues. a**, Deletion PCR across the dystrophin locus shows Cas9 activity in multiple somatic tissues, including the liver, spleen, lung, pancreas and brain, which all show evidence of targeted gene deletion. There is no detectable deletion in the kidney samples. Cas9 expression is driven by a constitutive CMV promoter. This result is consistent with AAV8 tissue tropism<sup>45</sup>. The off-target tissue-editing experiment was conducted once. **b**, Deletion PCR of genomic DNA from the testis is mostly negative and undetectable in sperm. AAV integration was only detected in one testis sample and no sperm samples. The germline experiment was conducted once. **c**, Deep sequencing of testis DNA indicates low levels of indel formation for both AAV8- and AAV9-injected mice that were injected as neonates. Data are mean  $\pm$  s.e.m.  $n = 4$  mice in all treated groups;  $n = 5$  untreated mice.



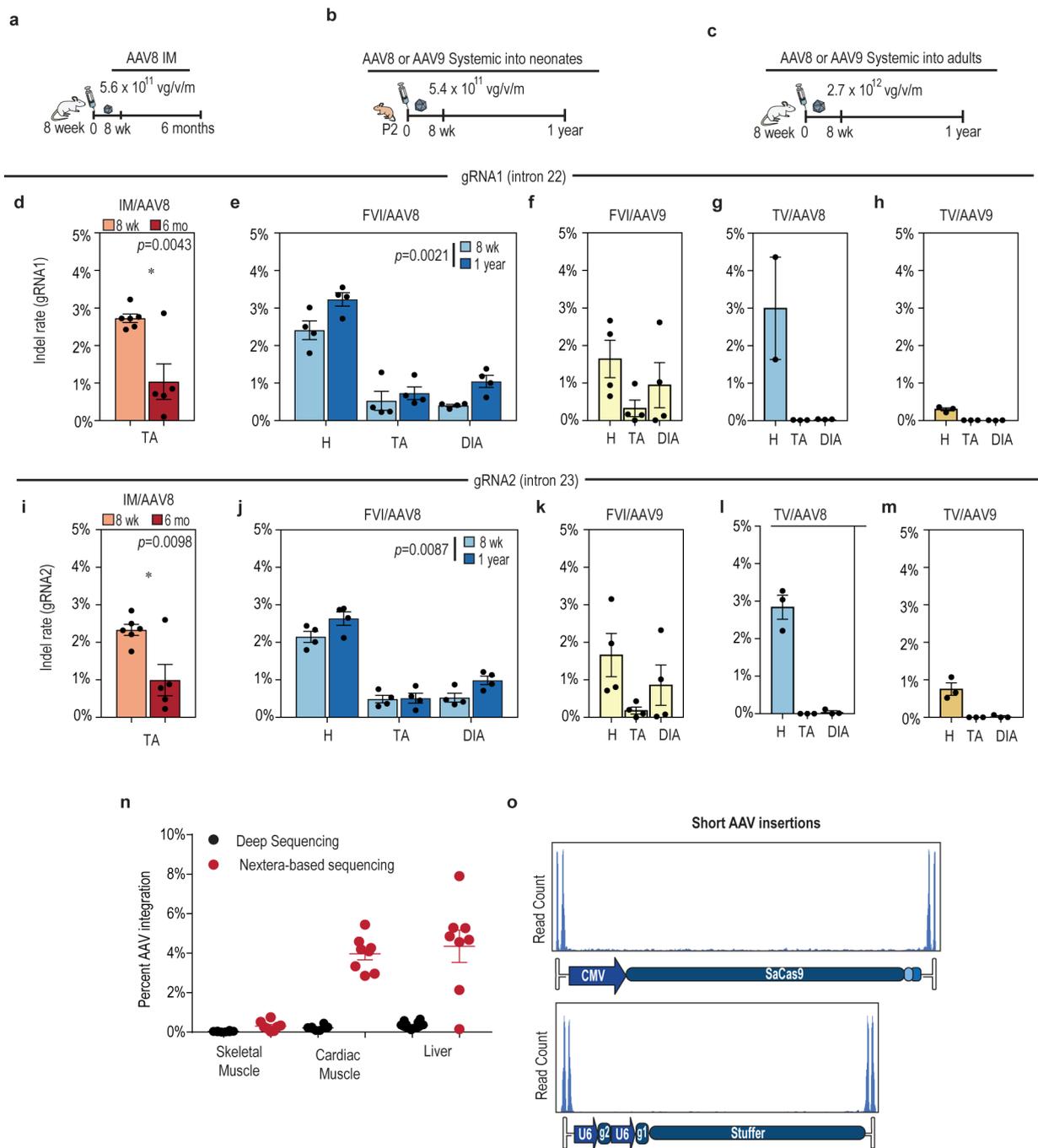
**Extended Data Fig. 3 | Complete panel of histology shows a decrease in dystrophin staining following local administration and an increase in dystrophin at the 1-year time point in systemically treated mice. a**, The histology images indicate a reduction in dystrophin after local injections, consistent with genomic and transcript data. **b**, Systemic samples show increased dystrophin expression after 1 year. Increased background at the 1-year time point may be a result of fibrosis in the tissue at the later time point. Representative images shown in Fig. 1 are highlighted in red. Scale bars, 200  $\mu\text{m}$ . Dystrophin-restoration experiments were conducted once for each treatment group.



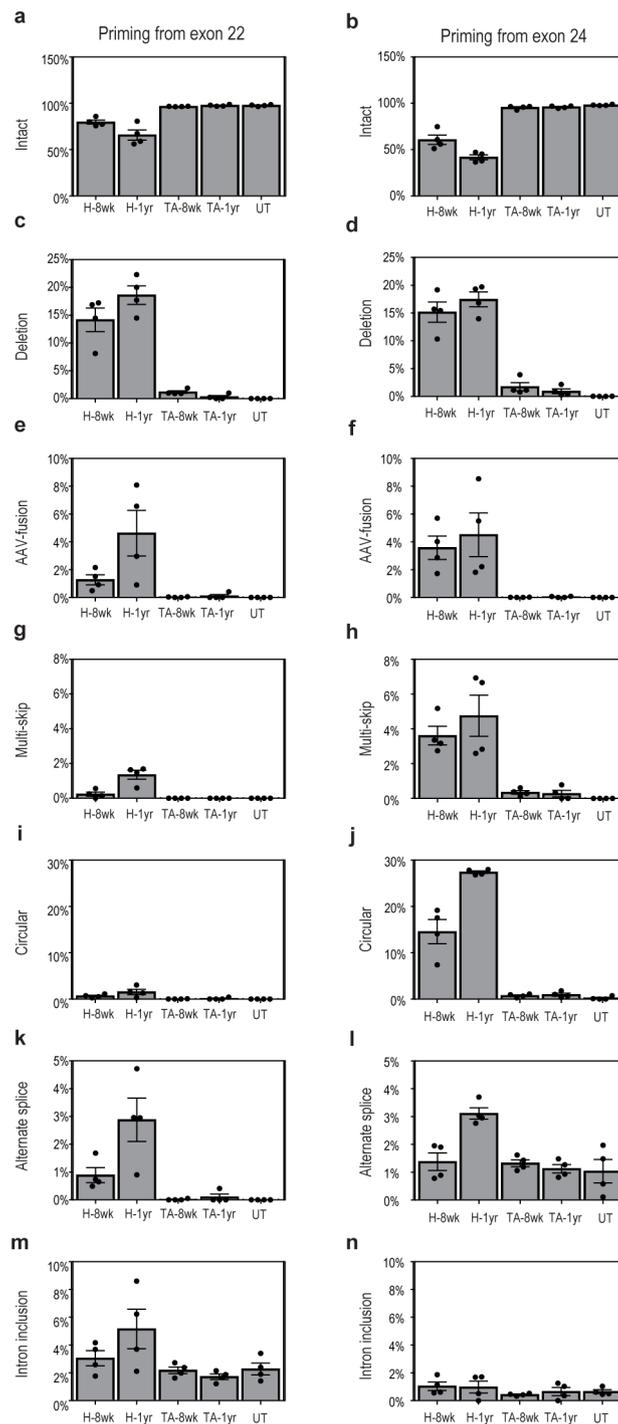
**Extended Data Fig. 4 | AAV delivery of SaCas9 elicits humoral and cellular immune responses.** **a**, Mice were administered with an AAV carrying a deactivated, nuclease-null SaCas9 transcriptional repressor (dSaCas9-KRAB) from a related study<sup>23</sup>. Serum was collected at 4, 8, 16 and 26 weeks. The IgG response invariably developed by 8 weeks in all tested mice and continued to increase until the 16-week time point. Data are mean  $\pm$  s.e.m. ( $n=4$  individual mice). The dotted line indicates the end of the linear range of the standard. **b**, ELISpot shows T cell responses in treated adults but not neonates regardless of administration route. Spot-forming cells (SFCs) were detected in mice injected with AAV9-SaCas9 in adults at 2 weeks and 8 weeks after systemic administration, as well as 8 weeks after intramuscular injection. SFCs were not detected in mice treated as neonates. **c**, Complete image panel of ELISpot data. Two separate plates are shown. Data are mean  $\pm$  s.e.m.  $n=3$  individual mice, biological replicates (BR) were used with two technical replicates (TR) as individual isolates from each mouse. **d**, Mice administered locally show increased IFN $\gamma$  and reduced FOXP3 and IL-12 $\beta$  expression, whereas mice administered systemically as adults or neonates show no significant changes. Statistics calculated compared to untreated,  $t$ -test with Holm-Bonferroni multiple comparisons correction was used.  $n=4$ , IM-AAV8;  $n=3$ , IV-AAV8;  $n=4$ , Neonate-AAV8.



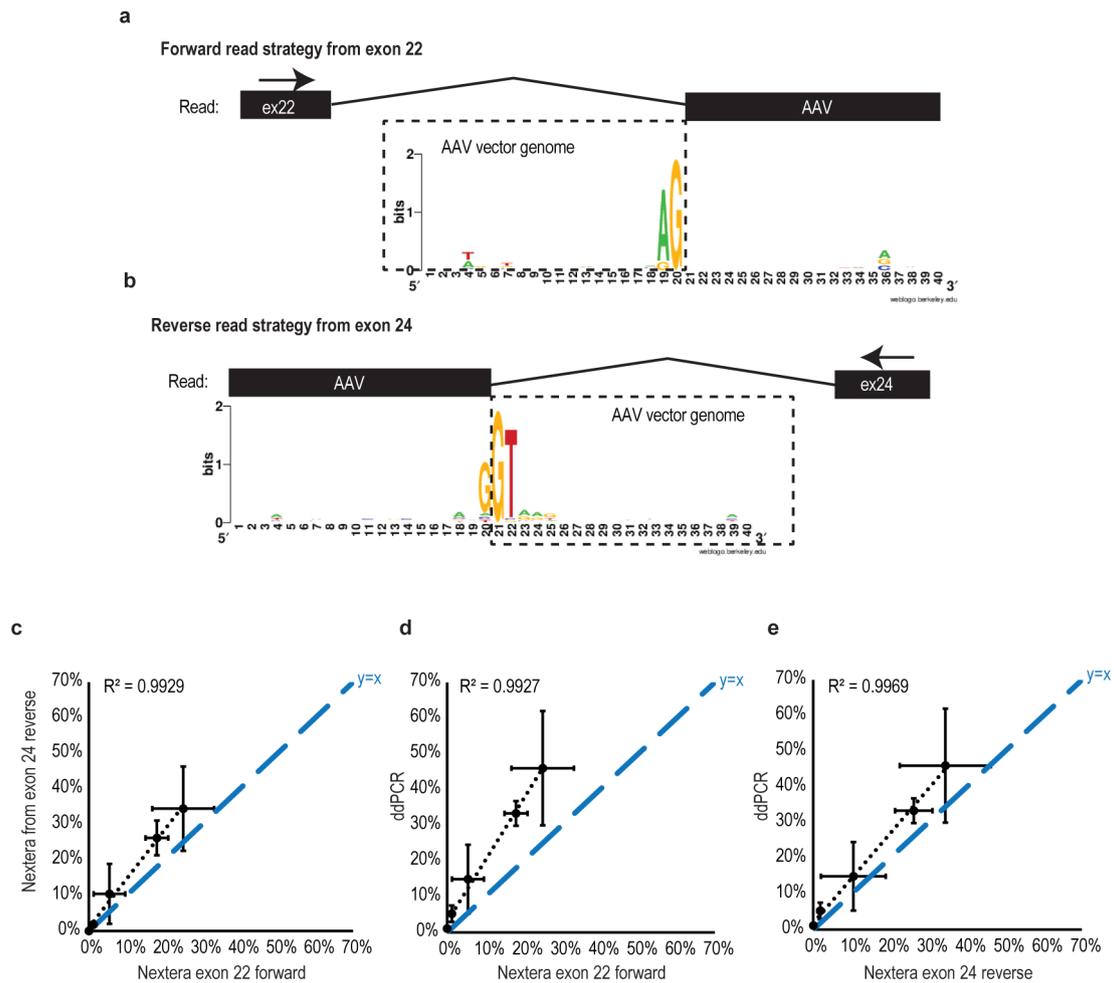
**Extended Data Fig. 5 | Complete quantitative and reproducibility data for the Illumina Nextera-based unidirectional sequencing measures.** Data associated with Fig. 3b. **a**, Quantitative data for genome-editing measurements are an average of  $n = 4$  individual mice. Skeletal and cardiac muscle are shown on a separate scale from liver samples. **b**, Comparison of deep sequencing for both gRNAs. **c–g**, Comparison of indel rates for gRNA1 to identify alternate modifications. **h–l**, Comparison of indel rates of gRNA2 with alternate modifications. Rare events have poorer correlations. An estimated limit of detection is given based on the inversions detected that range between 0.1% and 0.2%. The limit of detection could be decreased with more input DNA and increased number of reads to detect more rare events, possibly including translocations.



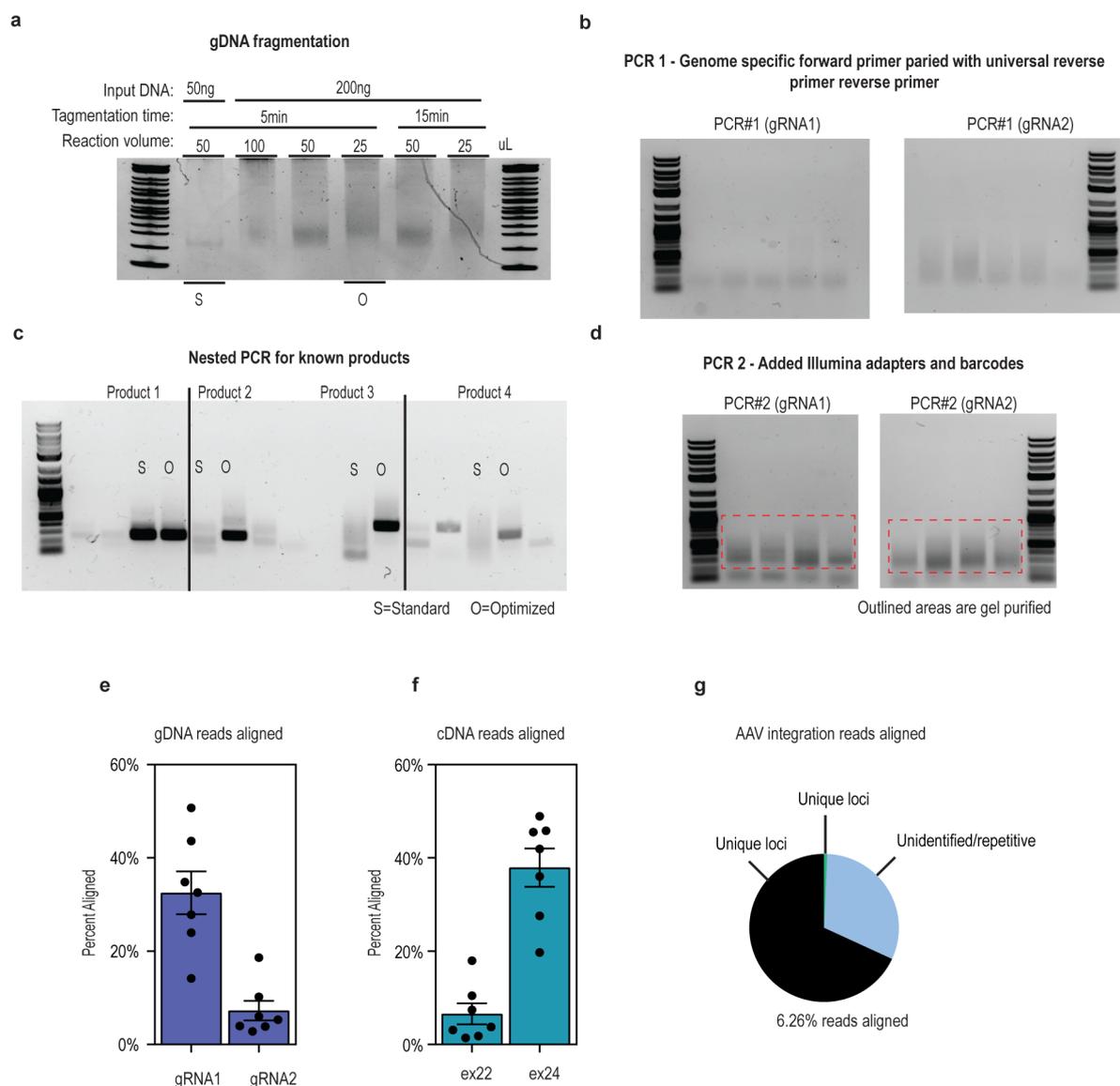
**Extended Data Fig. 6 | Deep sequencing of target loci for gRNA1 and gRNA2 in multiple tissues and treatment routes show indel formation and short AAV insertions.** **a**, Mice treated as 8-week-old adults by injection into the tibialis anterior were euthanized and tissues were collected at 8 weeks and 6 months after a single administration. **b**, Systemic administration in neonates by FVI of AAV8 or AAV9 was followed by analysis at 8 weeks and 1 year. **c**, Systemic administration in adults by tail-vein injection was followed by tissue collection at 12 weeks after the administration. **d**, Local administration with AAV8 ( $n=6$ ,  $n=5$ , two-tailed  $t$ -test). **e**, Systemic administration with AAV8 in neonates ( $n=4$ , two-way ANOVA). **f**, Systemic administration in neonates with AAV9 ( $n=4$ ). **g**, Tail-vein administration in adults with AAV8 ( $n=3$ ). **h**, Tail-vein administration in adults with AAV9 ( $n=3$ ). **i–m**, The same administration at the gRNA2 loci. **n**, Small AAV insertions were detected by deep sequencing for insertions that range from 10 to 45 bp in length. These insertions account for a small subset of integrations detected by Nextera-based sequencing. Nextera-based sequencing shows a higher detection rate of AAV genome insertions ( $n=8$ ). **o**, Short AAV insertions detected by indel sequencing are almost exclusively located within the ITR regions of the AAV vector genome. Data are mean  $\pm$  s.e.m.



**Extended Data Fig. 7 |** Illumina Nextera-based unidirectional sequencing of cDNA shows transcript changes over time in systemically administered neonates. There are notable differences in the number of circular RNA events and multi-skipping events when sequencing from the forward or reverse direction, which indicates that alternative splicing may be preferred in reverse direction. a-b) intact unedited transcripts, c-d) exon 23 deleted transcripts, e-f) AAV-fusion transcripts, g-h) transcripts with multiple skipped exons, i-j) circular transcripts, k-l) transcripts with alternative splicing, m-n) and transcripts with intron inclusion. Data are mean  $\pm$  s.e.m. ( $n=4$  for all samples).



**Extended Data Fig. 8 | Nextera-based sequencing reveals dystrophin-AAV transcript fusions and is a reproducible method. a**, Web logo map of nucleotide preference for splicing of dystrophin transcript to AAV vector genome shows canonical splicing is preferred. The forward read strategy priming from exon 22 shows that dystrophin-AAV splice fusions prefer an AG as the canonical splice acceptor. The sequencing read is shown as a black box. **b**, Similarly, the reverse priming strategy shows the preference for the canonical GT splice donor before the AAV-dystrophin fusion. The dotted-line box is not in the sequencing read so the AG or GT are revealed by alignment with the vector genome. Web logo maps were generated with the online tool<sup>46</sup> at <https://weblogo.berkeley.edu>. **c**, Deletions as measured by the Nextera method show a higher estimation by the reverse-sequencing method than the forward-sequencing method. **d,e**, ddPCR measures higher levels of gene deletion than either Nextera-based strategy. All comparisons were consistent ( $R^2 > 0.99$ ). The right Nextera-based method had an order of magnitude higher read count and there is potential bias for transposon recombination.  $y = x$  is plotted as a blue dotted line. Data are mean  $\pm$  s.e.m., all data are from  $n = 4$  mice.



**Extended Data Fig. 9 | Optimization of the Nextera-based sequencing method.** **a**, Multiple conditions were tested to find an optimized protocol. The standard method from the manufacturer's protocol is shown as 'S' and the optimized condition that we identified as 'O'. **b**, To test whether the random tagmentation works, a nested PCR was performed after the first PCR. The nested PCR used primer-binding sites of known amplicon size to detect the presence of expected products, including the unmodified target locus and the intended deletion. Only the optimized condition revealed all four predicted amplicons (5' and 3' enrichment for both products). We suspect that the standard condition generates fragments that are too short and lose the test-primer-binding site. By contrast, the optimized protocol generates longer DNA fragments that maintain the primer-binding site and would be better suited for unbiased sequencing. Optimization was performed only once. **c**, Gel showing amplicons after the first PCR. **d**, Gel showing amplicons after adding barcodes and adapters. Bands were purified within the outlined box. Gel images are representative of each sample analysed by deep sequencing ( $n=3$  independent experiments). **e**, Each method shows varying quantities of reads aligned to the reference because of mispriming. gDNA reads show that the gRNA1-intron22 strategy had a higher percentage of reads aligned to the genome ( $n=7$ ). **f**, The cDNA method shows a higher percentage of aligned reads for the exon 24 reverse priming strategy ( $n=7$ ). **g**, Sequencing out of the AAV shows 6.26% of reads aligned to the reference with the majority of aligned reads consisting entirely of AAV vector episomes with no novel junctions, represented in black. In blue are reads that did not align to the mouse reference genome or were within repetitive regions that made identification impossible. Green labels indicate the 0.04% of reads that aligned to unique loci within the mouse genome. These reads are listed in detail in Supplemental Table 1, the majority of the reads are in the targeted *Dmd* locus. Data are mean  $\pm$  s.e.m.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Sequencing data was collected by illumina sequencing.

Data analysis

Targeted deep sequencing was analyzed through online software (CRISPResso available from <http://github.com/lucapinello/CRISPResso>). Further interpretation was made using a custom matlab pipeline. Nextera-based sequencing and AAV integration sequencing used custom matlab code to align all reads the genome and then to a list of the expected products. For genome-wide AAV integration, 3' ends of the reads were aligned to the mm10 genome using BLAT. Matlab-based code can be made available upon request and has been uploaded to <https://github.com/chrisnelsonlab/CRISPR-Nextera/>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data has been deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) database (SRP157083). All other raw and interpreted data is either available in the supplemental information or upon request. Figure 3 and Figure 4 have data derived from SRP157083.

## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined based on our previous work for statistical significance. For biomolecular assays, n=4 was used as a minimum.
Data exclusions	Sequencing data - reads that did not align to the reference genome or expected products were excluded.
Replication	Multiple mice were examined for each endpoint. All findings can be reproduced with the same methodology in the multiple mice.
Randomization	Randomization was not used in this study. Mice were allocated into groups by cage or by litter in the case of neonatal injections. Covariates are not expected to affect dystrophin restoration in the mdx mouse. To help control for variation, all samples for each individual assay were processed by the same person.
Blinding	Single blinding was used for CK assay to eliminate potential bias in completing readings. Blinding was not necessary for other assays as bias cannot alter the outcome and all raw data is presented (i.e. gels/sequencing).

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials	Recombinant AAV used in this study can be produced commercially. Production of the AAV transfer plasmids is described in the methods and similar plasmids are available for purchased through Addgene. Specific plasmids used in this study can be made available upon request.
----------------------------	---

## Antibodies

Antibodies used	SaCas9 antibody (Supplier: diagenode, catalog number: C15200230, clone name: S. aureus CRISPR/Cas9 monoclonal antibody,
-----------------	---

Dilution: 1:4000 WB).

Mandys8 (Supplier: Sigma, Catalog number D8168, clone name: Mandys8, Dilution: 1:200 WB and IF)

#### Validation

Both antibodies have validation on their website available below. Mandys8 has an extensive record. The SaCas9 antibody was tested in our lab by IF in cell culture and western blot from bacterial lysate and transfected cell culture recreating the validation listed on the diagenode website (see Supplemental Figure 20).

SaCas9 (<https://www.diagenode.com/en/p/s-aureus-crispr-cas9-monoclonal-antibody>)

Mandys8 (<https://www.sigmaaldrich.com/catalog/product/sigma/d8168?lang=en&region=US>)

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

#### Laboratory animals

The study involved laboratory animals.

(1) Species: Mouse

(2) C57/Bl6 wild type mice and mdx model of DMD

(3) Males were used owing to the X-linked nature of the gene but females were also used as breeding pairs or when gender was not determined when injecting neonatal mice.

(4) Mice from two days of birth to one year are reported

#### Wild animals

The study did not involve wild animals

#### Field-collected samples

The study did not involve field-collected samples