Megabase-Scale Transgene De-Duplication to Generate a Functional Single-Copy Full-Length Humanized DMD Mouse Model.

Yu C. J. Chey[1,2], Mark A. Corbett[3], Jayshen Arudkumar[1,2], Sandra G. Piltz[1,2,4], Paul Q. Thomas[1,2,4*], Fatwa Adikusuma[1,2*]

1 School of Biomedicine and Robinson Research Institute, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide SA, Australia.

2 Genome Editing Program, South Australian Health and Medical Research Institute (SAHMRI), Adelaide SA, Australia.

3 Adelaide Medical School and Robinson Research Institute, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide SA, Australia.

4 South Australian Genome Editing (SAGE) Facility, SAHMRI, Adelaide SA, Australia.

* Corresponding authors:

P.Q.T.: paul.thomas@adelaide.edu.au

F.A.: fatwa.adikusuma@adelaide.edu.au

Abstract

The development of sequence-specific precision treatments like CRISPR gene editing therapies for Duchenne Muscular Dystrophy (DMD) requires sequence humanised animal models to enable the direct clinical translation of tested strategies. The current available integrated transgenic mouse

model containing the full-length human *DMD* gene, Tg(DMD)72Thoen/J (hDMDTg), has been found to have two copies of the transgene per locus in a tail-to-tail orientation, which does not accurately simulate the true copy number of the *DMD* gene. This duplication also complicates analysis when testing CRISPR therapy editing outcomes, as large genetic alterations and rearrangements can occur between the cut sites on the two transgenes. To address this, we performed long read nanopore sequencing on hDMDTg mice to better understand the structure of the duplicated transgenes. Following that, we performed a megabase-scale deletion of one of the transgenes by CRISPR zygotic microinjection to generate a single-copy, full-length, humanised DMD transgenic mouse model (hDMDTgSc). Functional, molecular, and histological characterisation shows that the single remaining human transgene retains its function and rescues the dystrophic phenotype caused by endogenous murine *Dmd* knockout. Our unique hDMDTgSc mouse model can potentially be used for the further generation of DMD disease models that would be better suited for the pre-clinical assessment and development of sequence specific CRISPR therapies.

Acknowledgements

Research was performed at SAHMRI on Kaurna Country. We pay respects to the Kaurna people of the Adelaide Plains and to all Aboriginal and Torres Strait Islander people. SAHMRI is committed to embracing knowledge and culture as we continue our working journey to incorporate Aboriginal health research across all our themes and further reconciliation.

This research is supported by the SAHMRI core facilities: Bioresources, Light Microscopy and Laser Capture, Histology and Cryogenics. Generation of mouse strains was carried out with the assistance of the South Australian Genome Editing (SAGE) Facility, the University of Adelaide and the South Australian Health and Medical Research Institute (SAHMRI). SAGE is supported by Phenomics

Australia. Phenomics Australia is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) program. We would also like to thank Melissa White for her comments and suggestions in writing this paper.

Funding

FA is supported by the Discovery Early Career Researcher Award (DECRA), Australian CSIRO Synthetic Biology Future Science Platform and the Emerging Leaders Development Award of Faculty of Health & Medical Science, University of Adelaide.

Conflict of interest statement: The authors declare that there is no conflict of interest.

Introduction

Duchenne Muscular Dystrophy (DMD) is an X-linked monogenic disorder characterised by progressive wasting of skeletal and heart muscles [1]. This condition has an incidence of around 1 in 5,000 newborn boys and is caused by loss-of-function mutations in the X-linked *DMD* gene, resulting in the absence of dystrophin protein expression [2, 3]. *DMD* is the largest human gene, encompassing 79 exons spanning 2.4 Mb. It encodes various short and long dystrophin isoforms that are differentially expressed in muscle and the nervous system [4]. The predominant muscle isoform, Dp427m, plays a crucial role in connecting intracellular actin filaments with extracellular laminin via the dystrophin associated glycoprotein complex, which is essential for maintaining and stabilising the muscle cell membrane (sarcolemma) during contractions. Without dystrophin, the sarcolemma is

69 prone to contraction-induced mechanical stress and damage, leading to the progressive replacement
70 of muscle with fat and fibrous tissue [5].

71 Most DMD-causing mutations are large exonic deletions clustering around deletion hotspots that
72 disrupt the reading frame [6]. Mutations that otherwise retain the *DMD* reading frame are associated
73 with the less common and symptomatically milder Becker Muscular Dystrophy (BMD), where
74 internally truncated but partially functional dystrophin protein is produced.

75 Pre-clinical testing of sequence-specific therapies for DMD, such as antisense oligonucleotide (ASO)
76 and CRISPR gene editing, require humanised animal models due to variations in the *DMD* gene
77 sequence across different species [7]. Strategies designed to target endogenous animal sequences
78 may not necessarily retain their efficiencies or specificities when converted to their corresponding
79 human-targeting sequence. Thus, humanised animal models that incorporate the entire human *DMD*
80 genomic sequence are invaluable for assessing patient-relevant therapeutic candidates that target
81 intronic sequences or modulate splicing.

82 The first humanised DMD mouse model with stable germline transmission, Tg(DMD)72Thoen/J
83 (hDMDTg), was created by integrating a yeast artificial chromosome (YAC) containing the full-length
84 human *DMD* (*hDMD*) transgene into chromosome 5 of mouse embryonic stem (ES) cells [8]. When
85 crossed with the *mdx* strain lacking murine dystrophin expression, the resulting hDMDTg/*mdx* mice
86 exhibited functional compensation for the absence of mouse dystrophin through human dystrophin
87 expression. Because it possesses the complete human *DMD*, including proximal regulatory
88 sequences (promoter), exons and introns, the hDMDTg/*mdx* strain serves as a platform for the
89 further creation of humanised DMD disease models by the introduction of mutations into the *hDMD*
90 transgene that mimic various DMD patient mutations. Examples of these mouse disease models
91 include the exon 52 deletion hDMDTgEx52Δ/*mdx* models [9, 10] and the exon 45 deletion
92 hDMDTgEx45Δ/*mdx* and hDMDTgEx45Δ/*mdxD2* models [11]. A second mouse model with the full
93 length human *DMD* gene carried on a non-integrated human artificial chromosome (DYS-HAC) has

also been generated, but mice with DYS-HACs containing DMD mutations has not yet been developed [12, 13].

During the detailed characterisation of their TALEN-generated hDMDTgEx52Δ/*mdx* mouse model, Yavas *et al.* (2020) identified a tail-to-tail duplication of the *hDMD* transgene present in both the hDMDTgEx52Δ/*mdx* and hDMDTg/*mdx* mice by interphase FISH and digital droplet PCR, indicating that a duplication event likely occurred during the initial generation of the hDMDTg model [14]. The two-copy number of this transgene does not accurately reflect the single-copy nature of the *DMD* gene in males and also makes the generation of mouse models harbouring patient mutations much more complicated.

In this study, we carried out a detailed genetic analysis of hDMDTg mice using short- and long-read whole genome sequencing to resolve the structure of the duplicated hDMD transgenes. We then de-duplicated the *hDMD* transgene in hDMDTg mice through the CRISPR microinjection of zygotes to generate a single-copy hDMD mouse strain, hDMDTgSc. Additionally, we provide characterisation of these hDMDTgSc mice and show that the single remaining transgene copy is functional and rescues the phenotype associated with the endogenous *Dmd* knock-out mouse.

Results

Sequence analysis of the hDMDTg model with duplicated hDMD transgenes

To identify the 5’ and 3’ boundaries of the *DMD* locus of the hDMDTg transgene, short-read genome sequencing was performed on genomic DNA extracted from the liver of homozygous hDMDTg mice. The sequencing data was mapped to the mouse (GRCm38_68) and human (GRCh38) genome reference sequences. Mapping to the mouse genome produced average depth of 27.68-fold (Figure 1). Mapping to the human genome produced an average depth of ~60 fold in the range

118 chrX:31,081,740-33,507,715 strongly suggesting that this region defines the locus included in the
119 original YAC clone used to make the hDMD transgene and confirming that the region was duplicated
120 based on the coverage depth compared to the mouse genome.

121 To identify the transgene integration site on mouse chromosome 5, we generated a custom mouse
122 reference genome incorporating the *DMD* locus identified from mapping to GRCh38 and the
123 sequence of pYAC4 the transgene vector. Next, we mapped short-read Illumina genome sequencing
124 (GS) data and Nanopore long-read sequence data from the hDMDTg mouse to this custom reference.
125 For short-read GS, 36,781 break ends corresponding to sites of structural variation (deletions,
126 duplications, translocations, or inversions) on chr5 were called with GRIDSS2. In the case of the long
127 read GS, 1769 structural variants (SV) were called on chr5. However, without a non-transgenic
128 littermate control genome, the exact insertion site could not be determined from these
129 investigations alone. There were no translocations between mouse chr5, the human *DMD* locus or
130 the pYAC4 identified from any of these analyses.

131 De novo assembly of hDMDTg long-read data revealed a collection of related contigs in the hDMDTg
132 assembly graph that corresponded to the transgene (Supplementary Figure 1A). Approximately 10.8
133 Mb of sequence in the assembly graph which included the *DMD* locus corresponded to yeast
134 chromosomal sequences (Supplementary Figure 1A). Mapping these transgene associated contigs
135 back to the custom GRCm38+*DMD*+pYAC4 reference revealed a breakpoint on chr5: 5:133,626,547
136 that linked to these inserted transgenic sequences. Re-examination of structural variant calls from
137 both long- and short-read mapped data confirmed the transgene insertion site on mouse chr5, which
138 included a 174 kb duplication of mouse chr 5 (5:g.133,626,547_133,800,724dup) at this location
139 (Figure 2).

140 Comparison of the mapped hDMDTg sequences revealed that the *DMD* loci within the hDMDTg
141 transgene were not exact duplications. One copy ranged from chrX:31,081,740-33,507,715, and the
142 other chrX:31,081,740-33,495,727 (Figure 2). The contig containing *DMD* from the de novo assembly

143 also captured the tail-to-tail orientation of the transgenes consisting of inverted repeats of *URA3,*
144 *HygR*, *HIS3*, and *β-lactamase* genes (Figure 2). Based on the assembly the NeoR antibiotic resistance
145 cassette is suggested to only occur once at the 5' end of one of the *hDMD* transgenes (Figure 2).

146

147 **Generation of the single copy hDMDTgSc mouse model**

148 To faithfully recapitulate the single-copy nature of the *DMD* gene, we aimed to generate a single-
149 copy hDMD transgenic mouse model (hDMDTgSc) by microinjecting hDMDTg$^{Tg/0}$ embryos with
150 SpCas9 mRNA and gRNAs targeting the NeoR and HygR cassettes on the YAC sequences. This was
151 designed to induce the megabase-scale intervening deletion of only a single transgene. Potential
152 founders were screened by PCR and by identified by amplification of the deletion interval between
153 the NeoR and HygR gRNA target sites, as well as the absence of NeoR amplification. One male
154 founder was identified to carry the desired deletion (Figure 3A).

155 Following transmission of the de-duplicated transgene to G1 mice, we quantified the hDMD
156 transgene copy number in hDMDTgSc$^{Tg/0}$ mice by qPCR using three sets of primers targeting the 5'
157 untranslated region (UTR), Exon 50 and 3' UTR of the *hDMD* transgene. Both G1 male and female
158 hDMDTgSc$^{Tg/0}$ mice were confirmed to carry a single copy of the autosomal *hDMD* transgene,
159 compared to two copies in the hDMDTg$^{Tg/0}$ control (Figure 3B).

160

161 **Sequence analysis of hDMDTgSc mice**

162 Mapping Nanopore long-read sequence data from hDMDTgSc mice to our custom reference mouse
163 genome with the human *DMD* locus and pYAC4 sequence, we identified 1462 SVs on chr5. Of all SV
164 called between the hDMDTg and hDMDTgSc data, 196 were concordant, one of which was the 174kb
165 duplication on chr5 containing the transgene insertions site. De novo assembly of hDMDTgSc long-
166 read data also captured the transgene insertion site on mouse chr5; however, a contig corresponding

167 to *DMD* was not assembled from these data by Shasta, and thus, this graph contained only mouse
168 and yeast chromosomal sequences (Supplementary Figure 1B). From comparison to mapped
169 hDMDTg sequences, only the shorter copy of the *DMD* locus (chrX:31,081,740-33,495,727) was
170 retained in the single-copy hDMDTgSc transgene (Figure 1).

171

172 **Characterisation of the hDMDTgSc mDmdKO mouse model**

173 To assess if the single copy hDMD transgene in hDMDTgSc mice retains functionality, we
174 characterised male hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ and hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice, both strains lacking
175 endogenous mouse dystrophin expression.

176 Skeletal and heart muscle tissues were collected from p60 mice for mRNA, protein, and histology
177 analysis. We performed quantitative RT-PCR analysis of the dystrophin transcript in quadriceps and
178 heart muscles using a set of primers that detects both the human and mouse transcript by targeting
179 shared sequences. In quadriceps muscle, the expression level of *hDMD* in hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice
180 is equivalent to that of the *mDmd* in WT C57BL/6J mice, whereas *hDMD* expression in hDMDTgSc$^{Tg/0}$
181 mDmd$^{KO/Y}$ quadriceps is approximately halved (Figure 4A). In heart muscles, *hDMD* transcript in
182 hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice is slightly higher than *mDmd* in WT C57BL/6J mice, while *hDMD*
183 expression in hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ heart is, although somewhat variable, more similar to that of
184 *mDmd* (Figure 4B).

185 Western blots were performed with the human dystrophin-specific MANDYS106 antibody. Full-length
186 dystrophin protein was detected in tibialis anterior (TA), quadriceps and heart muscles of
187 hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice at a lower level compared to hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice (Figure 5). The
188 quadriceps western blot was re-probed with MANDYS8, which is cross-reactive with both human and
189 mouse dystrophin, confirming that MANDYS106 only detects human dystrophin.

190 Immunofluorescence staining of dystrophin with MANDYS106 in TA, quadriceps and heart muscles
191 showed the expected sarcolemmal localisation of the human dystrophin protein in hDMDTgSc$^{Tg/0}$
192 mDmd$^{KO/Y}$ muscle, as for hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ muscle tissue (Figure 6, Supplementary Figure 2).
193 Histological analysis of hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ muscle sections stained with hematoxylin and eosin
194 (H&E) revealed normal muscle histology with nuclei localised at the sarcolemmal region, consistent
195 muscle fibre size, and no evidence of damage or infiltrating cells (Figure 7). Conversely, mDmd$^{KO/Y}$
196 muscle was dystrophic, with many centralised nuclei indicating muscle regeneration as well as the
197 presence of cellular infiltrates.

198 We performed forelimb grip strength testing followed by serum collection for creatine kinase analysis
199 on mice at postnatal day 32 (p32) and day 60 (p60). Forelimb grip strength testing showed that
200 hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice have similar maximum and mean grip strength to hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$
201 mice at p32 and that both these strains have no significant difference in strength compared to
202 wildtype C57BL/6J control mice, demonstrating functional rescue of the reduced grip strength
203 caused by the endogenous *mDmd* knock-out (Figure 8A,B). hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ and hDMDTg$^{Tg/0}$
204 mDmd$^{KO/Y}$ also do not experience the same onset of fatigue caused by repeated measurements as
205 observed in mDmd$^{KO/Y}$ mice (Figure 8C,D).

206 The serum samples collected after the forelimb grip strength testing were used to assess serum
207 creatine kinase (CK) levels, an indirect measure of muscle damage. Both hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ and
208 hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice have a similar reduction of serum CK levels to WT C57BL/6J controls at
209 p32 and p60 (Figure 9).

210

211

212 **Discussion**

213

Validated humanised animal models are important tools for pre-clinical evaluation of genetic therapies as they permit in vivo testing of sequence-specific reagents that can be directly translated into patients. Generation of a hDMDTg mouse model containing the entire *hDMD* locus randomly integrated into mouse chromosome 5 was a significant advance, particularly given the size of the transgene (2.4 Mb). However, it was subsequently shown that two copies of the transgene had inserted, which does not accurately reflect the copy number of the *DMD* gene [14]. Furthermore, derivatives of the hDMDTg line such as the del52 exon deletion disease model are more difficult to engineer and validate. While these double-copy humanised models can be used for the testing of ASOs [15], they are not ideal for evaluating strategies that are reliant on DNA double-stranded breaks such as CRISPR-Cas9 gene editing [16]. This is because large intervening deletions between the target sites of the two transgene copies can complicate the analysis of DNA repair outcomes.

Here, we describe the generation and characterisation of the hDMDTgSc mouse model carrying a single functional copy of the 2.4 Mb *hDMD* transgene. Large megabase-scale DNA modifications have been previously demonstrated to be achievable through CRISPR microinjection[17-20]. We performed a detailed genetic characterisation of the original duplicated hDMDTg transgene structure by analysis of Illumina short-read and Nanopore long-read sequencing data to determine unique targetable regions. Using gRNAs that flank one copy of the transgene, we readily achieved transgene de-duplication of the hDMDTg mouse model by CRISPR zygote microinjection and demonstrated stable germline transmission.

We also successfully resolved the exact YAC integration site in chromosome 5 and can confirm that no genes were disrupted at the integration site. Long-read sequencing helped us confirm that a single copy of the hDMD transgene remained intact in hDMDTgSc mice following the de-duplication process, and its reduced copy number was also reflected in the reduced mapping depths from sequencing.

238 There are some limitations to the genetic analyses we performed via next-generation sequencing.
239 Using only nanopore long-read sequences, we were not able to unambiguously assemble the full-
240 length hDMDTgSc and hDMDTg transgenes despite achieving a total assembly length similar to the
241 length of the mm10 genome reference and a relatively high contig N50 (Supp Table 1). The exact
242 assembly and ordering of the yeast chromosomal sequences that flank the DMD loci in the
243 duplicated and single-copy transgenes remains uncertain. Additional long-read sequencing with high-
244 quality PacBio circular consensus sequences would likely be sufficient to create a near-continuous
245 assembly of the transgenic mouse genome as has been done for the human T2T reference sequence
246 [21].

247 From phenotypic, biochemical, molecular, and histological characterisation of hemizygous hDMDTgSc
248 mice lacking endogenous mouse dystrophin, we found that the single hDMD transgene is sufficient
249 to rescue the dystrophic phenotype observed in mDmd knock-out mice at a similar level to the
250 hDMDTg mDmdKO strain. This is despite the reduced transgene transcription in skeletal muscle
251 driven by the human promoter sequence, which is likely not optimally regulated by the mouse
252 transcription machinery [22]. Additionally, there appears to be a slight reduction of hDMDTgSc
253 mDmdKO mouse strength compared to hDMDTg mDmdKO mice at p60 that might reflect a high but
254 incomplete level of functional rescue- more studies are needed to determine if it is due to transgene
255 expression not being as high as in hDMDTg mDmdKO mice or if it could be attributed to experimental
256 variation.

257 Our hDMDTgSc mouse model has considerable potential for the pre-clinical testing of sequence-
258 specific ASOs or CRISPR therapeutics as, in principle, any causative DMD mutation can be engineered
259 into the single-copy 2.4 Mb *hDMD* transgene, such as models lacking *DMD* exons 45 or 52 like those
260 previously generated from the double transgenic mice. This enables ready generation of humanised
261 DMD disease mouse models, which has an advantage over the more labour-intensive and inefficient
262 partial sequence humanisation methods where exon knock-ins must be performed for each separate

263 exonic target [23]. Crossing these human transgene knock-out mice with mice lacking endogenous
264 dystrophin expression would result in a disease model suitable for the testing of CRISPR therapy
265 strategies, and could also be inter-crossed with a mouse strain lacking both endogenous dystrophin
266 and utrophin expression to simulate an even more severe dystrophic phenotype [24].

267

268

269 **Materials and Methods**

270

271 **Animal studies**

272 All procedures and experiments using mice were approved by the South Australian Health and
273 Medical Research Institute (SAHMRI) Animal Ethics Committee and conducted per the Australian
274 code for the care and use of animals for scientific purposes.

275 Mice were housed in 13.5-hour light/10.5-hour dark cycles and were fed the Teklad Global 2918
276 rodent diet *Ad Libitum*. Except where indicated, male mice were used in characterisation
277 experiments. Tg(DMD)72Thoen/J (hDMDTg) were obtained from JAX (stock #018900) and maintained
278 in-house [8]. mDmdΔEx51 ($mDmd^{KO}$) and hDMDTgSc were generated as described in the sections
279 below and maintained in-house. We maintain hDMDTgSc mice as a tick-over colony of homozygotes.
280 C57Bl/6J mice are maintained in house (JAX stock #000664).

281

282 **Mouse generation**

283 To create the intervening deletion of a single copy of the *hDMD* transgene, we designed SpCas9
284 gRNAs targeting the Neomycin and Hygromycin antibiotic resistance cassettes in the integrated YAC
285 sequence. Complementary gRNA oligos with overhangs facilitating golden gate assembly were

286 phospho-annealed and inserted into pSpCas9(BB)-2A-Puro (PX459) V2.0 [25]. pSpCas9(BB)-2A-Puro
287 (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988 ; http://n2t.net/addgene:62988 ;
288 RRID:Addgene_62988). Following plasmid validation by *BbsI* diagnostic digest and Sanger sequencing
289 (Australian Genome Research Facility), gRNAs were amplified using NEB Phusion in HF buffer with
290 primers containing a 5' T7 promoter sequence. PCR amplicons purified using the QIAquick PCR
291 Purification Kit (Qiagen) were used to generate gRNA using the HiScribe T7 Quick High Yield RNA
292 Synthesis Kit (NEB) and purified using the RNeasy Mini Kit (Qiagen).

293 hDMDTg$^{Tg/Tg}$ mouse sperm was used for *in-vitro* fertilisation of C57BL/6J oocytes, and 2PN embryos
294 were injected with a microinjection mix containing a final concentration of 100 ng/µL SpCas9 mRNA
295 and 50 ng/µL each of NeoR and HygR gRNA. This was followed by oviduct transfer into pseudo-
296 pregnant females. Intervening transgene deletion was detected in founders by PCR amplification
297 using the forward NeoR and reverse HygR primers.

298 mDmd^{KO} mice were generated by deletion of *Dmd* Exon 51 via CRISPR microinjection of C57BL/6J
299 embryos (manuscript submitted).

300

301 **DNA PCR and Sanger sequencing**

302 Genomic DNA (gDNA) was isolated from mouse ear punch biopsies using the PCR Template
303 Preparation kit (Roche). Targets was PCR amplified using Taq Polymerase (Roche) and FailSafe™
304 Buffer J (Epicentre) and resolved on a 1% agarose TBE gel.

305 To prepare samples for sequencing, PCR reactions were cleaned up using the QIAquick PCR
306 Purification kit (Qiagen). Samples with sequencing primers added were sent to the Australian
307 Genome Research Facility (AGRF) for the purified DNA Sanger sequencing service.

308

309 **Short-read genome sequencing**

310 Library prep and sequencing was performed as a service by the Australian Genome Research Facility
311 (AGRF). Genomic DNA extracted from hDMDTg mice liver were sheared and PCR-free sequencing
312 libraries prepared. From these libraries, 150 base pair, paired-end reads were generated on a
313 NovaSeq 6000. Image analysis was performed in real time by the NovaSeq Control Software (NCS)
314 v1.7.5 and Real Time Analysis (RTA) v3.4.4. Then the Illumina bcl2fastq v2.20.0.422 pipeline was used
315 to generate the final fastq files.

316 *Mapping*

317 All mapping of short read data was carried out using BWA-MEM v 0.7.17 [26] setting option -K, the
318 number of input bases, to 100,000,000 per batch and all other options as default.

319 A custom reference genome was prepared which combined the GRCm38.p3 mouse genome
320 (GenBank: GCA_000001635.5), a segment of human chromosome X (NCBI RefSeq: NC000023.11;
321 chrX:31,081,740-33,507,715) and pYAC4 (GenBank: U01086.1). The chrX segment was defined by
322 identifying the extents of reads accurately mapped to the region flanking *DMD* on chrX in the
323 hs38DH human genome (created using the run-gen-ref script from bwakit) [27]. The pYAC4 vector
324 sequence was used as a proxy for the pRANT-11 and pLGTel-1 vectors that were originally used for
325 cloning of the hDMD YAC transgene and for which no sequence data are available [28].

326 Alignments were sorted by genome coordinate positions using samtools sort v 1.17 and duplicate
327 reads were marked using sambamba markdup v 0.8.2 [29].

328 *Variant calling*

329 Structural variants were called using GRIDSS v 2.12.0 with default settings [30].

330

331 **Oxford nanopore genome sequencing**

High molecular weight genomic DNA was extracted from homozygous hDMDTg and hDMDTgSc mice liver using the Monarch HMW DNA Extraction Kit for Tissue (NEB). Oxford Nanopore Technologies Promethion genome sequencing was performed as a service by Novogene, (Novogene HK). Sequence libraries were prepared using the DNA ligation sequencing protocol with the SQK-LSK-110 kit. Sequence data were generated using a Promethion 9.4.1 flow cell. Bases were called with bonito v0.6.2 (Oxford Nanopore Technologies) with the dna_r9.4.1_e8_sup@v3.3 super accurate base calling model.

Mapping

Oxford Nanopore long-read data were mapped to the custom GRCm38+*DMD*+pYAC4 genome described above and also the human genome reference GRCh38 using minimap2 v2.17-r941 with default parameters for nanopore sequencing.

Variant calling

Structural variants were called from mapped long read genome sequencing with Sniffles2 using default settings [31].

Assembly

Nanopore sequence reads were assembled de novo using Shasta v 0.11.1 [32] using the Nanopore-May2022 model adjusting the minimum length of input reads to 5 kb. To identify contigs associated with the hDMD transgene sequence, the Assembly.fasta outputs were mapped to the custom GRCm38+*DMD*+pYAC4 genome described above with minimap2 and assembled into super contigs using RagTag v 2.1.0 [33]. The contigs from the Assembly.fasta files were also converted into a BLAST v 2.14.1+ library that was queried with 500 bp segments from the 5’ and 3’limits of the *DMD* locus in the hDMDTgSc transgene. Assembly graphs were viewed using Bandage v 0.8.1.

Transgene copy number determination

356 qPCRs were performed to quantify the *hDMD* transgene copy number and were done in technical
357 triplicate. Each reaction well was set up using SYBR™ Green PCR Master Mix (Applied Biosystems), 50
358 nM of each forward and reverse primer and 10 ng gDNA in a 15 µL reaction volume. qPCR data was
359 normalised to endogenous *Sox1.* Reactions were run on the QuantStudio™ 3 Real-Time PCR System
360 (Applied Biosystems) using the Fast Run Mode with the following reaction conditions: 95°C for 20 s,
361 followed by 40 cycles of 95°C for 1 s and 60°C for 20 s, and a dissociation protocol to obtain a melt
362 curve for all samples running from 60°C to 95°C.

363

364 **RNA extraction and qRT-PCR**

365 Muscle tissues were bead homogenised in 500 µL TRIzol reagent (Invitrogen) using Lysing Matrix D
366 (MP Biomedicals). 100 µL of chloroform was added and samples were shaken vigorously, incubated
367 at RT for 3 mins then centrifuged at 10 000 g for 18 min at 4°C to separate phases. The aqueous
368 phase was isolated, mixed with an equal volume of ethanol, and spun through a RNeasy column from
369 the RNeasy Mini kit (Qiagen) at 8 000 g for 30 s. The column was washed with 700 µL buffer RW1 and
370 twice with 500 µL buffer RPE. RNA was eluted in 30 µL of nuclease-free water, quantified by
371 nanodrop and stored at -80°C. cDNA was generated from 1 µg RNA using the High-Capacity RNA-to-
372 cDNA kit (Applied Biosystems) as per manufacturer's protocol.

373 qRT-PCRs were performed to quantify dystrophin transcript expression in technical and biological
374 triplicate. The primers used were designed to target shared sequences in both *hDMD* and *mDmd*
375 transcript. Each reaction well was set up using SYBR™ Green PCR Master Mix (Applied Biosystems),
376 50 nM of each forward and reverse primer and 1 µL of cDNA in a 15 µL reaction volume. qPCR data
377 was normalised to beta-actin (*Actb*) expression. Reactions were run on the QuantStudio™ 3 Real-
378 Time PCR System (Applied Biosystems) with the same parameters listed in the transgene copy-
379 number determination section above. Analysis was performed using the Design and Analysis
380 Software 2.7.0. Transcript expression data presented as $\log_2$(RQ) to show proportional changes.

381

382 **Western blot**

383 Tissues were bead homogenised in disruption buffer (75 mM Tris HCl pH 6.8, 15% SDS, 1x cOmplete
384 Protease Inhibitor Cocktail EDTA (Roche)), using Lysing Matrix D (MP Biomedicals). 1/10 diluted
385 lysates were quantified using the Pierce BCA protein assay (Thermo Scientific). 30 µg of protein
386 lysates were mixed with equal parts 2x Loading buffer (21% w/v glycerol, 0.001% w/v bromophenol
387 blue, 5% β-Mercaptoethanol), boiled at 95°C for 5 mins and spun at 21 300 rcf for 5 mins prior to
388 loading in NuPAGE Tris-Acetate gels (Invitrogen). Gels were run at 150 V for 1 h and transferred onto
389 0.2 µm PVDF membranes using the High MW setting on the Trans-Blot Turbo (Bio-Rad). Membranes
390 were blocked with 10% skim milk in TBS-T for 2h with gentle agitation at RT. Dystrophin was probed
391 with 1:500 Anti-Dystrophin antibody, clone 2C6 MANDYS106 (Sigma-Aldrich) in 2% skim milk in TBS-T
392 overnight at 4°C. The following day, membranes were washed three times for 5 mins with TBS-T and
393 incubated with 1:5 000 secondary anti-mouse antibody in 2% skim milk in TBS-T for 1h at RT. After
394 another set of TBS-T washes, membranes were developed with SuperSignal West Pico Plus
395 Chemiluminescent Substrate (Thermo Scientific), and chemiluminescent detection was performed on
396 the ChemiDoc MP system (Bio-Rad) using automated exposure settings for intense bands.
397 Membranes were stripped with Restore PLUS Western Blot Stripping Buffer (Thermo Scientific) prior
398 to re-probing. All dystrophin was probed with 1:500 anti-dystrophin antibody MANDYS8. Vinculin
399 was probed with 1:1 000 Anti-vinculin antibody, V9131 (Sigma-Aldrich).

400

401 **Immunofluorescence**

402 Fresh muscle was snap-frozen in dry-ice-cooled isopentane and immediately embedded in Tissue-Tek
403 O.C.T. 10 µm thick cryostat-sliced transverse sections were permeabilised with PBS + 0.3% Triton X-
404 100 for 10 minutes, blocked with 10% FCS in PBS + 0.3% Triton X-100 for 1 h at RT, and treated with

405 1x ReadyProbes™ Mouse-on-Mouse IgG Blocking Solution (Invitrogen) as per manufacturer
406 instructions. Slides were incubated with 1:100 Anti-Dystrophin antibody, clone 2C6 MANDYS106
407 (Sigma-Aldrich) diluted in blocking solution overnight at 4°C in a humidified chamber. The next day
408 slides were washed with PBS for 3x 5mins and incubated with secondary anti-mouse antibody
409 conjugated with Alexa Fluor® 594 for 1h at RT. Slides were washed with PBS for 3x 5mins and
410 mounted with ProLong™ Gold Antifade Mountant with DAPI (Invitrogen) and coverslip. Images were
411 taken on a Nikon Eclipse Ti2 microscope with a DS-Qi2 camera. The same image processing was
412 applied to images from the same muscle group.

413

414 **Histological analysis of muscles**

415 Fresh muscle was mounted on 10% gum tragacanth on a piece of cork, snap frozen in liquid-nitrogen-
416 cooled isopentane and immediately stored at -80°C. For skeletal muscles, 10 µm thick cryostat-sliced
417 transverse sections were stained with Lillie Mayer's Hematoxylin and Eosin, as described by Wang et
418 al. (2017) with the following modifications to the dehydration steps: 2 dips of 70% ethanol, 4 dips of
419 95% ethanol, 2x 1 min of 100 ethanol and 2x 1 min of xylene. For heart muscles, 10 µm thick
420 cryostat-sliced transverse sections were stained with Lillie Mayer's Hematoxylin for 6 mins, rinsed
421 with running tap water for 1min, destained with 0.5% acid alcohol and developed in Scott's tap water
422 for 1 minute [34]. After rinsing with running water for 2 minutes, slides were stained with eosin for 3
423 mins and quickly rinsed with running tap water before dehydrating through an increasing
424 concentration of ethanol and xylene as described in the skeletal muscle protocol. Slides were
425 mounted with DPX mountant (Sigma-Aldrich) and coverslip. Images were taken using the 3DHISTECH
426 Pannoramic 250 Flash Slide Scanner at 40x magnification. The same image processing was applied to
427 all images.

428

429 **Forelimb grip strength test**

430 Muscle strength measurement was performed by a forelimb grip strength behaviour task [35]. Mice
431 were tested at p32 and p60 using a grid attached to a Centor Easy 10N (Andilog) force gauge. Mice
432 were suspended by the tail and allowed to grasp the grid tightly with both front paws, then pulled
433 along a straight horizontal line away from the force gauge until their grip on the grid was broken.
434 Measurements were repeated 4 times per set, and 3 sets were performed with a 1-minute rest
435 interval between sets. The highest measured rep was recorded for each set and normalised against
436 body weight. Max force reported is the highest measured rep among all sets, and mean force
437 reported is the average of the highest measured reps for each of all 3 sets.

438

439 **Serum creatine kinase measurement**

440 30 mins after grip strength testing, Emla 5% local anaesthetic cream was applied to the mice's tails,
441 and they were warmed in a Mini Thermacage (datesand) for 10 mins at 36°C to dilate the tail vein.
442 Blood samples were collected in microvettes from a nick to the lateral tail vein. The blood was
443 allowed to clot at room temperature for 30 mins, and the clot was separated from the serum by
444 centrifuging at 2 000 g for 10 mins at 4°C and immediately stored at -80°C. Mouse serum creatine
445 kinase (CK) levels were measured by Gribbles Veterinary Pathology using the ADVIA Creatine Kinase
446 CK_L assay on the ADVIA 1800 Chemistry System.

447

448 **Availability of data and materials**

449 The sequencing dataset supporting the conclusions of this article is available in the NCBI Sequence
450 Read Archive, [unique persistent identifier and hyperlink to dataset(s) in http:// format].

451

452 **List of figure legends**

453

454 **Figure 1: Coverage of the human *DMD* locus in the duplicated (hDMDTg) and single copy**
455 **(hDMDTgSc) transgenes.** Mapping depths at the *DMD* locus for short-read and long-read genomes
456 mapped to the human genome. The red box on the chromosome X ideogram at the top shows the
457 approximate location of the coverage shown. Graphs show the coverage of mapped reads for the
458 short-read (hDMDTg-SR) and long-read (hDMDTg-LR, hDMDTgSc-LR) genomes at this location. The
459 *DMD* exon model is shown below the coverage graphs. At the base of the image a zoomed in section
460 chrX: 33490727-33509715, shows the drop in coverage from position chrX:33,495,727 onwards in
461 the hDMDTg data and no coverage from this position in the hDMDTgSc mice.

462

463 **Figure 2: The structure of the duplicated hDMD transgene.** Simplified representation of the hDMD
464 duplicated transgene insertion site on mouse chromosome 5 showing the two copies of *DMD*
465 interspersed with YAC vector sequences and the mouse chr 5 duplication.

466

467 **Figure 3 : Genotyping hDMDTgSc mice. (A)** Detection of the single transgene deletion in G0 founder
468 mice by PCR. The intervening deletion between the Neomycin resistance (NeoR) and Hygromycin
469 resistance (HygR) cassettes was identified in Founder number 5 from the absence of PCR
470 amplification using primers flanking the NeoR cut site and the presence of amplicon using a forward
471 primer in NeoR and a reverse primer in HygR. The green text indicates the founder with the banding
472 expected from successful transgene deletion. **(B)** Real-time qPCR copy-number determination of the
473 human *DMD* transgene. The assay was performed on genomic DNA extracted from ear notch
474 biopsies of hDMDTg$^{Tg/0}$ mDmd$^{+/Y}$, hDMDTgSc$^{Tg/0}$ mDmd$^{+/Y}$ (male), hDMDTgSc$^{Tg/0}$ mDmd$^{+/+}$ (female) and
475 wildtype C57BL/6J (mDmd$^{+/Y}$) mice. Measurement of three targets, 5’ untranslated region (UTR),

476 exon 50 and 3' UTR, were normalised against the endogenous mouse *Sox1* gene. Transgene copy
477 number was successfully reduced from two to one in heterozygous hDMDTgSc mice. Error bars
478 indicate the 95% confidence interval of technical replicates.

479

480 **Figure 4: Dystrophin transcript measurement by qRT-PCR.** Quantification of (A) skeletal and (B)
481 heart tissues from p60 mice was performed using a set of primers that target both human *DMD* and
482 mouse *Dmd* transcript sequences, normalised against beta-actin (*Actb*). Data presented as $\log_2$
483 relative quantity (RQ) ± 95%CI, one-way ANOVA with Tukey's multiple comparisons test performed
484 on ddCt values, *** $P < 0.001$. n=3.

485

486 **Figure 5: Detection of human dystrophin protein in muscle tissue by western blot.** Analysis was
487 performed on skeletal hindlimb muscles, tibialis anterior and quadriceps, as well as heart muscle
488 protein lysates from p60 mice using the human-specific antibody against dystrophin MANDYS106.
489 The tibialis anterior blot was also re-probed with the cross-reactive antibody MANDYS8 to
490 demonstrate the species-specificity of MANDYS106. Vinculin was probed as the loading control.
491 Human dystrophin protein expression in hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice is reduced in all three tissues
492 compared to hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice.

493

494 **Figure 6: Immunofluorescence localisation of the human dystrophin protein.** Immunostaining was
495 performed on tibialis anterior, quadriceps and heart muscle tissue sections from p60 mice using the
496 MANDYS106 primary antibody. Scale bar 100 µm.

497

498 **Figure 7: Histological analysis of hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ and mDmd$^{KO/Y}$ mice.** Hematoxylin and eosin
499 staining was performed on quadriceps, tibialis anterior and triceps muscle tissue sections from p60
500 mice. Scale bar 50 µm.

501

502 **Figure 8: Forelimb grip strength testing to measure muscle function and performance of mice.**
503 Strength (grams of force) of hDMDTg$^{Tg/0}$ mDmd^{KO} and hDMDTgSc$^{Tg/0}$ mDmd^{KO} mice were compared
504 against WT C57BL/6J and mDmdΔEx51 (mDmd$^{KO/Y}$) controls, normalised to body weight at p32 and
505 p60. (A) Maximum force is the highest recorded measure among three sets of five pulls, and (B)
506 mean force is the average of the highest recorded measures of each of the three sets. Data
507 presented as mean ± SEM, two-way repeat measures ANOVA with Tukey's multiple comparisons test
508 ** $P < 0.01$, *** $P < 0.001$. Samples sizes (n) indicated in within bars in graph. (C,D) Averages of the
509 highest recorded measures for each set of pulls, presented as mean ± SEM, two-way repeat
510 measures ANOVA, Dunnett's multiple comparisons test with C57BL/6J as control group, ** $P < 0.01$,
511 *** $P < 0.001$.

512

513 **Figure 9: Serum creatine kinase (CK) biomarker level measurement of mice.** Serum CK was assayed
514 as an indirect measure of muscle damage and sarcolemma leakage at p32 and p60. The single-copy
515 transgene in hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice rescues the elevated serum CK caused by endogenous
516 *mDmd* knock-out to levels that are not significantly different than either WT C57BL/6J or hDMDTg$^{Tg/0}$
517 mDmd$^{KO/Y}$ mice. Data presented as mean ± SEM, two-way repeated measures ANOVA with Tukey's
518 multiple comparisons test *** $P < 0.001$. Samples sizes (n) indicated in first graph.

519

520

521 **Supplementary materials**

522

523 **Supplemental Table 1: ONT output and assembly statistics**

Sample	Sample Bases (bp)	Number of reads	Median read length	Read length N50	Median read quality	Number of reads >Q7 (%)	Number of reads > 5 kb	Number of contigs	Total assembly length	Longest contig	contig N50
hDMDTgSc (Single Tg)	77,630,478,340	4,776,412	11,923	27,794	14.2	4,327,887 (90.6%)	3,096,917	3,303	2,562,369,505	40,468,389	11,625,580
hDMDTg (Duplicated Tg)	76,267,366,191	5,071,821	11,392	24,219	14.1	4,531,600 (89.3%)	3,296,519	3,843	2,535,656,640	30,111,232	8,109,054

524

525

526

527

528

529 **Supplemental Table 2: DNA primers and oligos**

Name	Sequence 5'-3'
PCR NeoR F	tggattgcacgcaggttctccg
PCR NeoR R	aaggcgatagaaggcgatgcgc
PCR HygR F	gcgggtaaatagctgcgccgat
PCR HygR R	ctttgccctcggacgagtgctg
qPCR hDMD5p F	tagctggggtttgttgccgagc
qPCR hDMD5p R	gaggttggaggcaagcacaccc
qPCR hDMD3p F	cgcagcttgactggagagaggc
qPCR hDMD3p R	agctgaagaggaggcaggacgg
qPCR hDMD 50b F	ctgagctctgagtggaaggcgg
qPCR hDMD 50b R	tccaggagctaggtcaggctgc
qPCR Sox1 F	gacttgcaggctatgtacaacatc
qPCR Sox1 R	cctctcagacggtggagttatatt
qPCR cDNA DMD F	cagccagtgaagaggaagttag
qPCR cDNA DMD R	gaagatggcatttctagtttggaga
qPCR cDNA bActn F	ttgctgacaggatgcagaag
qPCR cDNA bActn R	gtacttgcgctcaggaggag
gRNA cacc Neo	caccgtgcgcaaggaacgcccgtcg
gRNA aaac Neo	aaaccgacgggcgttccttgcgcac
gRNA cacc Hyg	caccgtgcgacgcaatcgtccgatc
gRNA aaac Hyg	aaacgatcggacgattgcgtcgcac
PCR IVT Neo gRNA	ttaatacgactcactatagtgcgcaaggaacgcccgtcg
PCR IVT Hyg gRNA	ttaatacgactcactatagtgcgacgcaatcgtccgatc
PCR IVT TracR R	Aaaagcaccgactcggtgcc

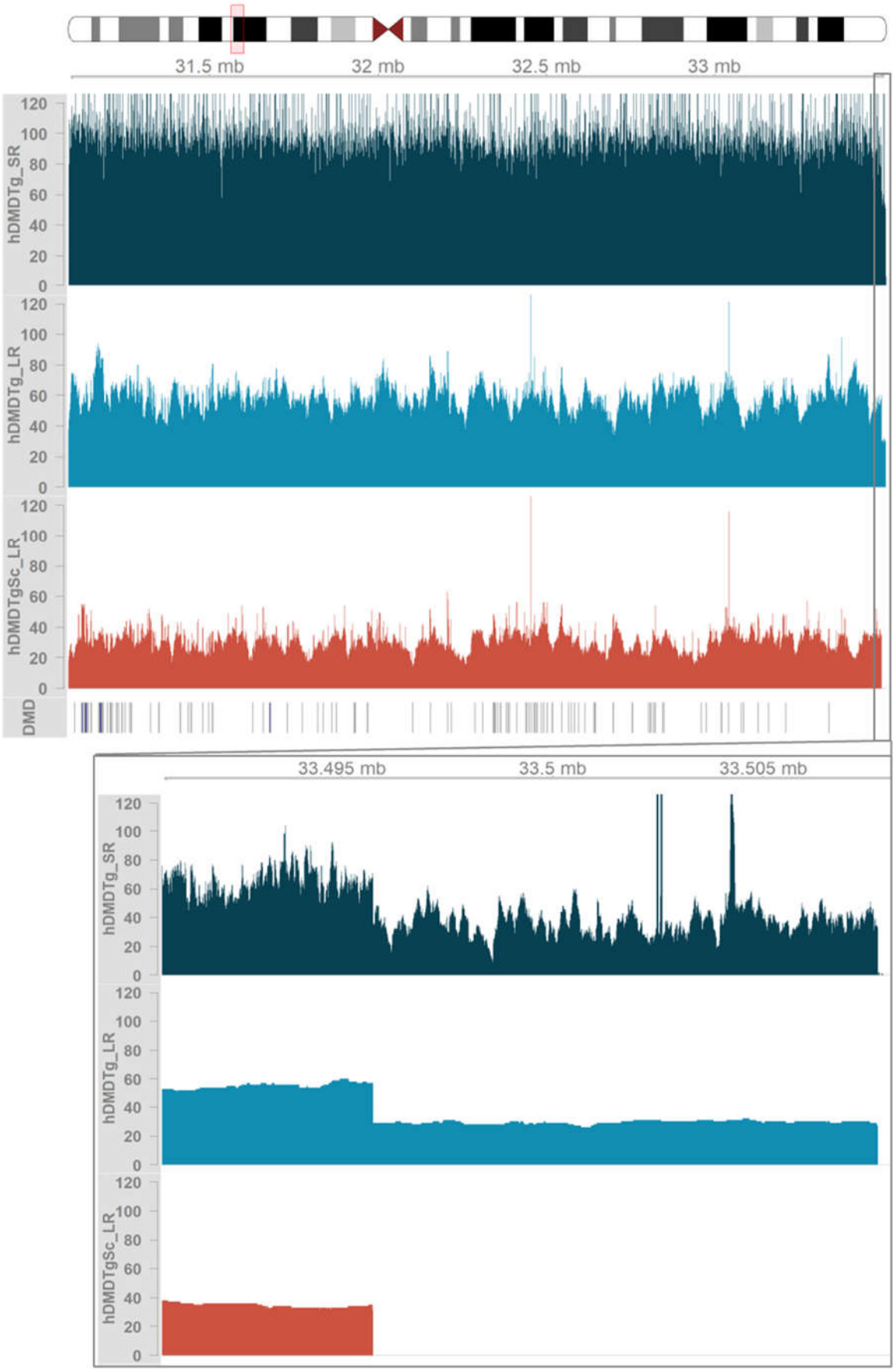
530 **Supplementary Figure legends**

531 **Supplementary Figure 1:** (A) Assembly sequence graph generated by Shasta of the sequences linked
532 to the human *DMD* locus. Each of the thicker coloured nodes represents a different contig from the
533 assembly. Labels indicate the best-matched identities of four sampled sequences of approximately
534 1.5 kb from the beginning, middle and end of each contig to BLAST searches of the non-redundant
535 nucleotide database from NCBI. The branched paths in the graph indicate the different sequences at
536 the beginning and end each copy of the DMD transgene. (B) The structure of the single copy hDMD
537 transgene integration site. Assembly sequence graph generated by Shasta of the sequences linked to
538 the mouse chromosome 5 insertion site (5: 133,626,547). Each of the thicker coloured nodes
539 represents a different contig from the assembly. Labels indicate the best-matched identities of four
540 sampled sequences of approximately 1.5 kb from the beginning, middle and end of each contig to
541 BLAST searches of the non-redundant nucleotide database from NCBI. The presence of mouse chr4
542 and chr7 in this graph are likely misassembled.

543

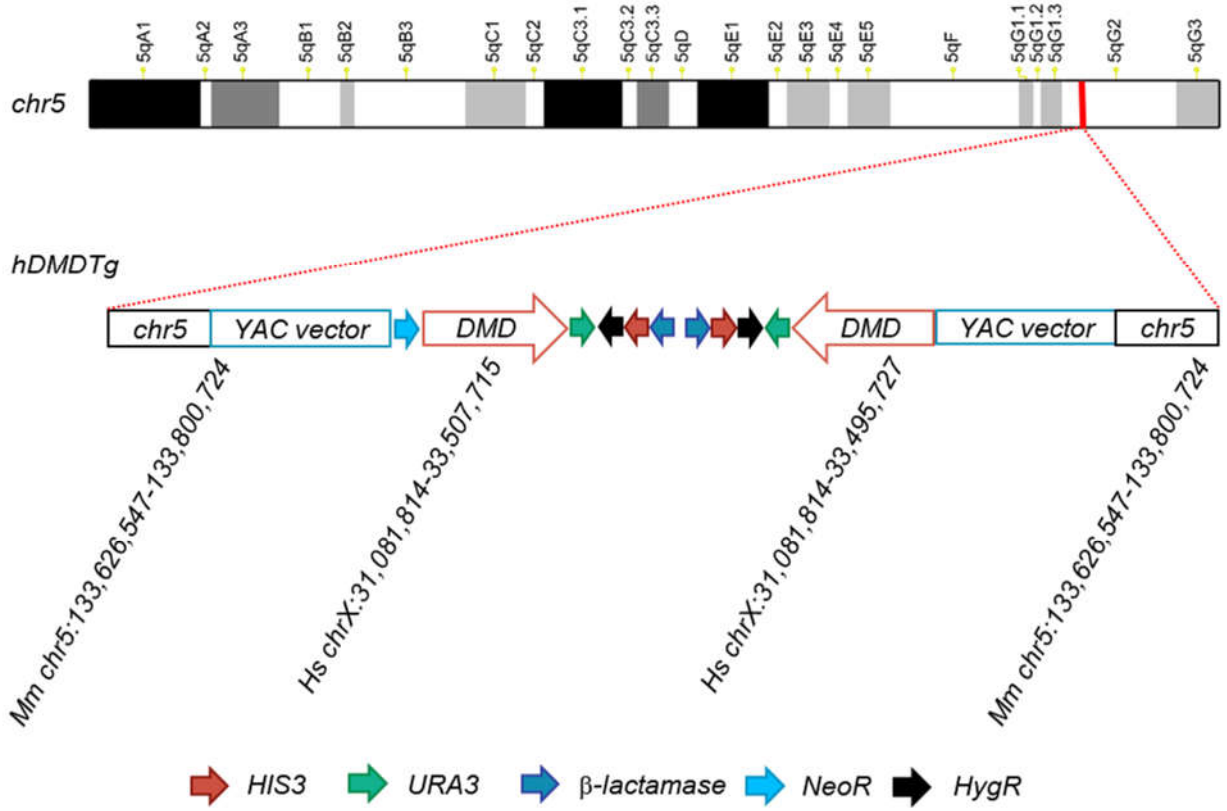
544 **Supplementary Figure 2:** Dapi and merged immunofluoresence images. Scale bar 100 µm.

545 Figure 1



546

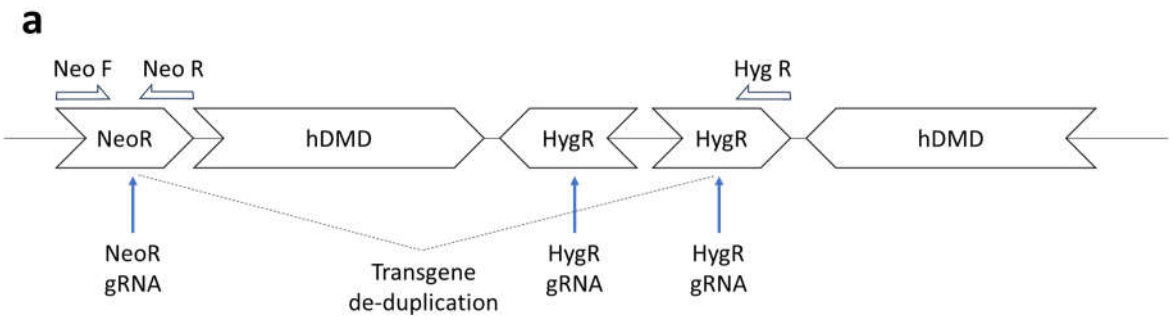
547 Figure 2



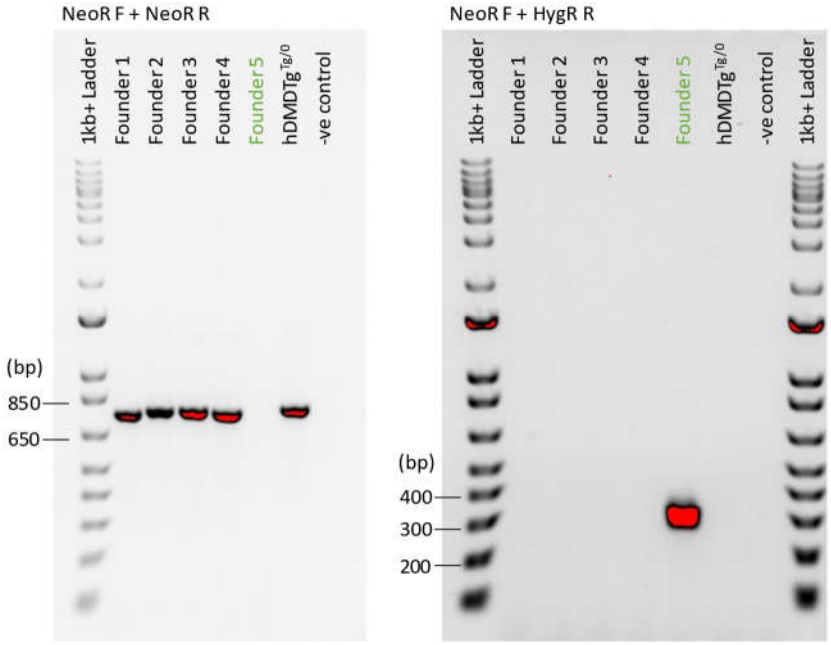
548

549

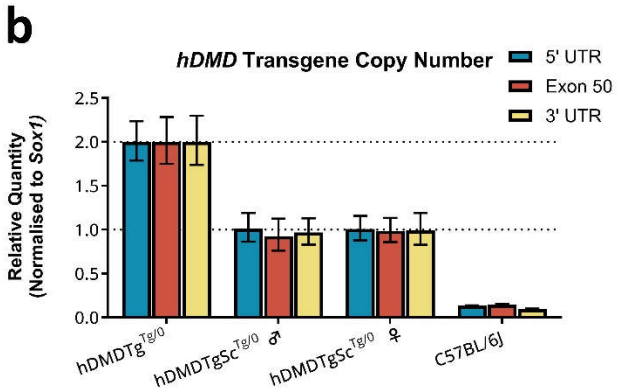
550 Figure 3



551



552

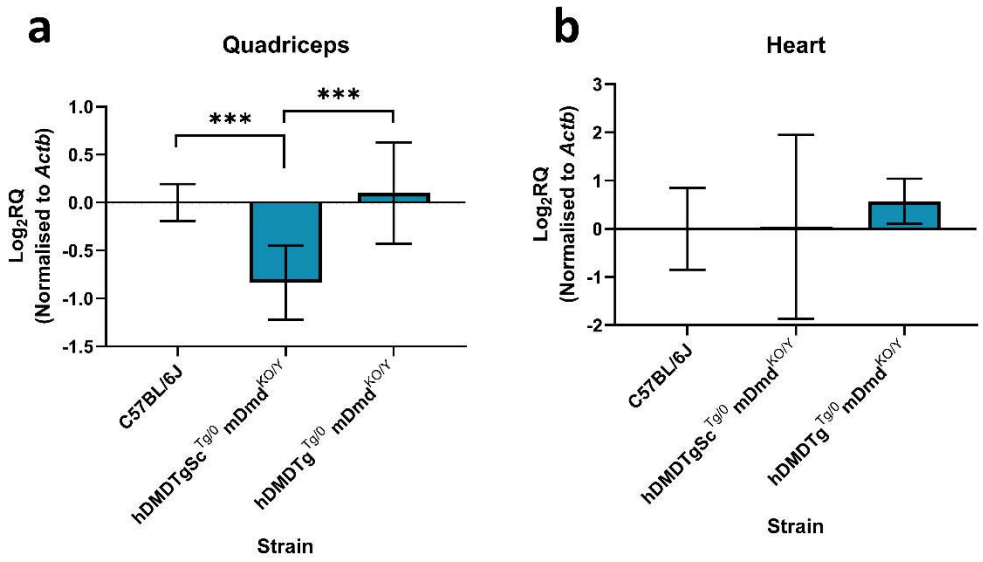


553

554

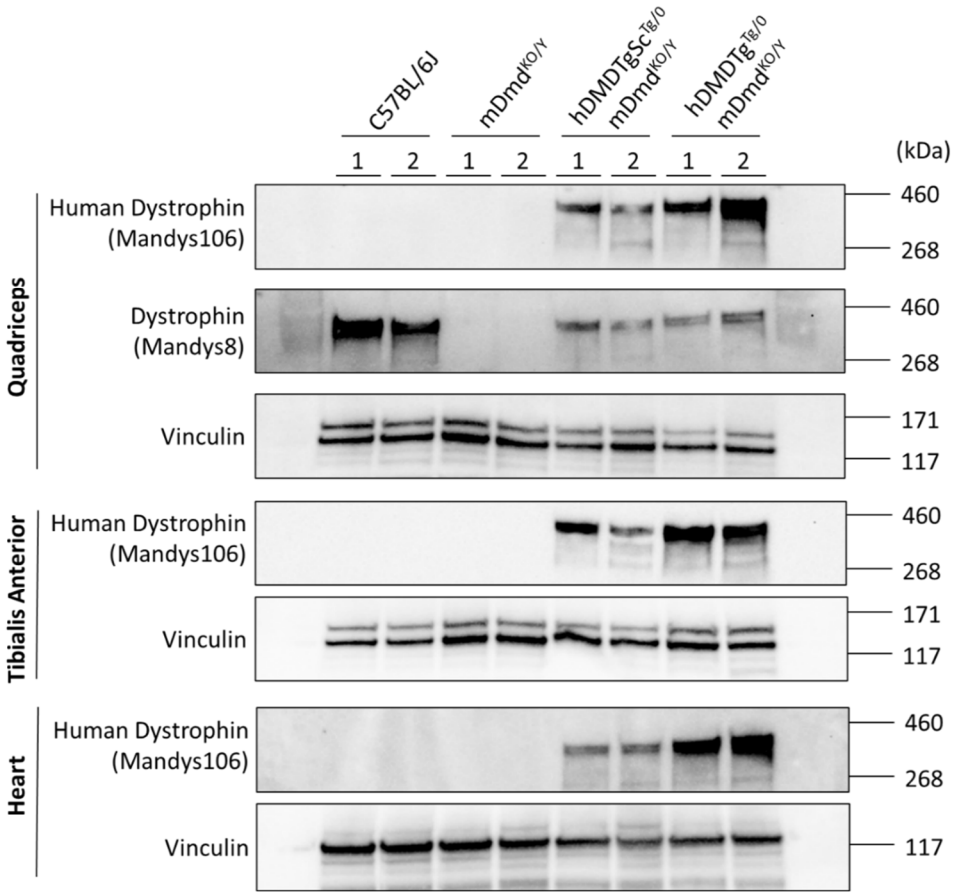
555

556 Figure 4



557

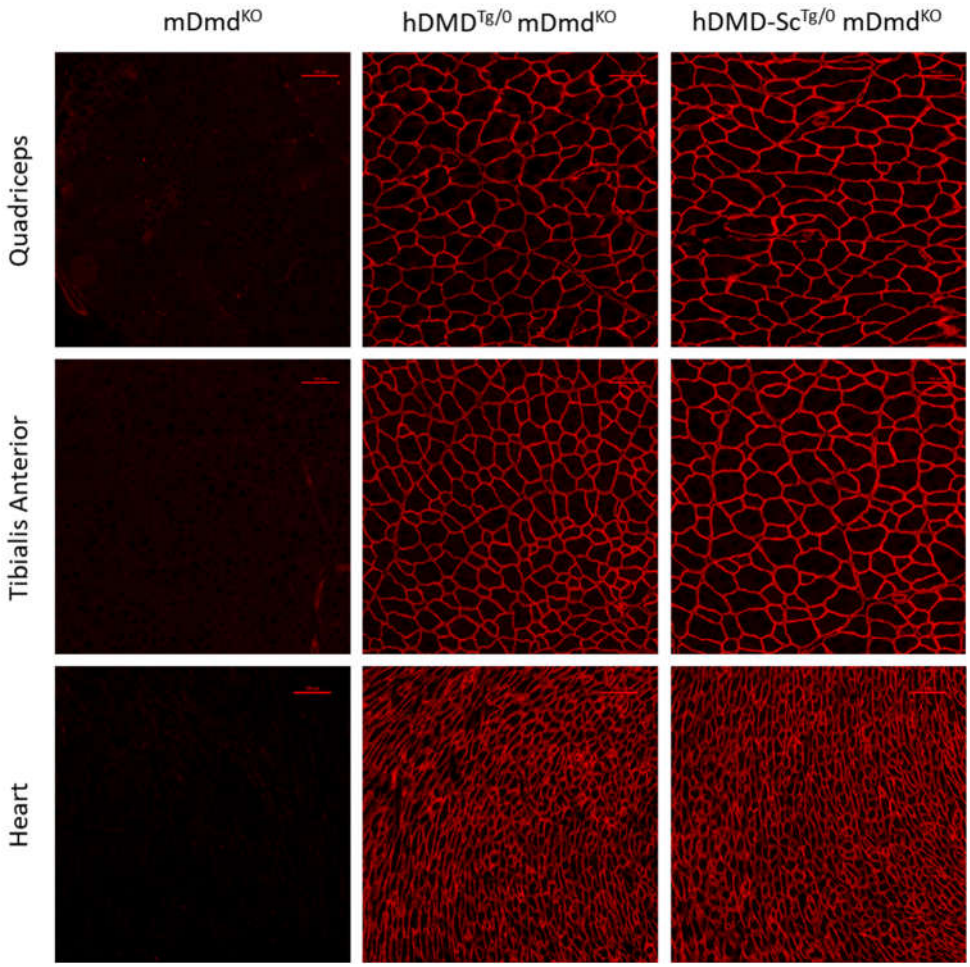
558 Figure 5



559

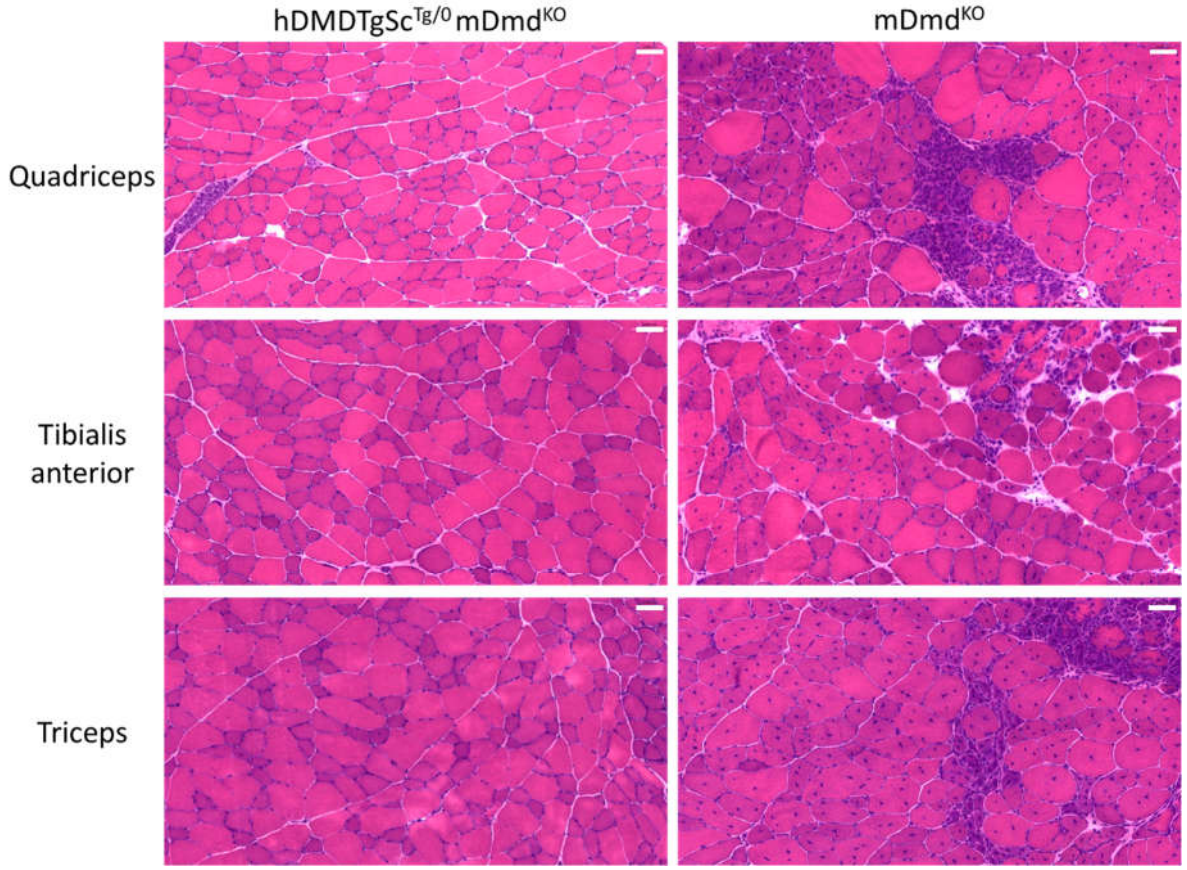
560

561 Figure 6



562

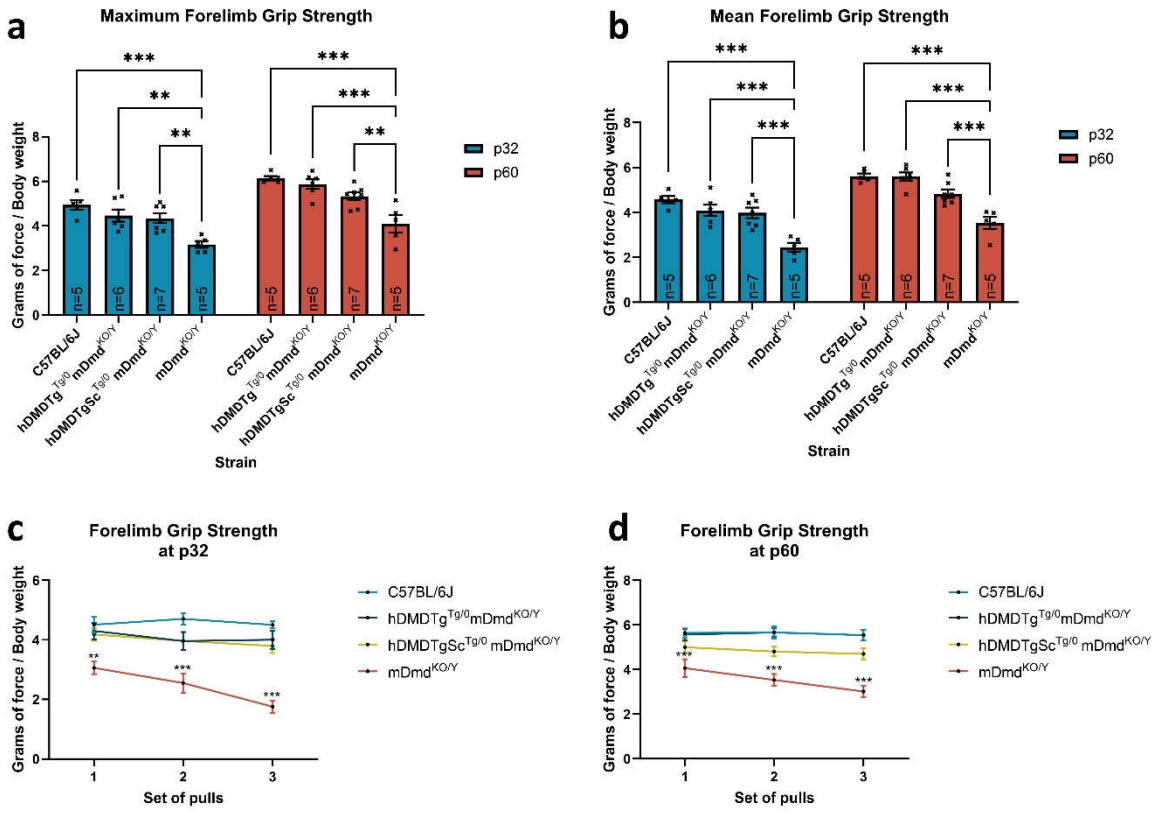
563 Figure 7



564

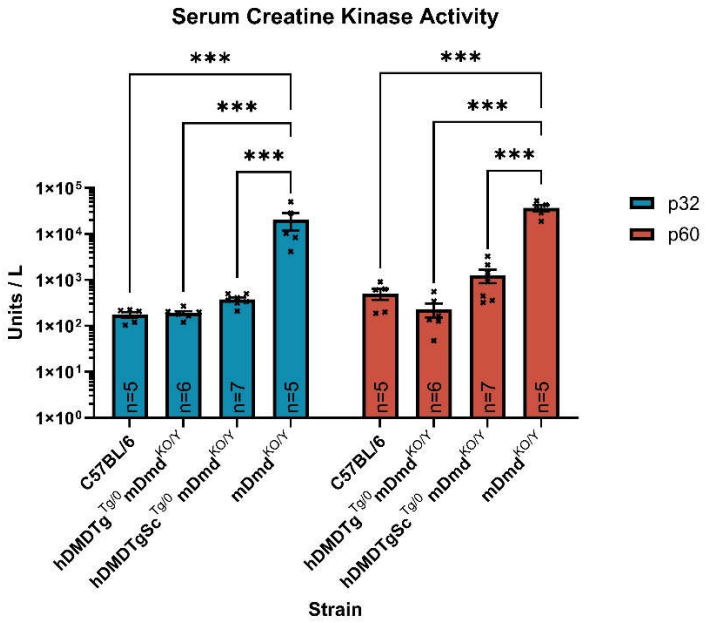
565

566 Figure 8



567

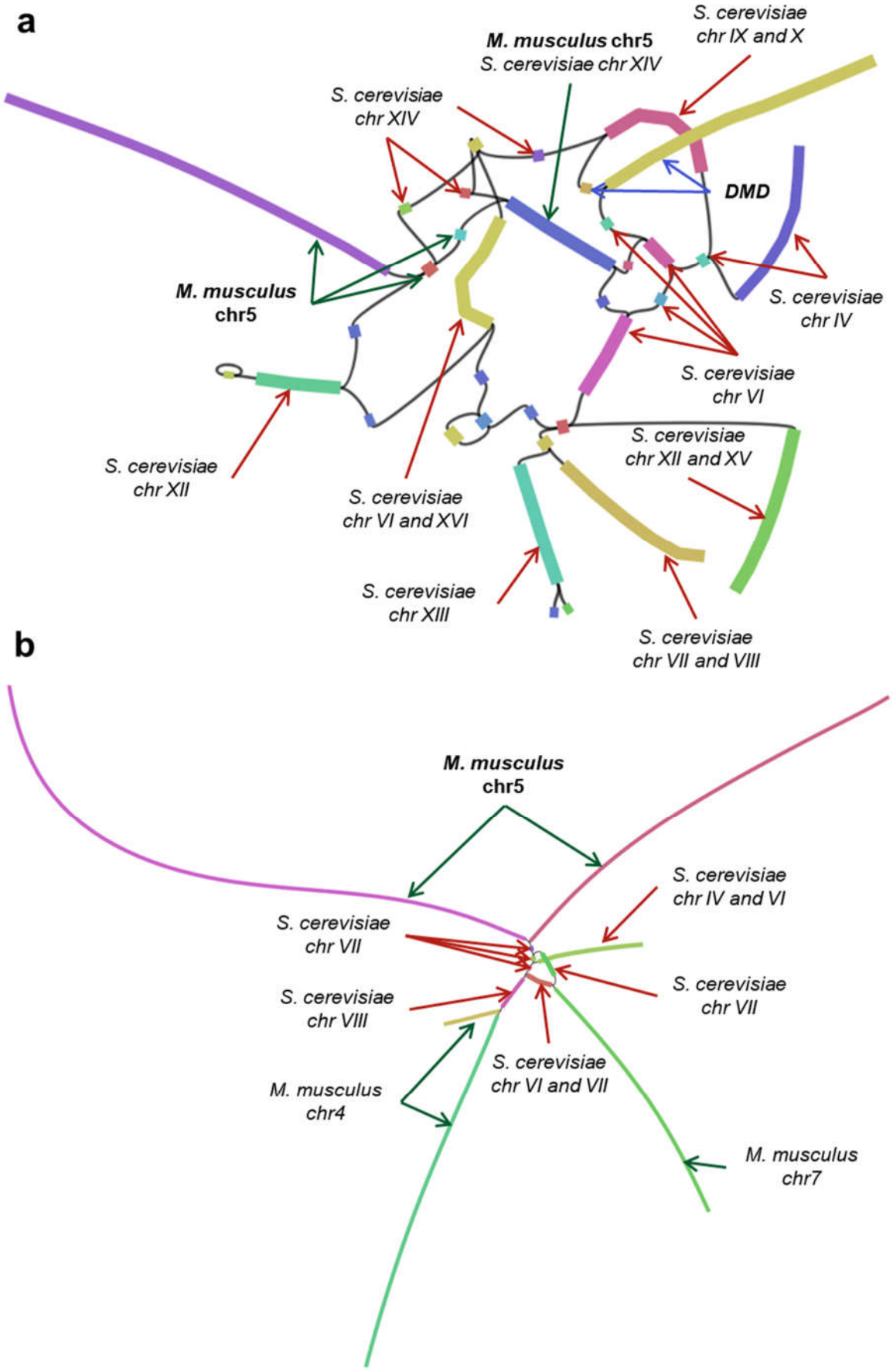
568 Figure 9



569

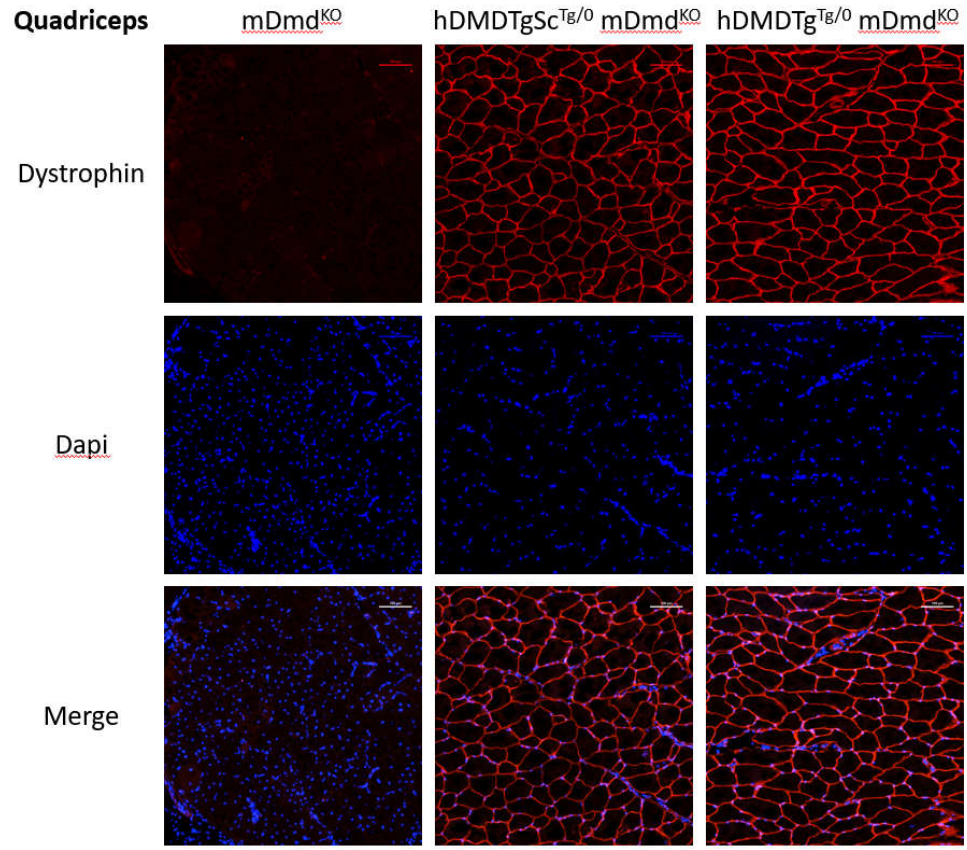
570

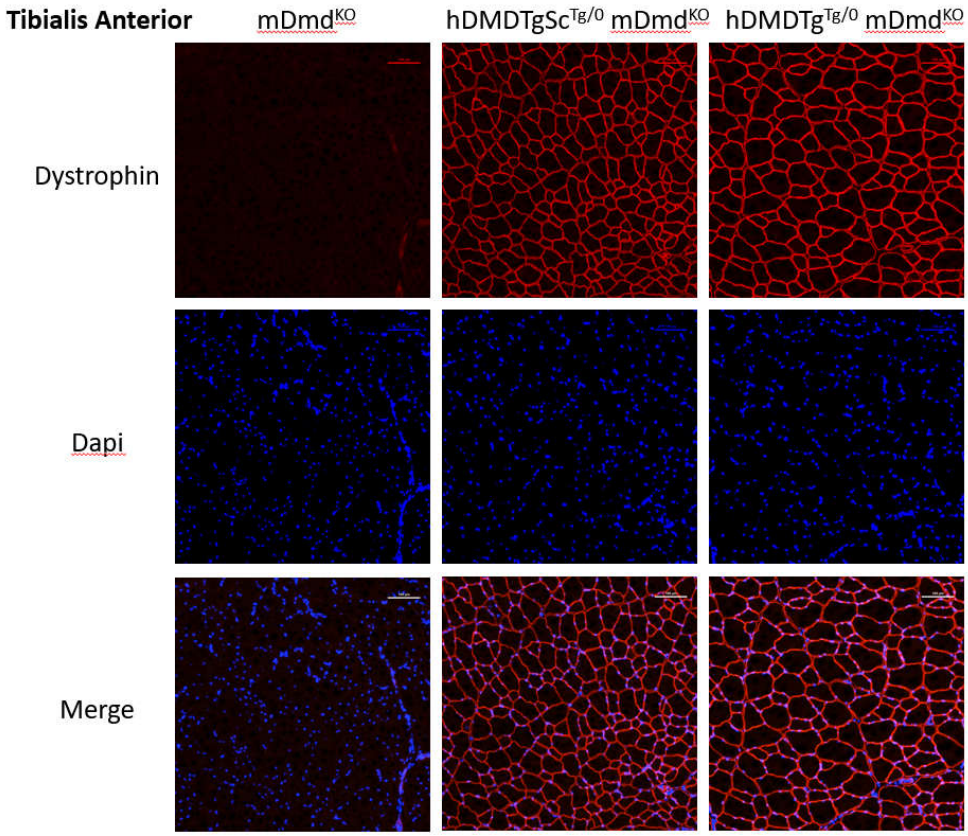
571 Supplementary figure 1

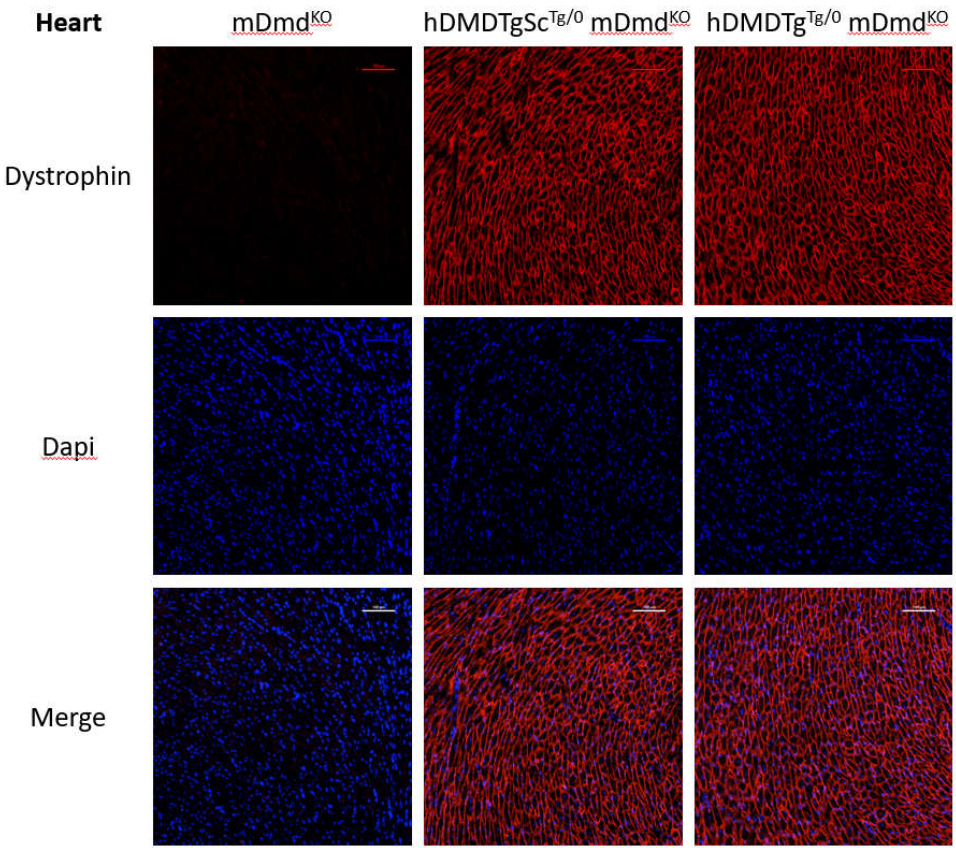


572

Supplementary figure 2







References

1. Min YL, Bassel-Duby R, Olson EN: **CRISPR Correction of Duchenne Muscular Dystrophy.** *Annu Rev Med* 2019, **70:**239-255.
2. Hoffman EP, Brown RH, Jr., Kunkel LM: **Dystrophin: the protein product of the Duchenne muscular dystrophy locus.** *Cell* 1987, **51:**919-928.
3. Kariyawasam D, D'Silva A, Mowat D, Russell J, Sampaio H, Jones K, Taylor P, Farrar M: **Incidence of Duchenne muscular dystrophy in the modern era; an Australian study.** *European Journal of Human Genetics* 2022, **30:**1398-1404.
4. Culligan KG, Mackey AJ, Finn DM, Maguire PB, Ohlendieck K: **Role of dystrophin isoforms and associated proteins in muscular dystrophy (review).** *Int J Mol Med* 1998, **2:**639-648.
5. Kharraz Y, Guerra J, Pessina P, Serrano AL, Muñoz-Cánoves P: **Understanding the process of fibrosis in Duchenne muscular dystrophy.** *Biomed Res Int* 2014, **2014:**965631.
6. Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, Dawkins H, Lamont L, Roy AJ, Chamova T, et al: **The TREAT-NMD DMD Global Database: Analysis of More than 7,000 Duchenne Muscular Dystrophy Mutations.** *Human Mutation* 2015, **36:**395-402.
7. Aartsma-Rus A, Van Putten M: **The use of genetically humanized animal models for personalized medicine approaches.** *Disease Models & Mechanisms* 2020, **13:**dmm041673.
8. t Hoen PA, de Meijer EJ, Boer JM, Vossen RH, Turk R, Maatman RG, Davies KE, van Ommen GJ, van Deutekom JC, den Dunnen JT: **Generation and characterization of transgenic mice with the full-length human DMD gene.** *J Biol Chem* 2008, **283:**5899-5907.

9. Veltrop M, Van Vliet L, Hulsker M, Claassens J, Brouwers C, Breukel C, Van Der Kaa J, Linssen MM, Den Dunnen JT, Verbeek S, et al: **A dystrophic Duchenne mouse model for testing human antisense oligonucleotides.** *PLOS ONE* 2018, **13:**e0193289.
10. Pickar-Oliver A, Gough V, Bohning JD, Liu S, Robinson-Hamm JN, Daniels H, Majoros WH, Devlin G, Asokan A, Gersbach CA: **Full-length dystrophin restoration via targeted exon integration by AAV-CRISPR in a humanized mouse model of Duchenne muscular dystrophy.** *Mol Ther* 2021, **29:**3243-3257.
11. Young CS, Mokhonova E, Quinonez M, Pyle AD, Spencer MJ: **Creation of a Novel Humanized Dystrophic Mouse Model of Duchenne Muscular Dystrophy and Application of a CRISPR/Cas9 Gene Editing Therapy.** *J Neuromuscul Dis* 2017, **4:**139-145.
12. Hiramuki Y, Abe S, Uno N, Kazuki K, Takata S, Miyamoto H, Takayama H, Morimoto K, Takehara S, Osaki M, et al: **Full-length human dystrophin on human artificial chromosome compensates for mouse dystrophin deficiency in a Duchenne muscular dystrophy mouse model.** *Scientific Reports* 2023, **13:**4360.
13. Hoshiya H, Kazuki Y, Abe S, Takiguchi M, Kajitani N, Watanabe Y, Yoshino T, Shirayoshi Y, Higaki K, Messina G, et al: **A highly stable and nonintegrated human artificial chromosome (HAC) containing the 2.4 Mb entire human dystrophin gene.** *Mol Ther* 2009, **17:**309-317.
14. Yavas A, Weij R, van Putten M, Kourkouta E, Beekman C, Puoliväli J, Bragge T, Ahtoniemi T, Knijnenburg J, Hoogenboom ME, et al: **Detailed genetic and functional analysis of the hDMDdel52/mdx mouse model.** *PLoS One* 2020, **15:**e0244215.
15. van Deutekom J, Beekman C, Bijl S, Bosgra S, van den Eijnde R, Franken D, Groenendaal B, Harquouli B, Janson A, Koevoets P, et al: **Next Generation Exon 51 Skipping Antisense Oligonucleotides for Duchenne Muscular Dystrophy.** *Nucleic Acid Ther* 2023, **33:**193-208.
16. Chey YCJ, Arudkumar J, Aartsma-Rus A, Adikusuma F, Thomas PQ: **CRISPR applications for Duchenne muscular dystrophy: From animal models to potential therapies.** *WIREs Mechanisms of Disease* 2023, **15**.
17. Mizuno S, Takami K, Daitoku Y, Tanimoto Y, Dinh TTH, Mizuno-Iijima S, Hasegawa Y, Takahashi S, Sugiyama F, Yagami K-i: **Peri-implantation lethality in mice carrying megabase-scale deletion on 5qc3.3 is caused by Exoc1 null mutation.** *Scientific Reports* 2015, **5:**13632.
18. Kato T, Hara S, Goto Y, Ogawa Y, Okayasu H, Kubota S, Tamano M, Terao M, Takada S: **Creation of mutant mice with megabase-sized deletions containing custom-designed breakpoints by means of the CRISPR/Cas9 system.** *Scientific Reports* 2017, **7:**59.
19. Korablev AN, Serova IA, Serov OL: **Generation of megabase-scale deletions, inversions and duplications involving the Contactin-6 gene in mice by CRISPR/Cas9 technology.** *BMC Genetics* 2017, **18:**112.
20. Boroviak K, Doe B, Banerjee R, Yang F, Bradley A: **Chromosome engineering in zygotes with CRISPR/Cas9.** *genesis* 2016, **54:**78-85.
21. Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, Taylor DJ, Shafin K, Shumate A, Xiao C, et al: **A complete reference genome improves analysis of human genetic variation.** *Science* 2022, **376:**eabl3533.
22. Zhu F, Nair RR, Fisher EMC, Cunningham TJ: **Humanising the mouse genome piece by piece.** *Nat Commun* 2019, **10:**1845.
23. Zhang Y, Li H, Nishiyama T, McAnally JR, Sanchez-Ortiz E, Huang J, Mammen PPA, Bassel-Duby R, Olson EN: **A humanized knockin mouse model of Duchenne muscular dystrophy and its correction by CRISPR-Cas9 therapeutic gene editing.** *Mol Ther Nucleic Acids* 2022, **29:**525-537.
24. Yucel N, Chang AC, Day JW, Rosenthal N, Blau HM: **Humanizing the mdx mouse model of DMD: the long and the short of it.** *NPJ Regen Med* 2018, **3:**4.
25. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F: **Genome engineering using the CRISPR-Cas9 system.** *Nature Protocols* 2013, **8:**2281-2308.

26. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25:**1754-1760.
27. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** *arXiv pre-print server* 2013.
28. Taylor SS, Larin Z, Tyler-Smith C: **Analysis of extrachromosomal structures containing human centromeric alphoid satellite DNA sequences in mouse cells.** *Chromosoma* 1996, **105:**70-81.
29. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P: **Sambamba: fast processing of NGS alignment formats.** *Bioinformatics* 2015, **31:**2032-2034.
30. Cameron DL, Baber J, Shale C, Valle-Inclan JE, Besselink N, van Hoeck A, Janssen R, Cuppen E, Priestley P, Papenfuss AT: **GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing.** *Genome Biol* 2021, **22:**202.
31. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC: **Accurate detection of complex structural variations using single-molecule sequencing.** *Nat Methods* 2018, **15:**461-468.
32. Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, et al: **Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes.** *Nat Biotechnol* 2020, **38:**1044-1053.
33. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S: **Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing.** *Genome Biol* 2022, **23:**258.
34. Wang C, Yue F, Kuang S: **Muscle Histology Characterization Using H&E Staining and Muscle Fiber Type Classification Using Immunofluorescence Staining.** *Bio Protoc* 2017, **7**.
35. Aartsma-Rus A, van Putten M: **Assessing functional performance in the mdx mouse model.** *J Vis Exp* 2014.