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

Yu C. J. Chey[1,2], Mark Corbett[3], Jayshen Arudkumar[1,2], Sandra 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 the 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 show 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 to further generation of DMD disease models, suited for the pre-clinical assessment of sequence-specific 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 mice strains were 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 a 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 prone to contraction-induced mechanical stress and damage, leading to the progressive replacement of muscle with fat and fibrous tissue [5].

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

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

81 The first humanised DMD mouse model with stable germline transmission, Tg(DMD)72Thoen/J
82 (hDMDTg), was created by integrating a yeast artificial chromosome (YAC) containing the full-length
83 human *DMD* (*hDMD*) transgene into chromosome 5 of mouse embryonic stem (ES) cells [8]. When
84 crossed with the *mdx* strain lacking murine dystrophin expression, the resulting hDMDTg/*mdx* mice
85 exhibited functional compensation for the absence of mouse dystrophin through human dystrophin
86 expression. Because it possesses the complete human *DMD*, including proximal regulatory
87 sequences (promoter), exons and introns, the hDMDTg/*mdx* strain serves as a platform for the
88 further creation of humanised DMD disease models by the introduction of mutations into the *hDMD*
89 transgene that mimic various DMD patient mutations. Examples of these mouse disease models
90 include the exon 52 deletion hDMDTgEx52Δ/*mdx* models [9, 10] and the exon 45 deletion
91 hDMDTgEx45Δ/*mdx* and hDMDTgEx45Δ/*mdxD2* models [11]. A second mouse model with the full
92 length human *DMD* gene carried on a non-integrated human artificial chromosome (DYS-HAC) has
93 also been generated, but mice with DYS-HACs containing DMD mutations has not yet been
94 developed [12, 13].

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

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

108

109

110 **Results**

111 **Sequence analysis of hDMDTg mice with duplicated hDMD transgenes**

112 To identify the 5' and 3' boundaries of the *DMD* locus of the hDMDTg transgene, short-read genome
113 sequencing was performed on genomic DNA extracted from the liver of homozygous hDMDTg mice.
114 The sequencing data was mapped to the mouse (GRCm38_68) and human (GRCh38) genome
115 reference sequences. Mapping to the mouse genome produced average depth of 27.68-fold (Figure
116 1). Mapping to the human genome produced an average depth of ~60 fold in the range
117 chrX:31,081,740-33,507,715 strongly suggesting that this region defines the locus included in the

118 original YAC clone used to make the hDMD transgene and confirming that the region was duplicated
119 based on the coverage depth compared to the mouse genome.

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

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

139 Comparison of the mapped hDMDTg sequences revealed that the *DMD* loci within the hDMDTg
140 transgene were not exact duplications. One copy ranged from chrX:31,081,740-33,507,715, and the
141 other chrX:31,081,740-33,495,727 (Figure 2). The contig containing *DMD* from the de novo assembly
142 also captured the tail-to-tail orientation of the transgenes consisting of inverted repeats of *URA3,*

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

145

146 **Generation of the single transgene copy hDMDTgSc mouse model**

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

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

159

160 **Sequence analysis of hDMDTgSc mice**

161 Mapping Nanopore long-read sequence data from hDMDTgSc mice to our custom reference mouse
162 genome with the human *DMD* locus and pYAC4 sequence, we identified 1462 SVs on chr5. Of all SV
163 called between the hDMDTg and hDMDTgSc data, 196 were concordant, one of which was the 174kb
164 duplication on chr5 containing the transgene insertions site. De novo assembly of hDMDTgSc long-
165 read data also captured the transgene insertion site on mouse chr5; however, a contig corresponding
166 to *DMD* was not assembled from these data by Shasta, and thus, this graph contained only mouse

167 and yeast chromosomal sequences (Supplementary Figure 1B). From comparison to mapped
168 hDMDTg sequences, only the shorter copy of the *DMD* locus (chrX:31,081,740-33,495,727) was
169 retained in the single-copy hDMDTgSc transgene (Figure 1).

170

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

172 To assess if the single copy hDMD transgene in hDMDTgSc mice retains functionality, we
173 characterised male hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ and hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice, both strains lacking
174 endogenous mouse dystrophin expression. We performed forelimb grip strength testing followed by
175 serum collection for creatine kinase analysis on mice at postnatal day 32 (p32) and day 60 (p60).
176 Forelimb grip strength testing showed that hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice have similar maximum and
177 mean grip strength to hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice at p32 and that both these strains have no
178 significant difference in strength compared to wildtype C57BL/6J control mice, demonstrating
179 functional rescue of the reduced grip strength caused by the endogenous *mDmd* knock-out (Figure
180 4A,B). hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ and hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ also do not experience the same onset of
181 fatigue caused by repeated measurements as observed in mDmd$^{KO/Y}$ mice (Figure 4C,D).

182 The serum samples collected after the forelimb grip strength testing were used to assess serum
183 creatine kinase (CK) levels, an indirect measure of muscle damage. Both hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ and
184 hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice have a similar reduction of serum CK levels to WT C57BL/6J controls at
185 p32 and p60 (Figure 5).

186 Skeletal and heart muscle tissues were collected from p60 mice for mRNA, protein, and histology
187 analysis. We performed quantitative RT-PCR analysis of the dystrophin transcript in quadriceps and
188 heart muscles using a set of primers that detects both the human and mouse transcript by targeting
189 shared sequences. In quadriceps muscle, the expression level of *hDMD* in hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice
190 is equivalent to that of the *mDmd* in WT C57BL/6J mice, whereas *hDMD* expression in hDMDTgSc$^{Tg/0}$

191 mDmd$^{KO/Y}$ quadriceps is approximately halved (Figure 6A). In heart muscles, *hDMD* transcript in
192 hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice is slightly higher than *mDmd* in WT C57BL/6J mice, while *hDMD*
193 expression in hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ heart is, although somewhat variable, more similar to that of
194 *mDmd* (Figure 6B).

195 Western blots were performed with the human dystrophin-specific MANDYS106 antibody. Full-length
196 dystrophin protein was detected in tibialis anterior (TA), quadriceps and heart muscles of
197 hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice at a lower level compared to hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice (Figure 7). The
198 TA western blot was re-probed with MANDYS8, which is cross-reactive with both human and mouse
199 dystrophin, confirming that MANDYS106 only detects human dystrophin (Figure 7A).

200 Immunofluorescence staining of dystrophin with MANDYS106 in TA, quadriceps and heart muscles
201 showed the expected sarcolemmal localisation of the human dystrophin protein in hDMDTgSc$^{Tg/0}$
202 mDmd$^{KO/Y}$ muscle, as for hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ muscle tissue (Figure 8, Supplementary Figure 2).
203 Histological analysis of hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ muscle sections stained with hematoxylin and eosin
204 (H&E) revealed normal muscle histology with nuclei localised at the sarcolemmal region, consistent
205 muscle fibre size, and no evidence of damage or infiltrating cells (Figure 9). Conversely, mDmd$^{KO/Y}$
206 muscle was dystrophic, with many centralised nuclei indicating muscle regeneration as well as the
207 presence of cellular infiltrates.

208

209

210 **Discussion**

211

212 Validated humanised animal models are important tools for pre-clinical evaluation of genetic
213 therapies as they permit in vivo testing of sequence-specific reagents that can be directly translated
214 into patients. Generation of a hDMDTg mouse model containing the entire *hDMD* locus randomly

215 integrated into mouse chromosome 5 was a significant advance, particularly given the size of the
216 transgene (2.4 Mb). However, it was subsequently shown that two copies of the transgene had
217 inserted, which does not accurately reflect the copy number of the *DMD* gene [14]. Furthermore,
218 derivatives of the hDMDTg line such as the del52 exon deletion disease model are more difficult to
219 engineer and validate. While these double-copy humanised models can be used for the testing of
220 ASOs [15], they are not ideal for evaluating strategies that are reliant on DNA double-stranded breaks
221 such as CRISPR-Cas9 gene-editing [16]. This is because large intervening deletions between the target
222 sites of the two transgene copies can complicate the analysis of DNA repair outcomes.

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

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

236 There are some limitations to the genetic analyses we performed via next-generation sequencing.
237 Using only nanopore long-read sequences, we were not able to unambiguously assemble the full-
238 length hDMDTgSc and hDMDTg transgenes despite achieving a total assembly length similar to the
239 length of the mm10 genome reference and a relatively high contig N50 (Supp Table 1). The exact

240 assembly and ordering of the yeast chromosomal sequences that flank the DMD loci in the
241 duplicated and single-copy transgenes remains uncertain. Additional long-read sequencing with high-
242 quality PacBio circular consensus sequences would likely be sufficient to create a near-continuous
243 assembly of the transgenic mouse genome as has been done for the human T2T reference sequence
244 [21].

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

255 Our hDMDTgSc mouse model has considerable potential for the pre-clinical testing of sequence-
256 specific ASOs or CRISPR therapeutics as, in principle, any causative DMD mutation can be engineered
257 into the single-copy 2.4 Mb *hDMD* transgene, such as models lacking *DMD* exons 45 or 52 like those
258 previously generated from the double transgenic mice. This enables ready generation of humanised
259 DMD disease mouse models, which has an advantage over the more labour-intensive and inefficient
260 partial sequence humanisation methods where exon knock-ins must be performed for each separate
261 exonic target [23]. Crossing these human transgene knock-out mice with mice lacking endogenous
262 dystrophin expression would result in a disease model suitable for the testing of CRISPR therapy
263 strategies, and could also be inter-crossed with a mouse strain lacking both endogenous dystrophin
264 and utrophin expression to simulate an even more severe dystrophic phenotype [24].

265

266

267 **Materials and Methods**

268

269 **Animal studies**

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

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

279

280 **Mouse generation**

281 To create the intervening deletion of a single copy of the *hDMD* transgene, we designed SpCas9
282 gRNAs targeting the Neomycin and Hygromycin antibiotic resistance cassettes in the integrated YAC
283 sequence. Complementary gRNA oligos with overhangs facilitating golden gate assembly were
284 phospho-annealed and inserted into pSpCas9(BB)-2A-Puro (PX459) V2.0 [25]. pSpCas9(BB)-2A-Puro
285 (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988 ; http://n2t.net/addgene:62988 ;
286 RRID:Addgene_62988). Following plasmid validation by *BbsI* diagnostic digest and Sanger sequencing
287 (Australian Genome Research Facility), gRNAs were amplified using NEB Phusion in HF buffer with

288 primers containing a 5' T7 promoter sequence. PCR amplicons purified using the QIAquick PCR
289 Purification Kit (Qiagen) were used to generate gRNA using the HiScribe T7 Quick High Yield RNA
290 Synthesis Kit (NEB) and purified using the RNeasy Mini Kit (Qiagen).

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

296 mDmd^{KO} mice were generated by deletion of *Dmd* Exon 51 via CRISPR microinjection of C57Bl/6J
297 oocytes (manuscript submitted).

298

299 **DNA PCR and Sanger sequencing**

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

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

306

307 ***Short-read genome sequencing***

308 Library prep and sequencing was performed as a service by the Australian Genome Research Facility
309 (AGRF). Genomic DNA extracted from hDMDTg mice liver were sheared and PCR-free sequencing
310 libraries prepared. From these libraries, 150 base pair, paired-end reads were generated on a

311 NovaSeq 6000. Image analysis was performed in real time by the NovaSeq Control Software (NCS)
312 v1.7.5 and Real Time Analysis (RTA) v3.4.4. Then the Illumina bcl2fastq v2.20.0.422 pipeline was used
313 to generate the final fastq files.

314 *Mapping*

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

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

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

326 *Variant calling*

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

328 ***Oxford nanopore genome sequencing***

329 High molecular weight genomic DNA was extracted from homozygous hDMDTg and hDMDTgSc mice
330 liver using the Monarch HMW DNA Extraction Kit for Tissue (NEB). Oxford Nanopore Technologies
331 Promethion genome sequencing was performed as a service by Novogene, (Novogene HK). Sequence
332 libraries were prepared using the DNA ligation sequencing protocol with the SQK-LSK-110 kit.
333 Sequence data were generated using a Promethion 9.4.1 flow cell. Bases were called with bonito

334 v0.6.2 (Oxford Nanopore Technologies) with the dna_r9.4.1_e8_sup@v3.3 super accurate base
335 calling model.

336 *Mapping*

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

340 *Variant calling*

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

343 *Assembly*

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

351

352 **Transgene copy number determination**

353 qPCRs were performed to quantify the *hDMD* transgene copy number and were done in technical
354 triplicate. Each reaction well was set up using SYBR™ Green PCR Master Mix (Applied Biosystems), 50
355 nM of each forward and reverse primer and 10 ng gDNA in a 15 µL reaction volume. qPCR data was
356 normalised to endogenous *Sox1.* Reactions were run on the QuantStudio™ 3 Real-Time PCR System

357 (Applied Biosystems) using the Fast Run Mode with the following reaction conditions: 95°C for 20 s,
358 followed by 40 cycles of 95°C for 1 s and 60°C for 20 s, and a dissociation protocol to obtain a melt
359 curve for all samples running from 60°C to 95°C.

360

361 **Forelimb grip strength test**

362 Muscle strength measurement was performed by a forelimb grip strength behaviour task [34]. Mice
363 were tested at p32 and p60 using a grid attached to a Centor Easy 10N (Andilog) force gauge. Mice
364 were suspended from the tail and allowed to grasp the grid tightly with both front paws, then pulled
365 along a straight horizontal line away from the force gauge until their grip on the grid was broken.
366 Measurements were repeated 4 times per set, and 3 sets were performed with a 1-minute rest
367 interval between sets. The highest measured rep was recorded for each set and normalised against
368 body weight. Max force reported is the highest measured rep among all sets, and mean force
369 reported is the average of the highest measured reps for each of all 3 sets.

370

371 **Serum creatine kinase measurement**

372 30 mins after grip strength testing, Emla 5% local anaesthetic cream was applied to the mice's tails,
373 and they were warmed in a Mini Thermacage (datesand) for 10 mins at 36°C to dilate the tail vein.
374 Blood samples were collected in microvettes from a nick to the lateral tail vein. The blood was
375 allowed to clot at room temperature for 30 mins, and the clot was separated from the serum by
376 centrifuging at 2 000 g for 10 mins at 4°C and immediately stored at -80°C. Mouse serum creatine
377 kinase (CK) levels were measured by Gribbles Veterinary Pathology using the ADVIA Creatine Kinase
378 CK_L assay on the ADVIA 1800 Chemistry System.

379

380 **RNA extraction and qRT-PCR**

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

389 qRT-PCRs were performed to quantify dystrophin transcript expression in technical and biological
390 triplicate. The primers used was designed to target shared sequences in both *hDMD* and *mDmd*
391 transcript. Each reaction well was set up using SYBR™ Green PCR Master Mix (Applied Biosystems),
392 50 nM of each forward and reverse primer and 1 µL of cDNA in a 15 µL reaction volume. qPCR data
393 was normalised to beta-actin (*Actb*) expression. Reactions were run on the QuantStudio™ 3 Real-
394 Time PCR System (Applied Biosystems) with the same parameters listed in the transgene copy-
395 number determination section above. Analysis was performed using the Design and Analysis
396 Software 2.7.0. Data presented as $\log_2$(RQ) to show proportional changes.

397

398 **Western blot**

399 Tissues were bead homogenised in disruption buffer (75 mM Tris HCl pH 6.8, 15% SDS, 1x cOmplete
400 Protease Inhibitor Cocktail EDTA (Roche)), using Lysing Matrix D (MP Biomedicals). 1/10 diluted
401 lysates were quantified using the Pierce BCA protein assay (Thermo Scientific). 30 µg of protein
402 lysates were mixed with equal parts 2x Loading buffer (21% w/v glycerol, 0.001% w/v bromophenol
403 blue, 5% β-Mercaptoethanol), boiled at 95°C for 5 mins and spun at 21 300 rcf for 5 mins prior to
404 loading in NuPAGE Tris-Acetate gels (Invitrogen). Gels were run at 150 V for 1 h and transferred onto
405 0.2 µm PVDF membranes using the High MW setting on the Trans-Blot Turbo (Bio-Rad). Membranes

406 were blocked with 10% skim milk in TBS-T for 2h with gentle agitation at RT. Dystrophin was probed
407 with 1:500 Anti-Dystrophin antibody, clone 2C6 MANDYS106 (Sigma-Aldrich) in 2% skim milk in TBS-T
408 overnight at 4°C. The following day, membranes were washed three times for 5 mins with TBS-T and
409 incubated with 1:5 000 secondary anti-mouse antibody in 2% skim milk in TBS-T for 1h at RT. After
410 another set of TBS-T washes, membranes were developed with SuperSignal West Pico Plus
411 Chemiluminescent Substrate (Thermo Scientific), and chemiluminescent detection was performed on
412 the ChemiDoc MP system (Bio-Rad) using automated exposure settings for intense bands.
413 Membranes were stripped with Restore PLUS Western Blot Stripping Buffer (Thermo Scientific) prior
414 to re-probing. All dystrophin was probed with 1:500 anti-dystrophin antibody MANDYS8. Vinculin
415 was probed with 1:1 000 Anti-vinculin antibody, V9131 (Sigma-Aldrich).

416

417 **Immunofluorescence**

418 Fresh muscle was snap-frozen in dry-ice-cooled isopentane and immediately embedded in Tissue-Tek
419 O.C.T. 10 µm thick cryotome-sliced transverse sections were permeabilised with PBS + 0.3% Triton X-
420 100 for 10 minutes, blocked with 10% FCS in PBS + 0.3% Triton X-100 for 1 h at RT, and treated with
421 1x ReadyProbes™ Mouse-on-Mouse IgG Blocking Solution (Invitrogen) as per manufacturer
422 instructions. Slides were incubated with 1:100 Anti-Dystrophin antibody, clone 2C6 MANDYS106
423 (Sigma-Aldrich) diluted in blocking solution overnight at 4°C in a humidified chamber. The next day
424 slides were washed with PBS for 3x 5mins and incubated with secondary anti-mouse antibody
425 conjugated with Alexa Fluor® 594 for 1h at RT. Slides were washed with PBS for 3x 5mins and
426 mounted with ProLong™ Gold Antifade Mountant with DAPI (Invitrogen) and coverslip. Images were
427 taken on a Nikon Eclipse Ti2 microscope with a DS-Qi2 camera. Same image processing was applied
428 to images from the same muscle group.

429

430 **Histological analysis of muscles**

431 Fresh muscle was mounted on 10% gum tragacanth on a piece of cork, snap frozen in liquid-nitrogen-
432 cooled isopentane and immediately stored at -80°C. For skeletal muscles, 10 µm thick cryotome-
433 sliced transverse sections were stained with Lillie Mayer's Hematoxylin and Eosin, as described by
434 Wang et al. (2017) with the following modifications to the dehydration steps: 2 dips of 70% ethanol,
435 4 dips of 95% ethanol, 2x 1 min of 100 ethanol and 2x 1 min of xylene. For heart muscles, 10 µm
436 thick cryotome-sliced transverse sections were stained with Lillie Mayer's Hematoxylin for 6 mins,
437 rinsed with running tap water for 1min, destained with 0.5% acid alcohol and developed in Scott's
438 tap water for 1 minute [35]. After rinsing with running water for 2 minutes, slides were stained with
439 eosin for 3 mins and quickly rinsed with running tap water before dehydrating through an increasing
440 concentration of ethanol and xylene as described in the skeletal muscle protocol. Slides were
441 mounted with DPX mountant (Sigma-Aldrich) and coverslip. Images were taken using the 3DHISTECH
442 Pannoramic 250 Flash Slide Scanner at 40x magnification. Same image processing was applied to all
443 images.

444

445 **List of figure legends**

446

447 **Figure 1: Coverage of the human *DMD* locus in the duplicated (hDMDTg) and single copy**
448 **(hDMDTgSc) transgenes.** Mapping depths at the *DMD* locus for short-read and long-read genomes
449 mapped to the human genome. The red box on the chromosome X ideogram at the top shows the
450 approximate location of the coverage shown. Graphs show the coverage of mapped reads for the
451 short-read (hDMDTg-SR) and long-read (hDMDTg-LR, hDMDTgSc-LR) genomes at this location. The
452 *DMD* exon model is shown below the coverage graphs. At the base of the image a zoomed in section

453 chrX: 33490727-33509715, shows the drop in coverage from position chrX:33,495,727 onwards in
454 the hDMDTg data and no coverage from this position in the hDMDTgSc mice.

455

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

459

460 **Figure 3 : Genotyping hDMDTgSc mice. (A)** Detection of the single transgene deletion in G0 founder
461 mice by PCR. Intervening deletion between the Neomycin resistance (NeoR) and Hygromycin
462 resistance (HygR) cassettes was identified in Founder number 5 from the absence of PCR
463 amplification using primers flanking the NeoR cut site and the presence of amplicon using a forward
464 primer in NeoR and a reverse primer in HygR. The green text indicates the founder with the banding
465 expected from successful transgene deletion. **(B)** Real-time qPCR copy-number determination of the
466 human *DMD* transgene. Assay performed on genomic DNA extracted from ear notch biopsies of
467 hDMDTg$^{Tg/0}$ mDmd$^{+/Y}$, hDMDTgSc$^{Tg/0}$ mDmd$^{+/Y}$ (male), hDMDTgSc$^{Tg/0}$ mDmd$^{+/+}$ (female) and wildtype
468 C57BL/6J (mDmd$^{+/Y}$) mice. Measurement of three targets, 5' untranslated region (UTR), exon 50 and
469 3' UTR, were normalised against the endogenous mouse *Sox1* gene. Transgene copy number was
470 successfully reduced from two to one in heterozygous hDMDTgSc mice. Error bars indicate the 95%
471 confidence interval of technical replicates.

472

473 **Figure 4: Forelimb grip strength testing to measure muscle function and performance of mice.**
474 Strength (grams of force) of hDMDTg$^{Tg/0}$ mDmd^{KO} and hDMDTgSc$^{Tg/0}$ mDmd^{KO} mice were compared
475 against WT C57BL/6J and mDmdΔEx51 (mDmd$^{KO/Y}$) controls, normalised to body weight at p32 and
476 p60. (A) Maximum force is the highest recorded measure among three sets of five pulls, and (B)

477 mean force is the average of the highest recorded measures of each of the three sets. Data
478 presented as mean ± SEM, two-way repeat measures ANOVA with Tukey's multiple comparisons test
479 ** P < 0.01, *** P < 0.001. Samples sizes (n) indicated in within bars in graph. (C,D) Averages of the
480 highest recorded measures for each set of pulls, presented as mean ± SEM, two-way repeat
481 measures ANOVA, Dunnett's multiple comparisons test with C57BL/6J as control group, ** P < 0.01,
482 *** P < 0.001.

483

484 **Figure 5: Serum creatine kinase (CK) biomarker level measurement of mice.** Serum CK was assayed
485 as an indirect measure of muscle damage and sarcolemma leakage at p32 and p60. The single-copy
486 transgene in hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice rescues the elevated serum CK caused by endogenous
487 *mDmd* knock-out to levels that are not significantly different than either WT C57BL/6J or hDMDTg$^{Tg/0}$
488 mDmd$^{KO/Y}$ mice. Data presented as mean ± SEM, two-way repeated measures ANOVA with Tukey's
489 multiple comparisons test *** P < 0.001. Samples sizes (n) indicated in first graph.

490

491 **Figure 6: Dystrophin transcript measurement by qRT-PCR.** Quantification of (A) skeletal and (B)
492 heart tissues from p60 mice was performed using a set of primers that target both human *DMD* and
493 mouse *Dmd* transcript sequences, normalised against beta-actin (*Actb*). Data presented as $\log_2$
494 relative quantity (RQ) ± 95%CI, one-way ANOVA with Tukey's multiple comparisons test performed of
495 ddCt values, *** P < 0.001. n=3.

496

497 **Figure 7: Detection of human dystrophin protein in muscle tissue by western blot.** Analysis was
498 performed on skeletal hindlimb muscles, tibialis anterior and quadriceps, as well as heart muscle
499 protein lysates from p60 mice using the human-specific antibody against dystrophin Mandys106. The
500 tibialis anterior blot was also re-probed with the cross-reactive antibody Mandys8 to demonstrate

501 the species-specificity of Mandys106. Vinculin was probed as the loading control. Human dystrophin
502 protein expression in hDMDTgSc$^{Tg/0}$ mDmd$^{KO/Y}$ mice is reduced in all three tissues compared to
503 hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ mice.

504

505 **Figure 8: Immunofluorescence localisation of the human dystrophin protein.** Immunostaining was
506 performed on tibialis anterior, quadriceps and heart muscle tissue sections from p60 mice using the
507 Mandys106 primary antibody. Scale bar 100 µm.

508

509 **Figure 9: Histological analysis of hDMDTg$^{Tg/0}$ mDmd$^{KO/Y}$ and mDmd$^{KO/Y}$ mice.** Hematoxylin and eosin
510 staining was performed on quadriceps, tibialis anterior and triceps muscle tissue sections from p60
511 mice. Scale bar 50 µm.

512

513

514

515

516

517

518

519

520

521

522 **Supplementary materials**

523

524 **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

525

526 **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

527

528 **Supplementary Figure legends**

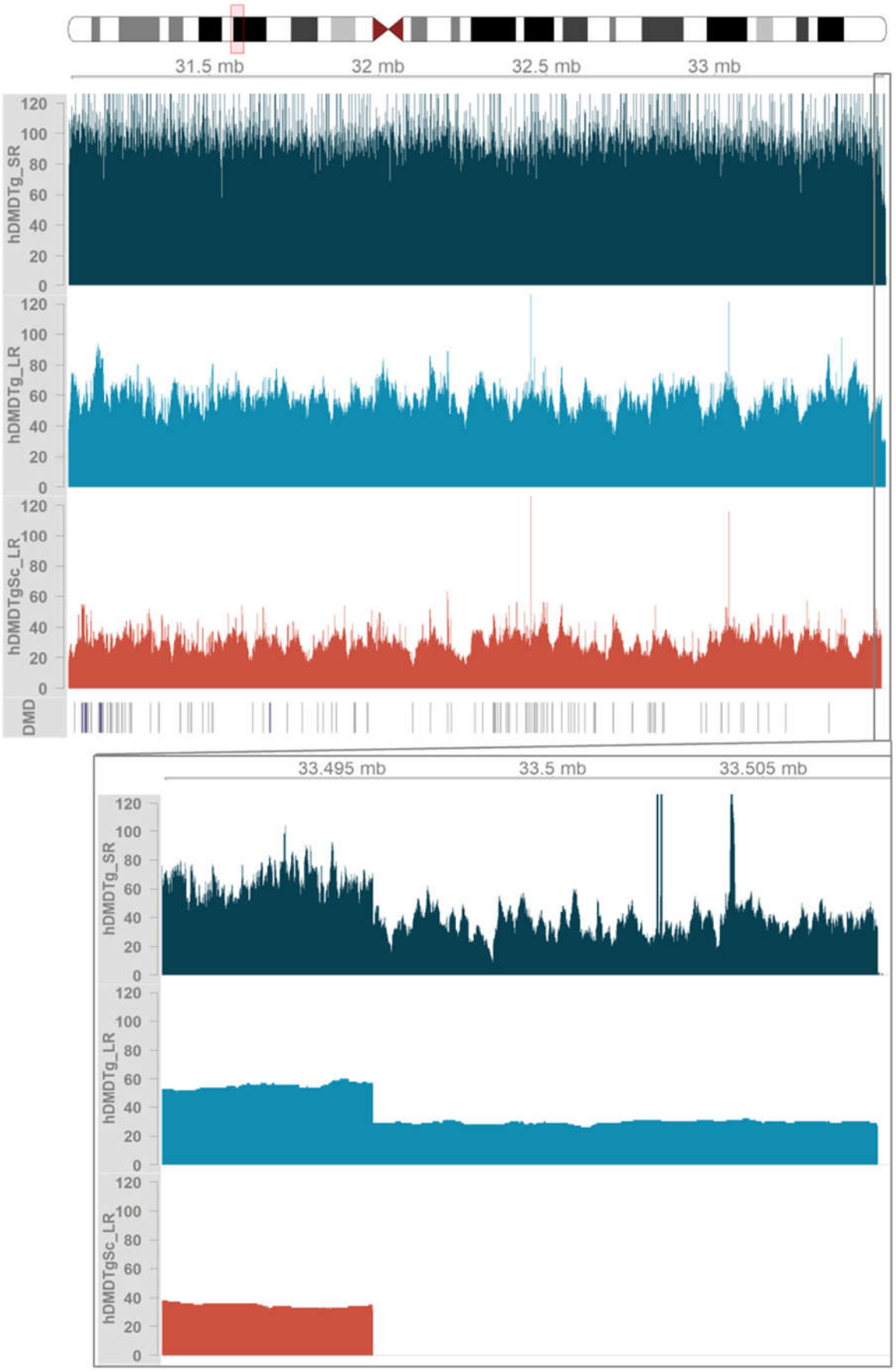
529 **Supplementary Figure 1:** (A) Assembly sequence graph generated by Shasta of the sequences linked
530 to the human *DMD* locus. Each of the thicker coloured nodes represents a different contig from the
531 assembly. Labels indicate the best-matched identities of four sampled sequences of approximately
532 1.5 kb from the beginning, middle and end of each contig to BLAST searches of the non-redundant

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

541

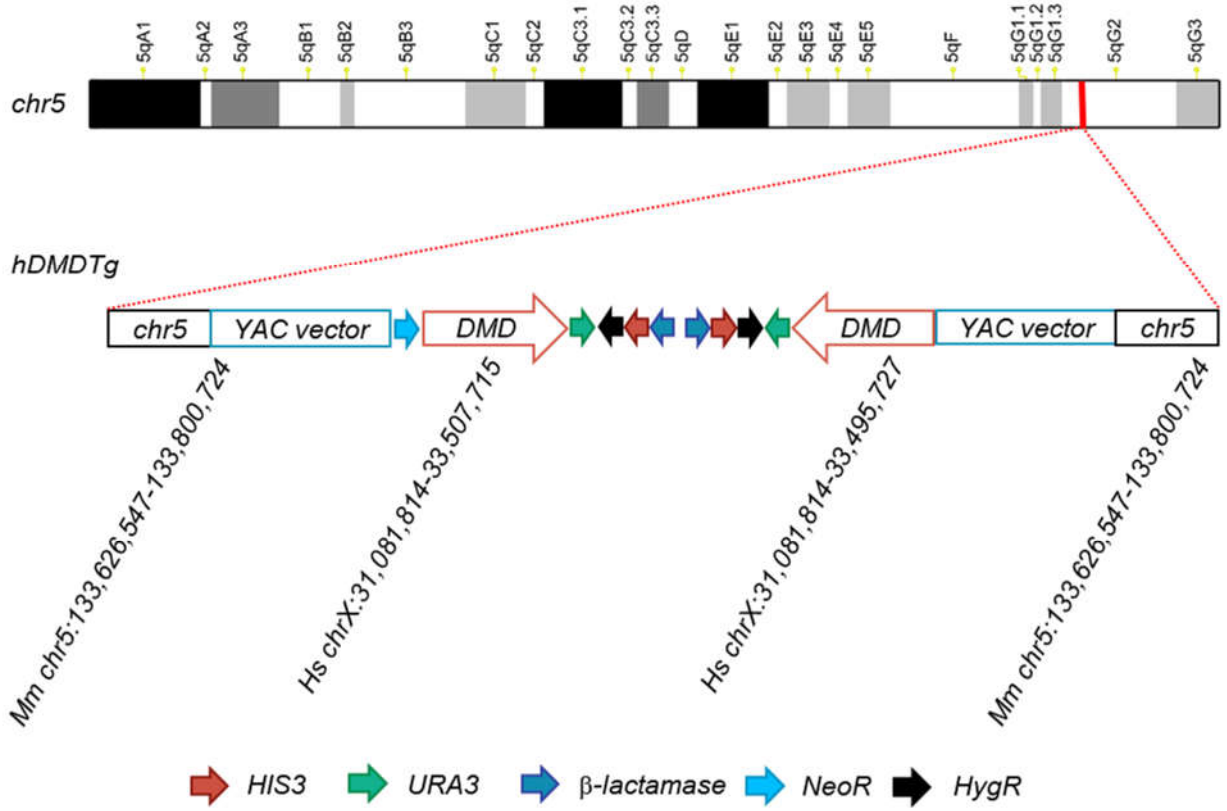
542 **Supplementary Figure 2:** Dapi and merged IF images. Scale bar 100 µm.

543 Figure 1



544

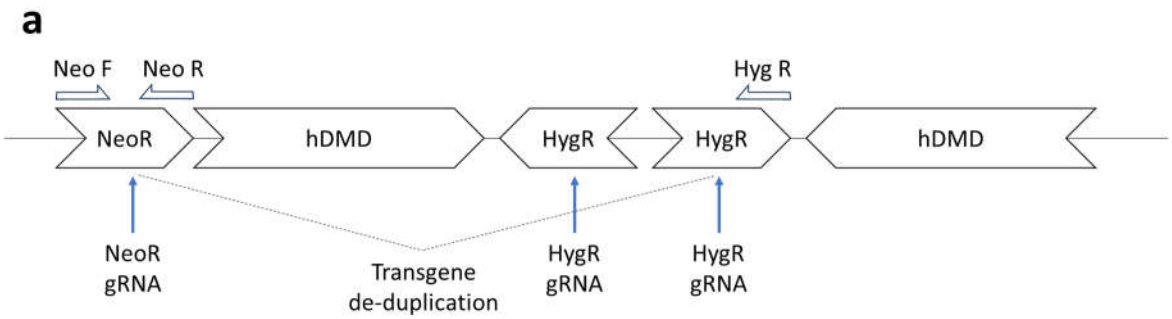
545 Figure 2



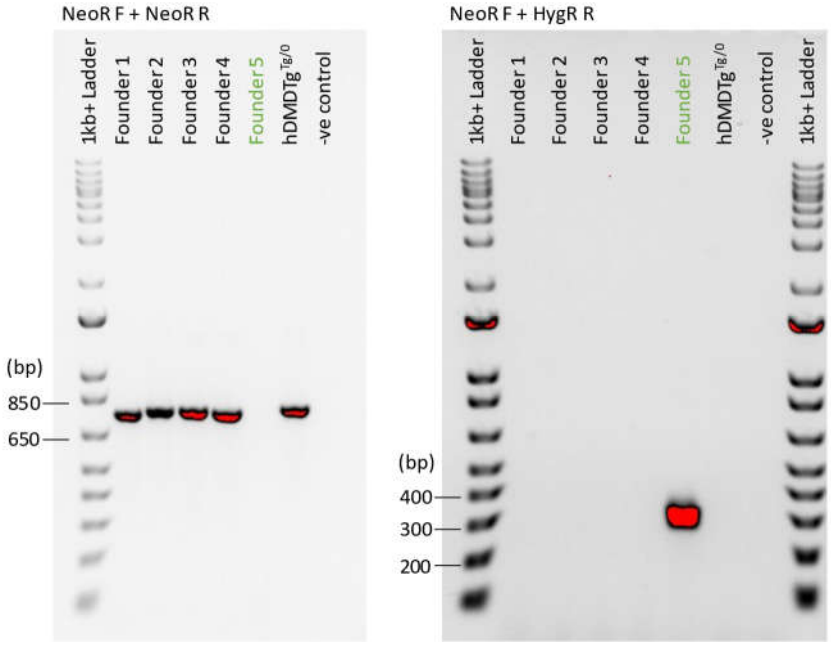
546

547

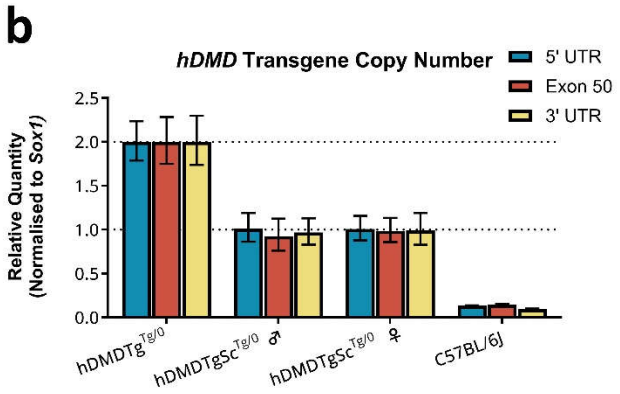
548 Figure 3



549



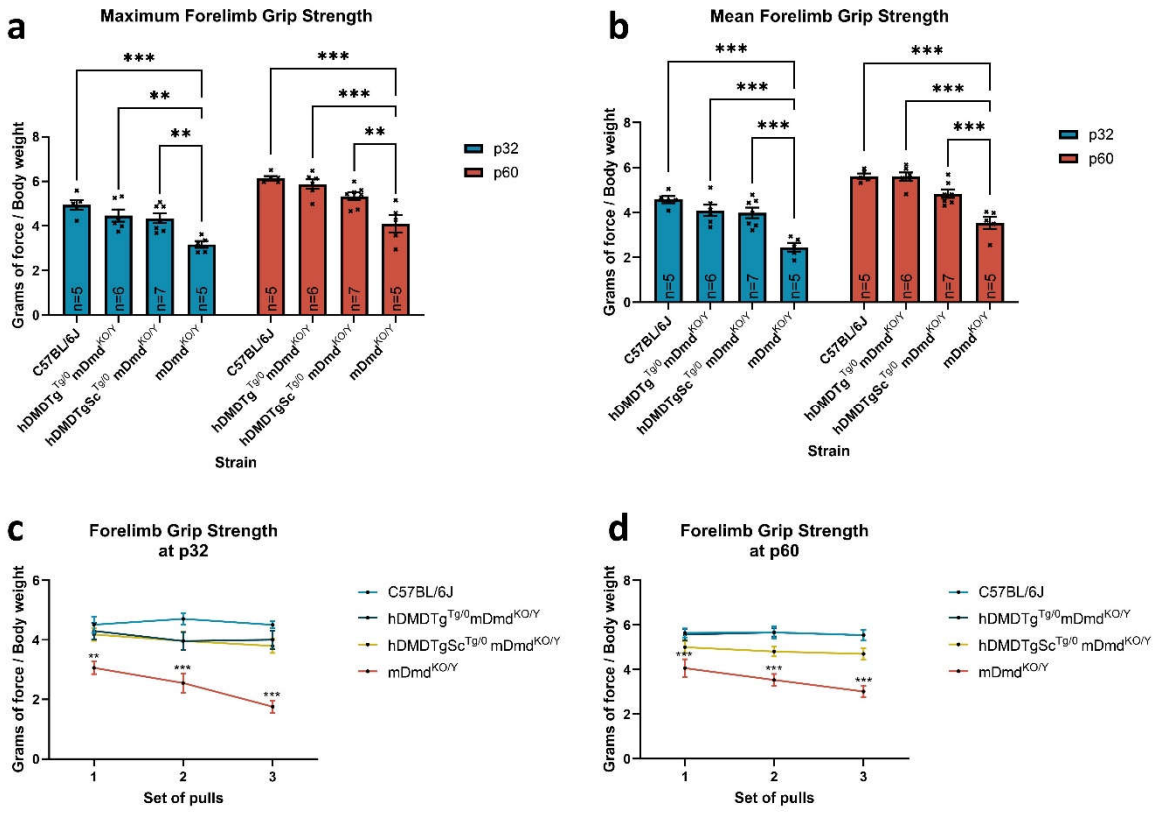
550



551

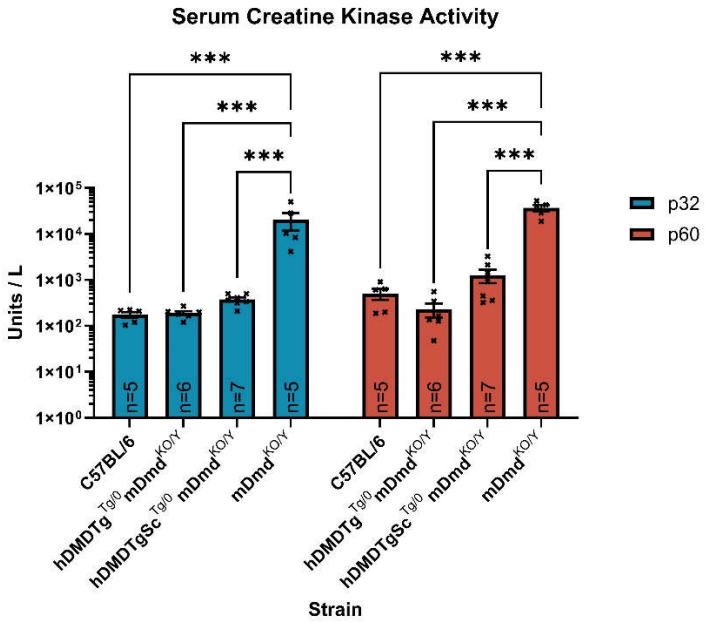
552

553 Figure 4



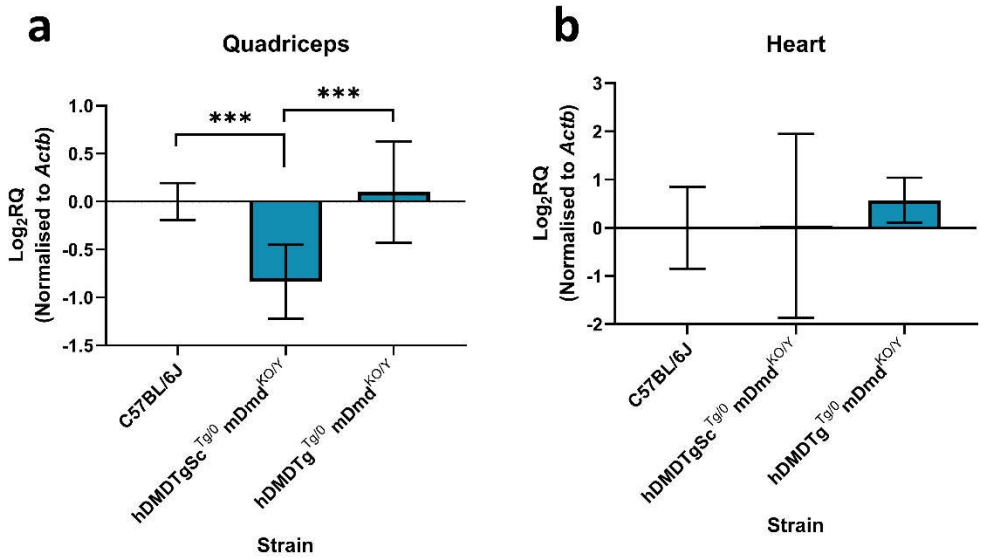
554

555 Figure 5



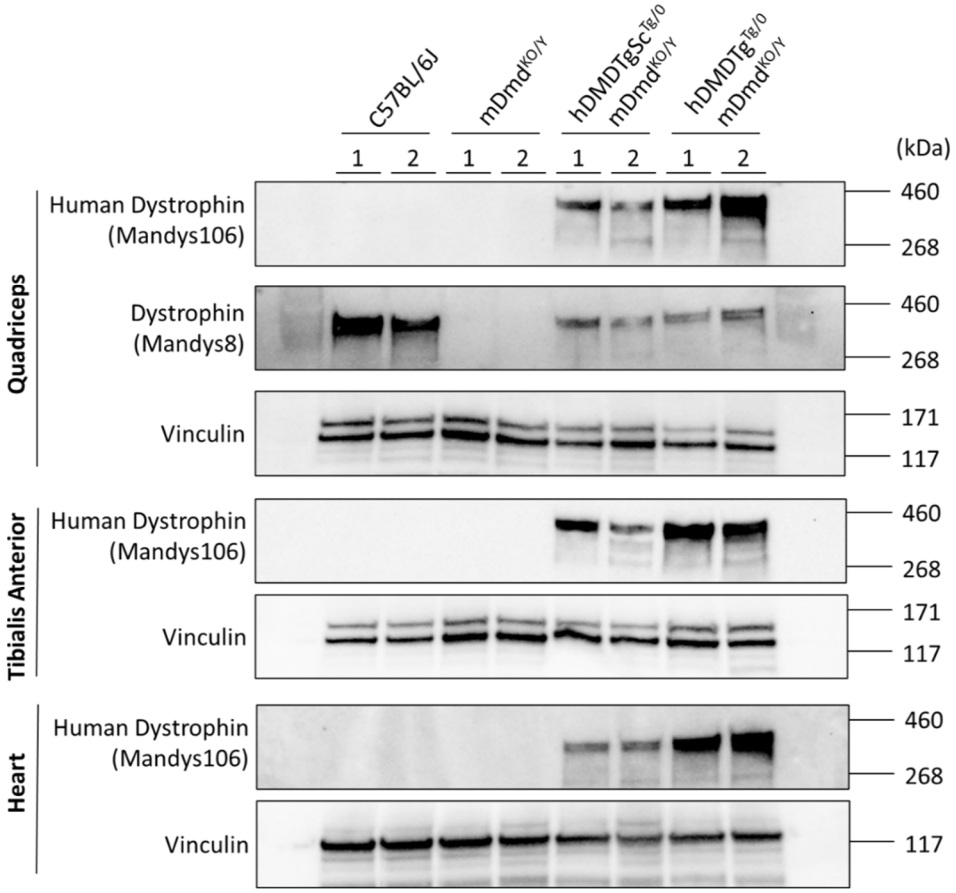
556

557 Figure 6



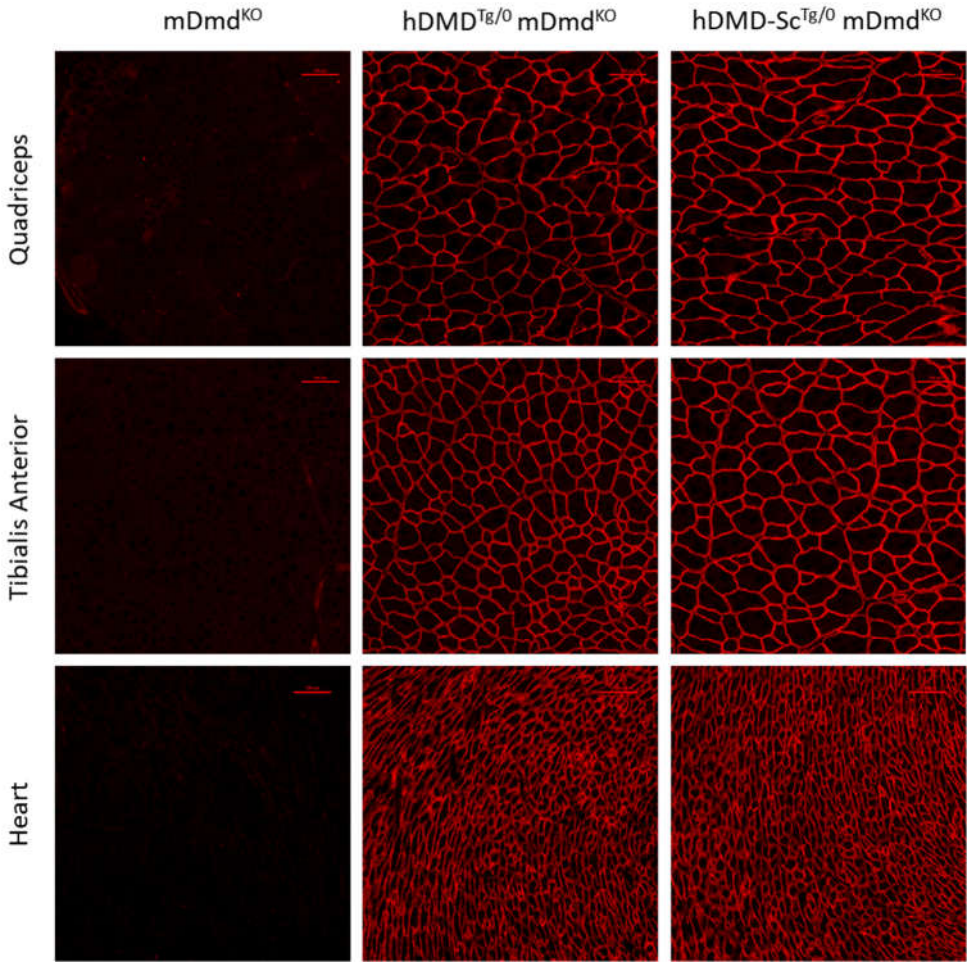
558

559 Figure 7



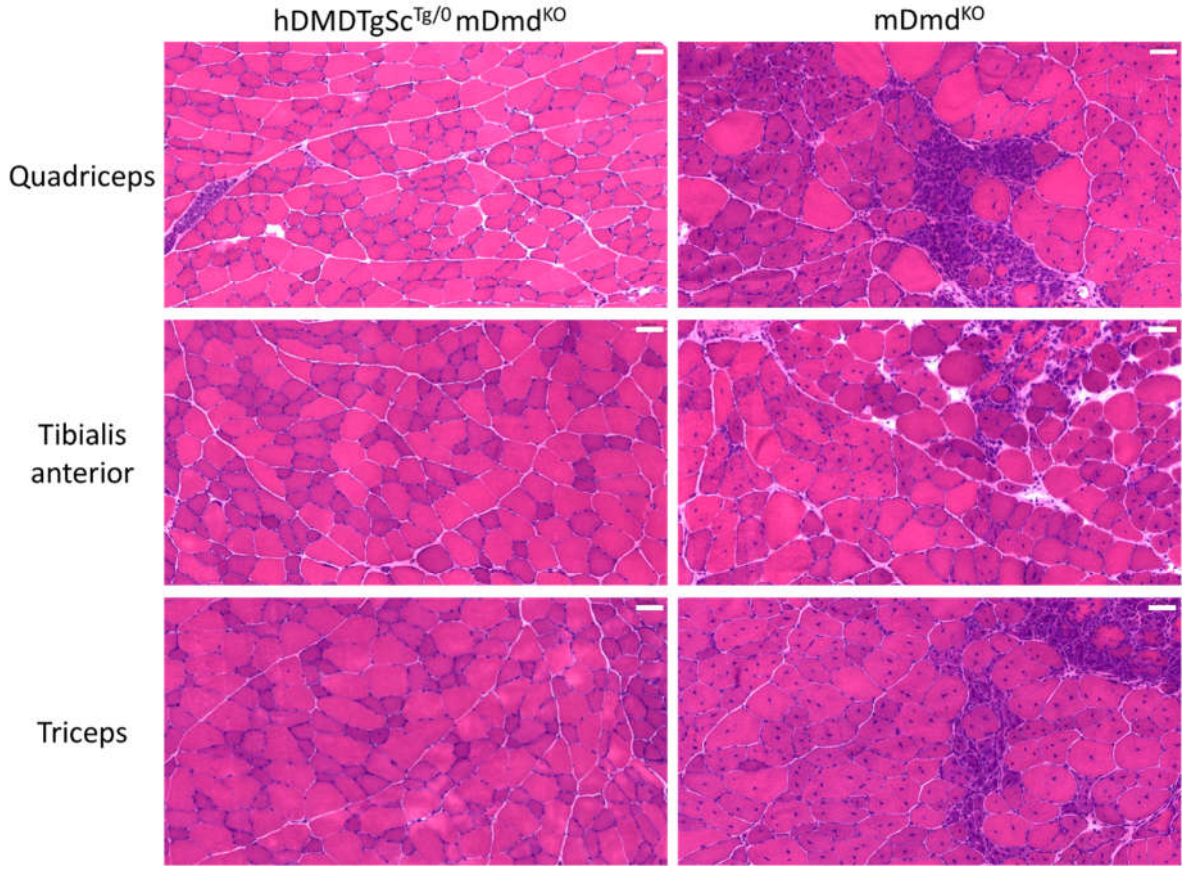
560

561 Figure 8


mDmd^{KO}
hDMD$^{Tg/0}$ mDmd^{KO}
hDMD-Sc$^{Tg/0}$ mDmd^{KO}
Quadriceps
Tibialis Anterior
Heart

562

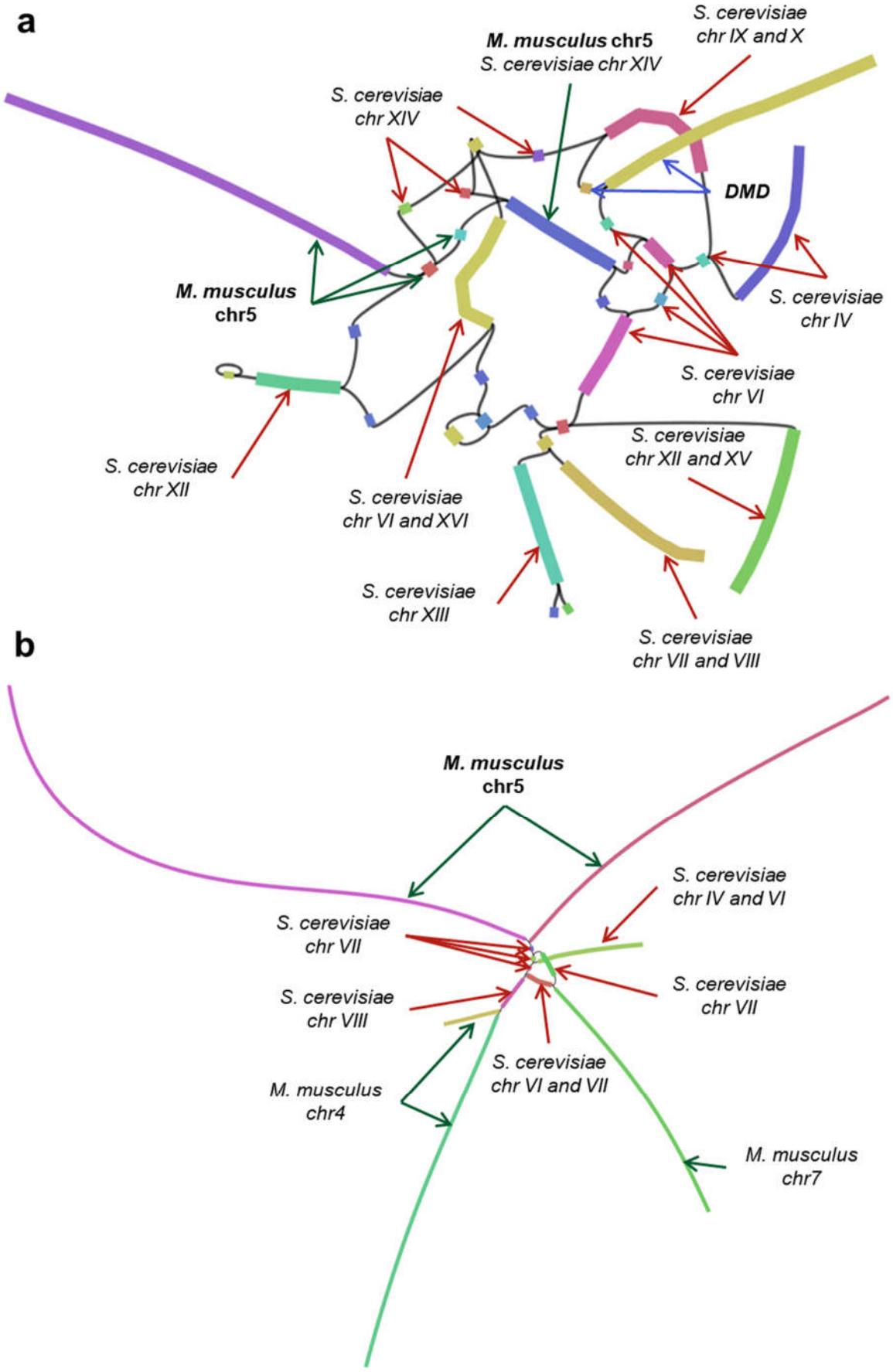
563 Figure 9



564

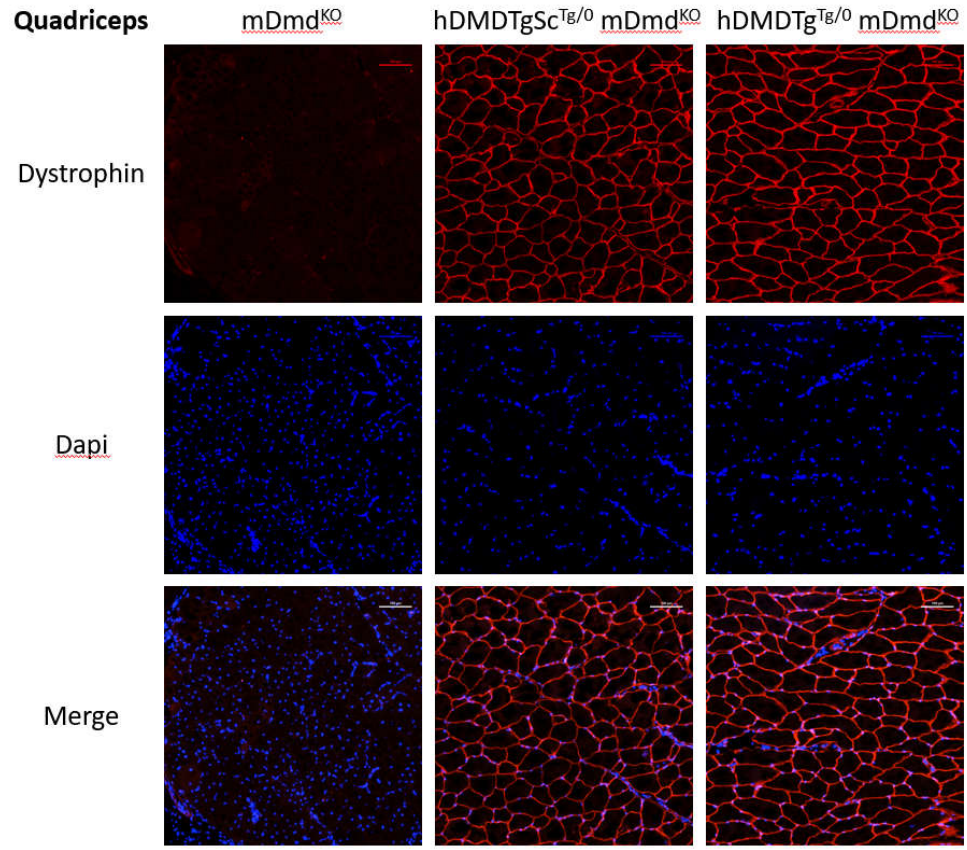
565

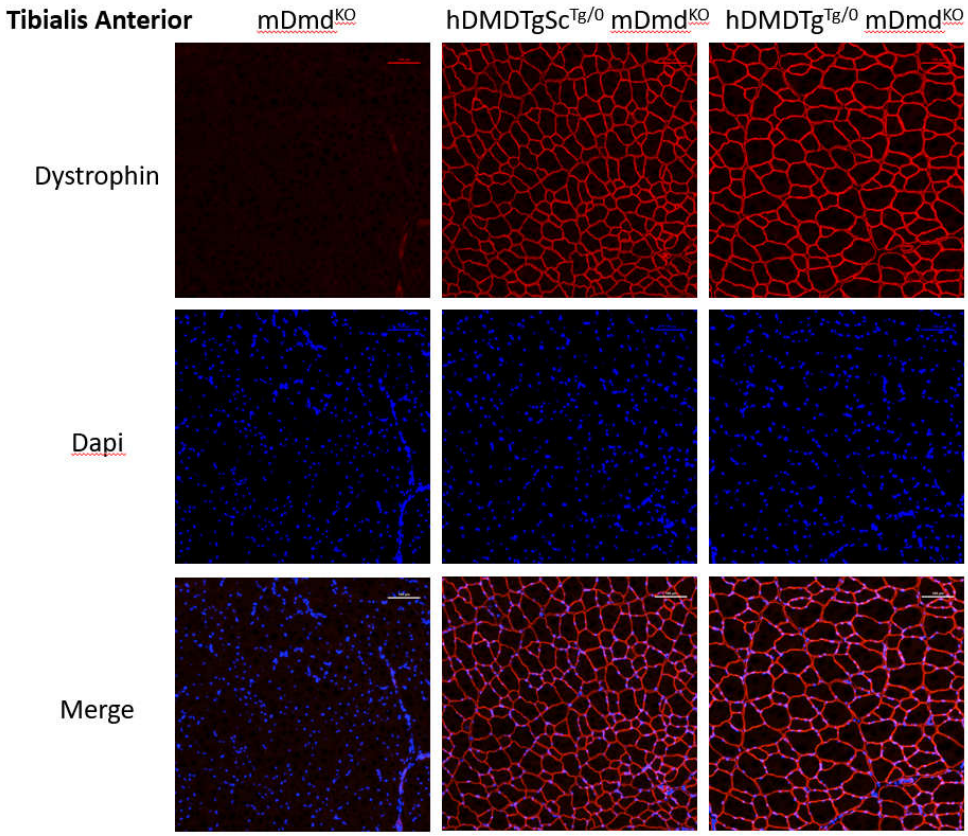
566 Supplementary figure 1

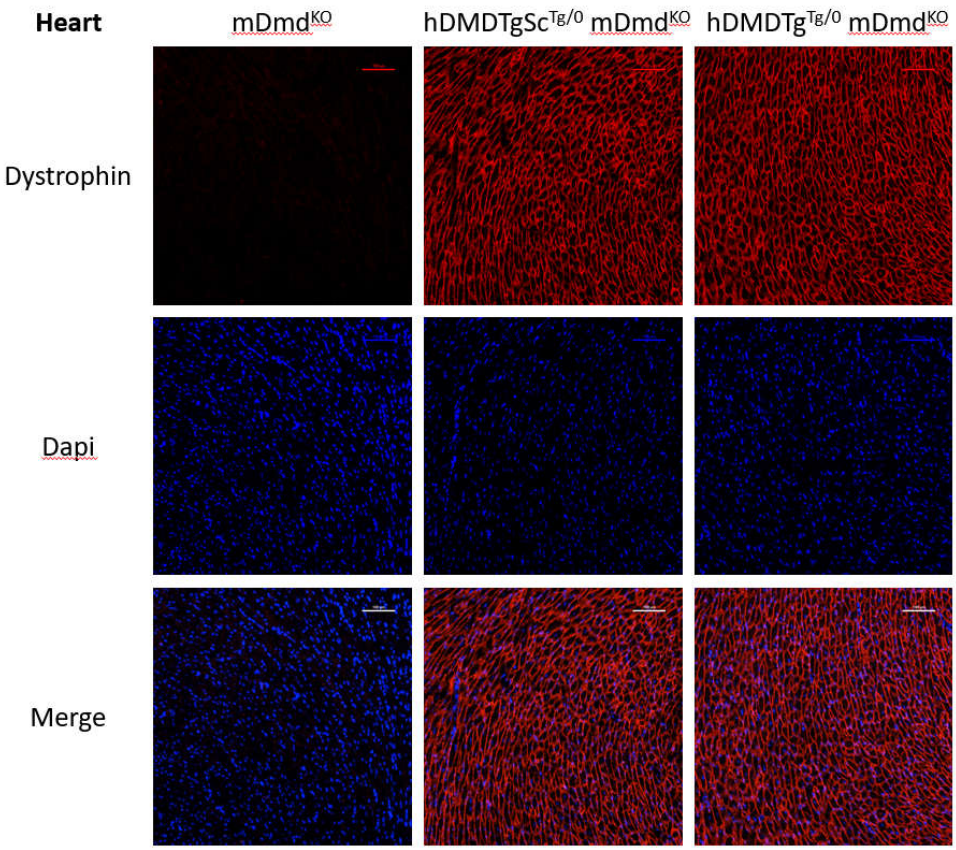


567

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.

585 6. Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, Dawkins H, Lamont L,
586 Roy AJ, Chamova T, et al: **The TREAT-NMD DMD Global Database: Analysis of More than**
587 **7,000 Duchenne Muscular Dystrophy Mutations.** *Human Mutation* 2015, **36:**395-402.
588 7. Aartsma-Rus A, Van Putten M: **The use of genetically humanized animal models for**
589 **personalized medicine approaches.** *Disease Models & Mechanisms* 2020, **13:**dmm041673.
590 8. t Hoen PA, de Meijer EJ, Boer JM, Vossen RH, Turk R, Maatman RG, Davies KE, van Ommen
591 GJ, van Deutekom JC, den Dunnen JT: **Generation and characterization of transgenic mice**
592 **with the full-length human DMD gene.** *J Biol Chem* 2008, **283:**5899-5907.
593 9. Veltrop M, Van Vliet L, Hulsker M, Claassens J, Brouwers C, Breukel C, Van Der Kaa J, Linssen
594 MM, Den Dunnen JT, Verbeek S, et al: **A dystrophic Duchenne mouse model for testing**
595 **human antisense oligonucleotides.** *PLOS ONE* 2018, **13:**e0193289.
596 10. Pickar-Oliver A, Gough V, Bohning JD, Liu S, Robinson-Hamm JN, Daniels H, Majoros WH,
597 Devlin G, Asokan A, Gersbach CA: **Full-length dystrophin restoration via targeted exon**
598 **integration by AAV-CRISPR in a humanized mouse model of Duchenne muscular dystrophy.**
599 *Mol Ther* 2021, **29:**3243-3257.
600 11. Young CS, Mokhonova E, Quinonez M, Pyle AD, Spencer MJ: **Creation of a Novel Humanized**
601 **Dystrophic Mouse Model of Duchenne Muscular Dystrophy and Application of a**
602 **CRISPR/Cas9 Gene Editing Therapy.** *J Neuromuscul Dis* 2017, **4:**139-145.
603 12. Hiramuki Y, Abe S, Uno N, Kazuki K, Takata S, Miyamoto H, Takayama H, Morimoto K,
604 Takehara S, Osaki M, et al: **Full-length human dystrophin on human artificial chromosome**
605 **compensates for mouse dystrophin deficiency in a Duchenne muscular dystrophy mouse**
606 **model.** *Scientific Reports* 2023, **13:**4360.
607 13. Hoshiya H, Kazuki Y, Abe S, Takiguchi M, Kajitani N, Watanabe Y, Yoshino T, Shirayoshi Y, Higaki
608 K, Messina G, et al: **A highly stable and nonintegrated human artificial chromosome (HAC)**
609 **containing the 2.4 Mb entire human dystrophin gene.** *Mol Ther* 2009, **17:**309-317.

610 14. Yavas A, Weij R, van Putten M, Kourkouta E, Beekman C, Puoliväli J, Bragge T, Ahtoniemi T,
611 Knijnenburg J, Hoogenboom ME, et al: **Detailed genetic and functional analysis of the**
612 **hDMDdel52/mdx mouse model.** *PLoS One* 2020, **15:**e0244215.
613 15. van Deutekom J, Beekman C, Bijl S, Bosgra S, van den Eijnde R, Franken D, Groenendaal B,
614 Harquouli B, Janson A, Koevoets P, et al: **Next Generation Exon 51 Skipping Antisense**
615 **Oligonucleotides for Duchenne Muscular Dystrophy.** *Nucleic Acid Ther* 2023, **33:**193-208.
616 16. Chey YCJ, Arudkumar J, Aartsma-Rus A, Adikusuma F, Thomas PQ: **CRISPR applications for**
617 **Duchenne muscular dystrophy: From animal models to potential therapies.** *WIREs*
618 *Mechanisms of Disease* 2023, **15**.
619 17. Mizuno S, Takami K, Daitoku Y, Tanimoto Y, Dinh TTH, Mizuno-Iijima S, Hasegawa Y, Takahashi
620 S, Sugiyama F, Yagami K-i: **Peri-implantation lethality in mice carrying megabase-scale**
621 **deletion on 5qc3.3 is caused by Exoc1 null mutation.** *Scientific Reports* 2015, **5:**13632.
622 18. Kato T, Hara S, Goto Y, Ogawa Y, Okayasu H, Kubota S, Tamano M, Terao M, Takada S: **Creation**
623 **of mutant mice with megabase-sized deletions containing custom-designed breakpoints by**
624 **means of the CRISPR/Cas9 system.** *Scientific Reports* 2017, **7:**59.
625 19. Korablev AN, Serova IA, Serov OL: **Generation of megabase-scale deletions, inversions and**
626 **duplications involving the Contactin-6 gene in mice by CRISPR/Cas9 technology.** *BMC*
627 *Genetics* 2017, **18:**112.
628 20. Boroviak K, Doe B, Banerjee R, Yang F, Bradley A: **Chromosome engineering in zygotes with**
629 **CRISPR/Cas9.** *genesis* 2016, **54:**78-85.
630 21. Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, Taylor DJ, Shafin K, Shumate A,
631 Xiao C, et al: **A complete reference genome improves analysis of human genetic variation.**
632 *Science* 2022, **376:**eabl3533.
633 22. Zhu F, Nair RR, Fisher EMC, Cunningham TJ: **Humanising the mouse genome piece by piece.**
634 *Nat Commun* 2019, **10:**1845.

635 23. Zhang Y, Li H, Nishiyama T, McAnally JR, Sanchez-Ortiz E, Huang J, Mammen PPA, Bassel-Duby
636 R, Olson EN: **A humanized knockin mouse model of Duchenne muscular dystrophy and its**
637 **correction by CRISPR-Cas9 therapeutic gene editing.** *Mol Ther Nucleic Acids* 2022, **29:**525-
638 537.
639 24. Yucel N, Chang AC, Day JW, Rosenthal N, Blau HM: **Humanizing the mdx mouse model of**
640 **DMD: the long and the short of it.** *NPJ Regen Med* 2018, **3:**4.
641 25. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F: **Genome engineering using the**
642 **CRISPR-Cas9 system.** *Nature Protocols* 2013, **8:**2281-2308.
643 26. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.**
644 *Bioinformatics* 2009, **25:**1754-1760.
645 27. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.**
646 *arXiv pre-print server* 2013.
647 28. Taylor SS, Larin Z, Tyler-Smith C: **Analysis of extrachromosomal structures containing human**
648 **centromeric alphoid satellite DNA sequences in mouse cells.** *Chromosoma* 1996, **105:**70-81.
649 29. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P: **Sambamba: fast processing of NGS**
650 **alignment formats.** *Bioinformatics* 2015, **31:**2032-2034.
651 30. Cameron DL, Baber J, Shale C, Valle-Inclan JE, Besselink N, van Hoeck A, Janssen R, Cuppen E,
652 Priestley P, Papenfuss AT: **GRIDSS2: comprehensive characterisation of somatic structural**
653 **variation using single breakend variants and structural variant phasing.** *Genome Biol* 2021,
654 **22:**202.
655 31. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC:
656 **Accurate detection of complex structural variations using single-molecule sequencing.** *Nat*
657 *Methods* 2018, **15:**461-468.
658 32. Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K,
659 Maurer N, Koren S, et al: **Nanopore sequencing and the Shasta toolkit enable efficient de**
660 **novo assembly of eleven human genomes.** *Nat Biotechnol* 2020, **38:**1044-1053.

661 33. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC,
662 Soyk S: **Automated assembly scaffolding using RagTag elevates a new tomato system for**
663 **high-throughput genome editing.** *Genome Biol* 2022, **23:**258.
664 34. Aartsma-Rus A, van Putten M: **Assessing functional performance in the mdx mouse model.** *J*
665 *Vis Exp* 2014.
666 35. Wang C, Yue F, Kuang S: **Muscle Histology Characterization Using H&E Staining and Muscle**
667 **Fiber Type Classification Using Immunofluorescence Staining.** *Bio Protoc* 2017, **7**.

668