Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells

Minchul Kim[1,4], Vedran Franke[2,4], Bettina Brandt[1], Simone Spuler[3], Altuna Akalin[2,*] and Carmen Birchmeier[1,*]

[1] Developmental Biology/Signal Transduction, Max Delbrueck Center for Molecular Medicine, Berlin, Germany

[2] Bioinformatics, Berlin Institute for Medical Science Biology, Max Delbrueck Center for Molecular Medicine, Berlin, Germany

[3] Muscle Research Unit, Experimental and Clinical Research Center, Charité Medical Faculty and Max Delbrueck Center for Molecular Medicine, Berlin, Germany

[4] These authors contributed equally to this work

[*] Co-correspondending authors: cbirch@mdc-berlin.de; altuna.akalin@mdc-berlin.de

Abstract

All cells must compartmentalize their intracellular space in order to properly function[1]. Syncytial cells face an additional challenge to this fundamental problem, which is to coordinate gene expression programs among many nuclei. Using the skeletal muscle fiber as a model, we systematically investigated the functional heterogeneity among nuclei inside a syncytium. We performed single nucleus RNA sequencing (snRNAseq) of isolated myonuclei from uninjured and regenerating muscle, which revealed a remarkable and dynamic heterogeneity. We identified distinct nuclear subtypes unrelated to fiber type diversity, completely novel subtypes as well as the expected neuromuscular[2-5] and myotendinous[6-8] junction subtypes. snRNAseq of fibers from the *Mdx* dystrophy mouse model uncovered additional nuclear populations. Specifically, we identified the molecular signature of degenerating fibers and found a nuclear population that expressed genes implicated in myofiber repair. We confirmed the existence of this population in patients with muscular dystrophy. Finally, modifications of our approach revealed an astonishing compartmentalization inside the rare and specialized muscle spindle fibers. In summary, our work shows how regional transcription shapes the architecture of multinucleated syncytial muscle cells, and provides an unprecedented roadmap to molecularly dissect distinct compartments of the muscle.

42 Multinucleated skeletal muscle cells possess functionally distinct nuclear compartments to meet
43 the needs for specialized gene expression in different parts of the large syncytium. The best
44 documented nuclear subtype locates at the neuromuscular junction (NMJ) where the nerve
45 instructs myonuclei to express genes involved in synaptic transmission[2-5]. Another specialized
46 compartment is found at the myotendinous junction (MTJ) where the myofibers attach to the
47 tendon. It has been shown that specialized cell adhesion and cytoskeletal proteins are enriched at
48 the MTJ to allow force transmission[6-8]. However, whether the corresponding transcripts are
49 specifically produced from MTJ myonuclei is often unclear, especially in mammals. More
50 generally, a systematic analysis of the heterogeneity among myonuclei is lacking, and we do not
51 know what kinds of nuclear subtypes exist. Such knowledge will provide important insights into
52 how myofibers can orchestrate their many functions.

53 Previous studies on gene expression in the muscle relied on the analysis of selected candidates
54 by *in situ* hybridization or on profiling the entire muscle tissue. The former is difficult to scale up,
55 whereas the latter averages the transcriptomes of all nuclei. More recently, several studies have
56 used single cell approaches to reveal the cellular composition of the entire muscle tissue[9-11].
57 However, these approaches did not sample nuclei in the myofiber and thus did not investigate
58 myonuclear transcriptomes. Single-nucleus RNA-Seq (snRNAseq) using cultured human
59 myoblasts failed to detect transcriptional heterogeneity among myotube nuclei[12], underscoring
60 the importance of studying the heterogeneity in an *in vivo* context where myofibers interact with
61 surrounding cell types.

62

63 **Single-nucleus RNA-Seq analysis of uninjured and regenerating muscles**

64 We genetically labeled mouse myonuclei by crossing a myofiber-specific Cre driver (*HSA-Cre*)
65 with a Cre-dependent H2B-GFP reporter. H2B-GFP is deposited at the chromatin, which allows
66 us to isolate single myonuclei using flow cytometry. Nuclei of regenerating fibers were also
67 efficiently labeled 7 days after cardiotoxin-induced injury (7 days post injury; 7 d.p.i.) (Extended
68 Data Fig. 1a). We confirmed the efficiency and specificity of the labeling (Extended Data Fig. 1a
69 and 1b). H2B-GFP was absent in endothelia (Cd31+), Schwann cells (Egr2+), tissue resident
70 macrophages (F4/80+) and muscle stem cells (Pax7+) (Extended Data Fig. 1c).

71 We next established a protocol for rapid isolation of intact myonuclei. Conventional methods
72 involve enzymatic dissociation of muscle fibers at 37°C, which causes secondary changes in
73 gene expression[13,14]. In contrast, our procedure uses mechanical disruption on ice and takes only
74 20 minutes from dissection to flow cytometry. Indeed, we did not detect nuclear populations
75 expressing stress-induced genes in the subsequent analysis (data not shown).

76 For snRNAseq profiling, we used the CEL-Seq2 technology[15], which has superior gene detection
77 sensitivity over other existing methods[16]. We used a stringent filter and considered only those
78 genes that were detected in at least 5 nuclei for further analysis. As a result, we detected 1000-
79 2000 genes per nucleus and ~5000 genes per cluster. Throughout our study, we extensively
80 tested multiple algorithms and parameters, the choice of which did not affect any of the
81 biological findings. We analyzed nuclei from uninjured (476 nuclei) and regenerating *tibialis*
82 *anterior* (TA) muscle (7 and 14 d.p.i., 958 and 1,675 nuclei, respectively). T-distributed
83 stochastic neighbor embedding (t-SNE) projection of these datasets revealed heterogeneity
84 among myonuclei (Fig. 1a). All nuclei expressed high levels of *Ttn*, a pan-muscle marker (Fig.
85 1b). The TA muscle contains three different fiber types (IIA - intermediate, IIB - very fast and
86 IIX - fast) that express distinct myosin genes. The major population (cluster 1) could be sub-
87 divided into nuclei from distinct fiber types (Fig. 1b); *Myh1*- (IIX) or *Myh4* (IIB)-positive nuclei
88 were most abundant and present roughly in a ratio of 1:1. *Myh2* (IIA)-expressing nuclei
89 represented a minor population, consistent with the reported proportion of fiber types[17].

90 To understand the molecular processes associated with regeneration, we selected top ~50 genes
91 that distinguish each time point and performed gene ontology (GO) term analysis. Genes
92 expressed specifically in uninjured muscle were generally associated with metabolism (e.g.
93 *Gdap1, Gpx3* and *Smox*) and negative regulation of signal transduction (e.g. *Igfbp5*). Genes
94 expressed at 7 d.p.i. were associated with processes related to Pol II transcription (e.g. Med16),
95 suggesting that fiber growth is occurring. Indeed, genes known to induce fiber growth such as
96 Igf1 and Mettl21c were up-regulated at 7 d.p.i[18,19]. Lastly, genes expressed at 14 d.p.i. were
97 enriched for terms like muscle contraction and sarcomere organization (e.g. *Actn2, Lmod2* and
98 *Dmd*), suggesting that normal muscle function is being recovered at late regeneration stage.

99 In addition to the major population, we detected small nuclear populations with very distinct
100 transcriptomes (clusters 2-8, Fig. 1c). Different myosin genes were evenly distributed in these,
101 indicating that the heterogeneity is not caused by fiber type differences (Fig. 1b). We first
102 searched for and identified a cluster specifically expressing known NMJ marker genes such as
103 *Prkar1a, Chrne, Chrna*1, *Ache*, *Colq*, *Utrn* and *Lrp4* (cluster 4; Fig. 1d and 1e)[20]. Further, our
104 data identified other genes not previously known to be specifically expressed at the NMJ such as
105 *Ufsp1*, *Vav3* and *Ablim2.* (Fig. 1e). Double fluorescence *in situ* hybridization (FISH) of newly
106 identified markers and *Prkar1a* confirmed their specific expression at the NMJ (Fig. 1f).

107

108 **Two distinct nuclear populations at the myotendinous junction**

109 We found two nuclear populations that expressed MTJ-related genes, and designated them MTJ-
110 A and MTJ-B (Fig. 2a and 2b). MTJ-A expressed genes whose protein products are known to be
111 enriched at the MTJ (e.g. *Itgb1*)[21] as well as specific collagens (e.g. *Col22a1* and *Col24a*).
112 *Col22a1* has been functionally characterized in zebrafish using morpholino knockdown causing
113 a disruption of MTJ development[22]. MTJ-B nuclei expressed an alternative set of collagens that
114 are known to be deposited at the MTJ such as *Col6a1*, *Col6a3* and *Col1a2*[23]. *Col6a1* expression
115 was particularly notable because its mutation causes Bethlem myopathy, which is characterized
116 by deficits at the MTJ[24].

117 We validated the two top marker genes of MTJ-A (*Tigd4* and *Col22a1*) using single molecule
118 FISH (smFISH), and observed their expression in nuclei at the fiber endings (Fig. 2c and
119 Extended Data Fig. 2). These transcripts were exclusively expressed from H2B-GFP positive
120 myonuclei and present only at the MTJ. Their expression became much more pronounced at 14
121 d.p.i. compared to uninjured muscle, indicating that at this stage of regeneration the MTJ is re-
122 forming (Fig. 2c). To visualize heterogeneity within the syncytium, we isolated single fibers and
123 performed double FISH (Fig. 2d). As expected, *Tigd4* FISH signals were detected at the fiber
124 ends, whereas *Ufsp1* transcripts appeared at the middle of the fiber where the NMJ is located.

125 We detected markers of MTJ-B (*Pdgfrb* and *Col6a3* transcripts) expressed from H2B-GFP
126 nuclei at the MTJ in both uninjured muscle and at 14 d.p.i. (Extended Data Fig. 3). Unlike the
127 genes of MTJ-A, these transcripts were also found in H2B-GFP negative nuclei located outside

128 the fiber and distally to the MTJ (Extended Data Fig. 3). These signals likely originate from
129 interstitial fibroblasts or muscle stem cells[9,10,25]. Unlike MTJ-A, MTJ-B nuclei were not found in
130 every fiber. In addition, we confirmed that Ebf1 protein, another marker of MTJ-B, is present in
131 H2B-GFP positive myonuclei (Fig. 2c).

132

133 **Identification of unprecedented nuclear populations**

134 Other clusters represent novel myonuclear compartments. The top markers of cluster 5 were
135 maternally imprinted lncRNAs that are located at one genomic locus (also known as Dlk1-Dio3
136 locus) (Fig. 3a). This locus encodes a very large number of microRNAs expressed from the
137 maternal allele, among them microRNAs known to target transcripts of mitochondrial proteins
138 encoded by the nucleus[26]. smFISH against one of the lncRNAs (*Rian*) showed a clear and strong
139 expression in a subset of myonuclei (Fig. 3b), which was observed regardless of the animal's sex
140 (data not shown). smFISH on isolated fibers showed dispersed localization of *Rian* expressing
141 nuclei without clear positional preference (Fig. 3c). A previous study reported that the Dlk1-
142 Dio3 locus becomes inactive during myogenic differentiation[27]. However, our results show that
143 some myonuclei retain the expression, which might be important to shape the metabolisms of the
144 fiber.

145 Intriguingly, cluster 2 contained many genes that function in endoplasmic reticulum (ER)-
146 associated protein translation and trafficking (Fig. 3a). Tmem170a induces formation of ER
147 sheets, which is associated with active protein translation[28], and Rab40b is known to localize to
148 the Golgi/endosome and regulate trafficking[29]. Furthermore, srpRNA (signal recognition particle
149 RNA), an integral component of ER-bound ribosomes[30], was markedly enriched in this
150 population (Fig. 3a). The enrichment of srpRNA was also observed when expression of repeat
151 elements was quantified (Extended Data Fig. 4). smFISH against *Gssos2*, one of the marker
152 transcripts, showed a heterogeneous and strong expression in a subset of myonuclei (Fig. 3b).

153 Re-clustering of the major population identified in Figure 1a revealed additional nuclear
154 subpopulations with distinct transcriptomes (clusters 1a and 1b; Fig. 3d, 3e and Extended Data 5).
155 We verified cluster 1a using smFISH against the top marker gene *Muc13*. Myonuclei expressing
156 *Muc13* were always located at the very outer part of the muscle tissue near the perimysium (Fig.

157 3f and 3g). One ultrastructural study using bovine muscle suggested that myofibers and
158 perimysium establish specialized adhesion structures[31], and we suggest that we have detected the
159 myonuclear compartment participating in this process.

160

161 **snRNAseq of fibers in *Mdx* dystrophy model**

162 To begin to understand whether and how myonuclei heterogeneity is altered in muscle disease,
163 we conducted snRNAseq on *Mdx* fibers (1,939 nuclei), a mouse model of Duchenne muscular
164 dystrophy (DMD) (Fig. 4a). DMD is a degenerative muscle disease caused by mutations in the
165 *dystrophin* gene, which is also mutated in *Mdx* mice. To examine how the *Mdx* transcriptome is
166 related to those of uninjured and regenerating muscle, we calculated gene signature scores of
167 each nucleus based on the top 50 genes that distinguish different time points after injury. This
168 showed that the overall gene signature of *Mdx* resembles nuclei of the regenerating muscle (14
169 d.p.i.) (Fig. 4b).

170 We next tried to identify NMJ and MTJ populations in *Mdx* fibers. However, nuclei with strong
171 NMJ signature could not be identified. Consistently, double smFISH against two NMJ markers
172 (*Ufsp1* and *Prkar1a1*) showed that the strict co-localization seen in control muscle was lost, but
173 these genes were instead expressed in a dispersed manner in *Mdx* muscles (Extended Data Fig. 6).
174 Notably, the histological structure of the NMJ is known to be fragmented in *Mdx* mice[32,33], and
175 our data suggest that also postsynaptic nuclei are incompletely specified. The histology of the
176 MTJ was reported to be disrupted in patients and mice with *Dystrophin* mutations[34-36].
177 Interestingly, we did not find nuclei with an MTJ-B signature in *Mdx* mice, whereas MTJ-A
178 nuclei were present but did not cluster in the t-SNE map (Extended Data Fig. 6). Nevertheless,
179 marker genes of MTJ-A (*Tigd4* and *Col22a1*) robustly labeled the MTJ of *Mdx* muscle. We
180 interpret that fiber-level heterogeneity (e.g. dying fibers, newly formed fibers, fibers in various
181 stages of maturity, repairing fibers) strongly drives the shape of the t-SNE map in *Mdx*,
182 interfering with the clustering of MTJ-A nuclei.

183

184 **Emergence of new nuclear populations in *Mdx* dystrophy model**

185 Several populations newly emerged in the *Mdx* muscle (especially clusters 1, 3 and 5), and we
186 further characterized two of these. Cluster 3 was highly enriched in various non-coding
187 transcripts (Fig. 4c and Extended Data 7a) and further experiments demonstrated that these
188 transcripts were located in dying fibers. In particular, staining of mouse IgG that identifies
189 degenerating fibers with leaky membranes demonstrated the expression of these ncRNAs in
190 IgG+ fibers (Fig. 4c and Extended Data Fig. 7b). Further, fibers strongly expressing these
191 ncRNAs were highly infiltrated with H2B-GFP negative cells likely corresponding to
192 macrophages (Extended Data Fig. 7c). Whether the ncRNAs are the consequence or active
193 contributors to fiber death needs further investigation.

194 Marker genes of cluster 5 showed high enrichment of ontology terms related to human muscle
195 disease (Fig. 4d). Indeed, many top marker genes were previously reported to be mutated in
196 human myopathies (*Klhl40*, *Flnc* and *Fhl1*)[37-39] or to directly interact with proteins whose
197 mutation causes disease (Ahnak interacts with Dysferlin, and Hsp7b or Xirp1 with Flnc)[40-42].
198 Combinatorial smFISH in tissue sections confirmed their co-expression in a subset of nuclei of
199 *Mdx* muscle, but such co-expressing nuclei were not present in control muscle (Fig. 4e, Extended
200 Data Fig. 8a and 8b). Further, nuclei expressing these genes were frequently closely spaced. We
201 also verified that *FLNC* and *XIRP1* were co-expressed in muscle tissues of Becker's patients, a
202 muscular dystrophy caused by *DYSTROPHIN* mutation, but not in healthy human muscle (Fig.
203 4f and Extended Data Fig. 8a).

204 Previous studies have established that Flnc and Xirp1 proteins localize to sites of myofibrillar
205 damage to repair such insults[41,43,44], whereas Dysferlin, an interaction partner of Ahnak,
206 functions during repair of muscle membrane damage[45]. Our analysis shows that these genes are
207 transcriptionally co-regulated, and suggests that this occurs in response to micro-damage. To
208 substantiate that this signature is not specific to muscular dystrophy caused by *Dystrophin*
209 mutation, we investigated whether these nuclei are also present in *Dysferlin* deficient muscle
210 where the continuous micro-damage to the membrane is no longer efficiently repaired[45]. Again,
211 we observed nuclei co-expressing *Flnc* and *Xirp1* in this mouse disease model and in biopsies
212 from human patients (Extended Data Fig. 8c and 8d). We propose that this cluster represents a
213 'repair signature'.

214

215 **Nuclear heterogeneity in muscle spindle fibers**

216 In principle, our approach can be used to explore nuclear heterogeneity in specific fiber types.
217 We thus aimed to investigate heterogeneity in muscle spindles that detect muscle stretch and
218 function in motor coordination[46]. Muscle spindles are extremely rare and ~10 spindles exist in a
219 TA muscle of the mouse[47]. They contain bag and chain fibers, and their histology forecasts
220 further compartmentalization (Fig. 5a). *HSA-Cre* labels myonuclei of the spindle (Fig. 5b), but
221 the overwhelming number of nuclei derive from extrafusal fibers. To overcome this, we used
222 *Calb1-Cre* to specifically isolate spindle myofiber nuclei (Fig. 5b) and discovered a pronounced
223 transcriptional heterogeneity (Fig. 5c).

224 Bag fibers are slow fibers and express *Myh7b*[48,49], whereas chain fibers are fast. A cluster
225 expressing *Myh7b* and *Tnnt1*, a slow type troponin isoform, was assigned to mark Bag fibers. In
226 contrast, two clusters expressed *Myh13*, a fast type Myosin, or *Tnnt3*, a fast type troponin, which
227 we named Chain1 and Chain2. Strikingly, we identified a cluster that expressed overlapping but
228 not identical set of genes as those identified in NMJ nuclei of extrafusal fibers, e.g. *Chrne*, *Ufsp1*
229 and *Ache,* which we assign as the NMJ of the spindle (spdNMJ) (Fig. 5d and 5e). Furthermore,
230 the spindle myotendinous nuclei (spdMTJ) expressed an overlapping but not identical set of
231 genes as those identified in MTJ-B nuclei of extrafusal fibers (Fig. 5d and 5e). MTJ-A markers
232 were not detected. Notably, the clusters Bag and spdNMJ expressed the mechanosensory channel
233 *Piezo2*[50]. To verify the assignment and to define the identity of an additional large compartment
234 (labeled as Sens), we validated the expression of different marker genes in H2B-GFP positive
235 fibers of *Calb1-Cre* muscle in tissue sections (Extended Data Fig. 9) and after manual isolation
236 of fibers under a fluorescent dissecting microscope (Fig. 5f). smFISH of *Calcrl*, a marker of the
237 Sens cluster, showed specific localization in the central part of spindle fibers containing densely
238 packed nuclei, the site where sensory neurons innervate (Fig. 5f). In the same fiber, transcripts of
239 the spdNMJ marker gene *Ufsp1* located laterally as distinct foci. In contrast, *Piezo2* was
240 expressed throughout the lateral contractile part of the fiber, but was excluded from the central
241 portion. Thus, the central non-contractile part of the muscle spindle that is contacted by sensory
242 neurons represents a distinct fiber compartment with specialized myonuclei clearly
243 distinguishable from the spdNMJ.

244

245 **Profiling transcriptional regulators across distinct compartments**

246 To gain insights into the transcriptional control of the distinct nuclear compartments, we
247 investigated the expression profile of transcription factors and epigenetic regulators (Extended
248 Fig. 10a). Notably, Etv5 (also known as Erm), a transcription factor known to induce the NMJ
249 transcriptome[51], was identified as one of the top factor in NMJ nuclei. To begin to assess the
250 functionality of new candidates, we chose Ebf1, the most strongly enriched transcription
251 activator in MTJ-B. ChIP-Seq data of Ebf1 (ENCODE project) showed that Ebf1 directly binds
252 to about ~70% of the genes we identified as MTJ-B markers (Extended Fig. 10b). We generated
253 a C2C12 cell line in which Ebf1 expression is induced by doxycycline (Extended Fig. 10c). RT-
254 qPCR analysis of selected markers showed that many of them were induced in a dose-dependent
255 manner upon Ebf1 over-expression (Extended Fig. 10d). We also verified that ColVI protein was
256 induced by Ebf1 over-expression (Extended Fig. 10c). To validate these findings *in vivo*, we
257 analyzed *Ebf1* mutant muscle. Using smFISH for *Ttn* to identify myonuclei, we observed a
258 strong reduction of *Col6a3* and *Fst1l* transcripts in the *Ebf1* mutant compared to control MTJ
259 myonuclei (Extended Fig. 10e). Their expression from non-myonuclei that also express Ebf1 was
260 also diminished. Thus, our dataset provides a template for the identification of regulatory factors
261 that establish or maintain these compartments.

262

263 **Discussion**

264 Here, we used single-nucleus sequencing to define the transcriptome of myofiber nuclei, which
265 identified nuclear populations at anatomically distinct locations such as NMJ, MTJ and an
266 extremely rare nuclear population at the perimysium. Intriguingly, we found two distinct
267 populations at the MTJ in the uninjured muscle, but only one of these were present in *Mdx* and
268 spindle fibers. Moreover, we identified two new nuclear populations that are present throughout
269 the fiber. Thus, myonuclear populations are not always associated with distinctive anatomical
270 features. How these nuclear subtypes emerge and are regulated, e.g. whether they are associated
271 with other cell types in the muscle tissue, induced by certain local signals or arise stochastically,
272 needs further study. By investigating the *Mdx* dystrophy model, we identified the molecular

signature of degenerating fibers and a transcriptional program that appears to be associated with fiber repair. These newly identified gene signatures might be useful for a quantitative and rapid assessment of muscle damage in the clinic. Lastly, by restricting recombination to muscle spindle fibers, we demonstrated compartmentalization inside these rare fibers, especially the presence of a distinct compartment at the site of innervation by proprioceptive sensory neurons. Taken together, our results show how regional transcription shapes the architecture of multinucleated skeletal muscle cells, reveal the complexity of gene expression regulation in the syncytium, and provide a comprehensive resource for studying distinct muscle compartments.

Given the large size and complexity of the muscle tissue, the full diversity of myonuclei likely needs further exploration. In this aspect, our strategy of genetic labeling to enrich nuclei of interest appears advantageous as demonstrated by our success with the identification of perimysium and muscle spindle myonuclei, which are very rare and therefore hard to detect when the entire tissue is used. Although we concentrated here on the analysis of uninjured and regenerating muscle in health and disease, our strategy should be promising in identifying new nuclear subtypes in other contexts. In the accompanying manuscript, snRNAseq successfully discovered novel nuclear subtypes during early development and in aged muscle (Petrany MJ, Swoboda CO, Sun C, Chetal K, Chen X, Weirauch MT, Salomonis N, Millay DP. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. Submitted to *BioRxiv*). More generally, our approach should be instrumental in investigating other syncytial cell types such as the placental trophoblasts or osteoclasts. In addition to charting and dissecting new nuclear populations, another important question will be to understand the next layers towards establishing specialized compartments such as RNA/protein trafficking.

295 **Methods**

296 **Mouse lines and muscle injury**

297 All experiments were conducted according to regulations established by the Max Delbrück
298 Centre for Molecular Medicine and LAGeSo (Landesamt für Gesundheit und Soziales), Berlin.
299 *HSA-Cre*, *Calb1-IRES-Cre* and *Mdx* mice were obtained from the Jackson laboratory. *Rosa26-*
300 *Lsl-H2B-GFP* reporter line was a kind gift from Martin Goulding (Salk institute) and was
301 described[52]. For experiments regarding uninjured and regenerating muscles, homozygous
302 *Rosa26-LSL-H2B-GFP* mice with heterozygous *HSA-Cre* were used. For *Calb1-IRES-Cre* and
303 *Mdx* experiments, the *H2B-GFP* allele was heterozygous. Nuclei were isolated from muscle of
304 2.5 months old mice. Genotyping was performed as instructed by the Jackson laboratory. For
305 genotyping of the *Rosa26-Lsl-H2B-GFP* reporter, the following primers were used. Rosa4; 5'-
306 TCA ATGGGCGGGGGTCGTT-3', Rosa10; 5'-CTCTGCTGCCTCCTGGCTTCT-3', Rosa11;
307 5'-CGAGGCGGATCACAAGCAATA-3'. Ebf1 mutants[53] and Dysferlin mis-sense mutants[54]
308 were described. To induce muscle injury, 30µl of cardiotoxin (10µM, Latoxan, Porte les Vaence,
309 France) was injected into the tibialis anterior (TA) muscle.

310

311 **Isolation of nuclei from TA muscle**

312 For each sorting, we pooled two TA muscles from two individual mice. Dissected muscle was
313 cut into small pieces, suspended in 1ml hypotonic buffer (250mM sucrose, 10mM KCl, 5mM
314 $MgCl_2$, 10mM Tris-HCl pH 8.0, 25mM HEPES pH 8.0, 0.2mM PMSF and 0.1mM DTT
315 supplemented with protease inhibitor tablet from Roche), and transferred to a 'Tissue
316 homogenizing CKMix' tube (Bertin instruments) containing ceramic beads of mixed size. After
317 incubating on ice for 15 minutes, samples were homogenized with Precellys 24 tissue
318 homogenizer (Bertin instruments) for 20 seconds at 5,000 rpm. Homogenized samples were
319 passed once through 70µm filter (Sysmex), twice through 20µm filter (Sysmex), and once
320 though 5ml filter cap FACS tube (Corning 352235). DAPI (Sigma) was added to final
321 concentration of 300nM to label DNA. GFP and DAPI double positive nuclei were sorted using
322 ARIA Sorter III (BD).

323

324 **Isolation of nuclei from muscle spindle**

325 We pooled all TA muscles from 3-4 mice for each experiment. Two different protocols were
326 employed. The first procedure (protocol 1 in Extended Data Fig. 9a) employed the same
327 procedure as the one used to isolate TA nuclei, pooling up to 4 TA muscles/homogenizer tube,
328 and yielded 96 spindle nuclei.

329 The second procedure (protocol 2 in Extended Data Fig. 9a), we shorten isolation time. Briefly,
330 0.1% Triton X-100 was added to the hypotonic buffer to solubilize the tissue debris, which are
331 otherwise detected as independent particles during FACS. After homogenization and filtration,
332 nuclei were pelleted by centrifuging at 200g for 10 minutes at 4°C. After discarding the
333 supernatants, nuclei were resuspended in 300ul hypotonic buffer and passed through the FACS
334 tube. This yielded 192 spindle nuclei.

335

336 **Library generation and sequencing**

337 96-well plates for sorting were prepared using an automated pipetting system (Integra Viaflo).
338 Each well contained 1.2µl of master mix (13.2µl 10% Triton X-100, 25mM dNTP 22µl, ERCC
339 spike-in 5.5µl and ultrapure water up to 550µl total) and 25ng/µl barcode primers. Plates were
340 stored at -80°C until use.

341 After sorting single nuclei into the wells, plates were centrifuged at 4,000g for 1 minute,
342 incubated on incubated on 65°C for 5 minutes and immediately cooled on ice. Subsequent library
343 generation was performed using the CEL-Seq2 protocol as described by Hashimshony et al[15].
344 After reverse transcription and second strand synthesis, products of one plate were pooled into
345 one tube and cleaned up using AMPure XP beads (Beckman Coulter). After in vitro transcription
346 and fragmentation, aRNA was cleaned up using RNAClean XP beads (Beckman Coulter) and
347 eluted in 7µl ultrapure water. 1ul of aRNA was analyzed on Bioanalyzer RNA pico chip for a
348 quality check. To construct sequencing library, 5µg aRNA was used for reverse transcription
349 (Superscript II, Thermofisher) and library PCR (Phusion DNA polymerase, Thermofisher). After
350 clean up using AMPure XP beads, 1µl sample was ran on Bioanalyzer using a high sensitivity

351 DNA chip to measure size distribution, which we demonstrated the presence of a peak of around
352 400bp length. Sequencing was performed using Illumina HiSeq2500 (for uninjured and
353 regenerating muscle) or HiSeq4000 (for *Mdx* and spindle).

354

355 **Bioinformatics analysis**

356 Single nucleus RNA-Sequencing data was processed using PiGx-scRNAseq pipeline - a
357 derivative of a CellRanger pipeline, but enabling deterministic analysis reproducibility (version
358 0.1.3)[55]. In short, polyA sequences were removed from the reads. The reads were mapped to the
359 genome using STAR[56]. Number of nuclei, for each sample, was determined using dropbead[57].
360 Finally, a combined digital expression matrix was constructed, containing all sequenced
361 experiments.

362 Digital expression matrix post processing was performed using Seurat[58]. The raw data was
363 normalized using the NormalizeData function. The expression of each nucleus was then
364 normalized by multiplying each gene with the following scaling factor: 10000/(total number of
365 raw counts), log(2) transformed, and subsequently scaled. Number of detected genes per nucleus
366 was regressed out during the scaling procedure.

367 Variable genes were defined using the FindVariableGenes function with the default parameters.
368 Samples were processed in three groups with differing parameters. Samples originating from
369 uninjured, 7 d.p.i. and 14 d.p.i. were processed as one group, samples from the *Mdx* mouse as the
370 second group and muscle spindle nuclei as the third group.

371 For samples of the first group, nuclei with less than 500 detected genes were filtered out.
372 Subsequently, genes which were detected at least in 5 nuclei were kept for further analysis. To
373 remove the putative confounding effect between time of sample preparation and biological
374 variable (injury), the processed expression matrices were integrated using the
375 FindIntegrationAnchors function with reciprocal PCA, from the Seurat package. The function
376 uses within batch covariance structure to align multiple datasets.

377 The integration was based on 2000 top variable features, and first 30 principal components. T-
378 SNE was based on the first 15 principal components. Outlier cluster detection was done with
379 dbscan (10.18637/jss.v091.i01.), with the following parameters eps = 3.4, minPts = 20.

380 *Mdx* samples were processed by filtering out all nuclei with less than 100 detected genes. Top
381 100 most variable genes were used for the principal component analysis. t-SNE and Leuvain
382 clustering were based on the first 10 principal components. Resolution parameter of 1 was used
383 for the Leuvain clustering.

384 Spindle cell samples were processed by filtering out all cells with less than 300 detected genes.
385 Top 200 most variable genes were used for the principal component analysis. t-SNE and Leuvain
386 clustering were based on the first 15 principal components. Resolution parameter of 1 was used
387 for the Leuvain clustering.

388 For all datasets, multiple parameter sets were tested during the analysis, and the choice of
389 parameters did not have a strong influence on the results and the derivative biological
390 conclusions. Genes with cluster specific expression were defined using Wilcox test, as
391 implemented in the FindAllMarkers function from the Seurat package. Genes that were detected
392 in at least 25% of the cells in each cluster were selected for differential gene expression analysis.

393 NMJ Nuclei Definition: NMJ nuclei were identified based on the expression of three previously
394 known markers (*Prkar1a*, *Chrne* and *Ache*), using the AddModuleScore function from the Seurat
395 package. All cells with a score greater than 1 were selected as NMJ positive cells. NMJ marker
396 set was expanded by comparing the fold change of gene expression in averaged NMJ positive to
397 NMJ negative cells.

398 MTJ A and B Nuclei Definition: The original gene sets were extracted from cluster specific
399 genes detected in the uninjured, 7 d.p.i. and 14 d.p.i. experiment. Cells were scored as MTJ A/B
400 using the aforementioned gene set, with the AddModuleScore function. All cells which had a
401 respective score greater than 1 were labeled as MTJ A/B positive cells.

402 *Mdx* nuclei scoring by uninjured and regenerating signatures: First, gene signatures specific for
403 each time point were selected using the FindAllMarkers function from the Seurat library, using

404 the default parameters. *Mdx* samples were scored using the top 50 genes per time point with the
405 AUCell method (10.1038/nmeth.4463).

406 Repetitive element annotation was downloaded from the UCSC Browser database
407 (10.1093/nar/gky1095) on 21.01.2020. Pseudo-bulk bigWig tracks were constructed for each
408 defined cluster in uninjured, 7.d.p.i, and 14.d.p.i. The tracks were normalized to the total number
409 of reads. Repetitive element expression was quantified using the ScoreMatrixBin function from
410 the genomation (10.1093/bioinformatics/btu775) package, which calculates the average per-base
411 expression value per repetitive element. The expression was finally summarized to the repetitive
412 element family (class) level by calculating the average expression of all repeats belonging to the
413 corresponding family (class).

414 Transcription factor compendium, used in all analyses was downloaded from AnimalTFDB
415 (https://doi.org/10.1093/nar/gky822). The expression map (Fig. 5d) for spindle fibers was created
416 using the DotPlot from the Seurat package on a selected set of cluster specific genes. Gene
417 ontology analysis was performed using the Enrichr program (10.1093/nar/gkw377).
418 (http://amp.pharm.mssm.edu/Enrichr/).

419

420 **Preparation of tissue sections**

421 Freshly isolated TA muscles were embedded in OCT compound and processed as previously
422 described[59]. Frozen tissue blocks were sectioned to 12-16µm thickness, which were stored at -
423 80°C until future use.

424

425 **Single-molecule FISH (RNAscope)**

426 RNAscope_V2 kit was used according to manufacturer's instructions (ACD/bio-techne). We
427 used Proteinase IV. When combined with antibody staining, after the last washing step of RNA
428 Scope, the slides were blocked with 1% horse serum and 0.25% BSA in PBX followed by
429 primary antibody incubation overnight on 4°C. The subsequent procedures were the same as
430 regular immunohistochemistry. Slides were mounted with Prolonged Antifade mounting solution

431 (Thermofisher). The following probes were used in this study; Ttn (483031), Rian (510531),
432 Vav3 (437431), Col6a3 (552541), Egr1 (423371), Gm10800 (479861) and Calcrl (452281). The
433 following probes were newly designed; Prkar1a (c1), Tigd4 (c1), Muc13 (c1), Gssos2 (c1), Flnc
434 (c1), Klhl40 (c1), Col22a1 (c2), Ufsp1 (c2), Gm10801 (c2), Xirp1 (c2), Fst1l (c2), human Flnc
435 (c2), Ablim2 (c3) and human Xirp1 (c3).

436

437 **Preparation of fluorescent in situ hybridization (FISH) probes**

438 Probes of 500-700bp length spanning exon-exon junction parts were designed using software in
439 NCBI website. Forward and reverse primers included Xho1 restriction site and T3 promoter
440 sequence, respectively. cDNA samples prepared from E13.5-E14.5 whole embryos were used to
441 amplify the target probes using GO Taq DNA polymerase (Promega). PCR products were cloned
442 into pGEM-T Easy vector (Promega) according to manufacturer's guideline, and the identity of
443 the inserts was confirmed by sequencing. 2µg of cloned plasmid DNA was linearized, 500ng
444 DNA was subjected to *in vitro* transcription with T3 polymerase and DIG- or FITC- labeled
445 ribonucleotides (All Roche) for 2 hours at 37°C. Synthesized RNA probes were purified using
446 RNeasy kit (Qiagen). Probes were eluted in 50µl ultrapure water (Sigma), and 50µl formamide
447 was added. We checked the RNA quality and quantity by loading 5µl RNA to 2% agarose gel.
448 Until future use, probes were stored in -80°C. The followings are the annealing sequences of the
449 FISH probes used in this study. Prkar1a For; 5'-CTGCACCACTGAATTACCGTA-3', Prkar1a
450 Rev; 5'-ACAAGATAAGTGTGCCTCGTT-3', Ufsp1 For; 5'-CACTGCCCGTAGCTAGTCC-
451 3', Ufsp1 Rev; 5'-CAGGCCCAAACTTGATCGC-3', Phldb2 For; 5'-
452 GATGAGGAGTCTGTGTTCGAGG-3', Phldb2 Rev; 5'-GCCTCTCTAGCTTTTCCCTCTC-3',
453 Col6a3 For; 5'-GCAGGCTAGTGTGTTCTCGT-3', Col6a3 Rev; 5'-
454 CGTTGTCGGAGCCATCCAAA, Pdgfrb For; 5'-AGTGATACCAGCTTTAGTCCT-3', Pdgrfb
455 Rev; 5'-GTTCACACTCACTGACACGTT-3', Tigd4 For; 5'-
456 CAATGGTTGGCTGGATCGTTTT-3' and Tigd4 Rev; 5'-CTTTGGAGACCTGAATCCTGCT-
457 3'.

458

459 **Fluorescence in situ hybridization (FISH) and immunohistochemistry**

460 Basic procedure for FISH was described before[59] with minor modifications to use fluorescence
461 for final detection. After hybridizing the tissue sections with DIG- labeled probes, washing,
462 RNase digestion and anti-DIG antibody incubation, amplification reaction was carried out using
463 TSA-Rhodamine (1:75 and 0.001% H_2O_2). After washing, slides were mounted with Immu-
464 Mount (Thermo Scientific). When applicable, GFP antibody was added together with anti-DIG
465 antibody.

466 When conducting double FISH, the tissue was hybridized with DIG- and FITC-labeled probes,
467 after detection of the DIG signal, slides were treated with 3% H_2O_2 for 15 minutes and then with
468 4% PFA for one hour at room temperature to eliminate residual peroxidase activity. The second
469 amplification reaction was performed using anti-FITC antibody and TSA-biotin (1:50), which
470 was visualized using Cy5-conjugated anti-streptavidin.

471 Antibodies used for this study were: GFP (Aves labs, 1:500), Col3 (Novus, 1:500), ColIV
472 (Millipore, 1:500), CD31-PE (Biolegend, 1:200), F4/80 (Abcam ab6640, 1:500) and Laminin
473 (Sigma L9393, 1:500). Pax7 (1:50) and Ebf1 (1:20) anti-serum was described[59,60]. Egr2 (1:10000)
474 anti-serum was a kind gift from Michael Wegner (Friedrich-Alexander University, Erlangen,
475 Germany). For Pax7, we used an antigen retrieval step. For this, after fixation and PBS washing,
476 slides were incubated in antigen retrieval buffer (diluted 1:100 in water; Vector) pre-heated to
477 80°C for 15 minutes. Slides were washed in PBS and continued at permeabilization step. Cy2-,
478 Cy3- and Cy5-conjugated secondary antibodies were purchased from Dianova.

479

480 **FISH experiments using isolated single fibers**

481 We isolated single extensor digitorum longus (EDL) muscle fibers as described before[61]. Isolated
482 EDL fibers were immediately fixed with 4% PFA, and were subjected hybridization in 1.5ml
483 tubes. After DAPI staining, fibers were transferred on slide glasses and mounted. Spindle fibers
484 were vulnerable to collagenase treatment. Thus, we pre-fixed the EDL tissue and peeled off
485 spindles under the fluorescent dissecting microscope (Leica).

486

487 **Acquisition of fluorescence images**

488 Fluorescence was visualized by laser-scanning microscopy (LSM700, Carl-Zeiss) using Zen
489 2009 software. Images were processed using ImageJ and Adobe Photoshop, and assembled using
490 Adobe Illustrator.

491

492 **Cell culture**

493 C2C12 cell line was purchased from ATCC, and cultured in high glucose DMEM (Gibco)
494 supplemented with 10% FBS (Sigma) and Penicilllin-Streptomycin (Sigma). To engineer C2C12
495 cells with doxycycline (Sigma) inducible Ebf1, mouse Ebf1 cDNA (Addgene) was cloned into
496 pLVX Tet-One Puro plasmid (Clontech), packaged in 293T cells (from ATCC) using psPAX2
497 and VsvG (Addgene), followed by viral transduction to C2C12 cells with 5µg/µl polybrene
498 (Millipore). Transduced cells were selected using 3µg/µl puromycin (Sigma).

499

500 **Western blotting**

501 Cell pellets were resuspended in NP-40 lysis buffer (1% NP-40, 150mM NaCl, 50mM Tris-Cl
502 pH 7.5, 1mM MgCl2 supplemented with protease (Roche) and phosphatase (Sigma) inhibitors),
503 and incubated on ice for 20 minutes. Lysates were cleared by centrifuging in 16,000g for 20 min
504 at 4°C. Protein concentration was measured by Bradford assay (Biorad), and lysates were boiled
505 in Laemilli buffer with beta-mercaptoethanol for 10 minutes. Denatured lysates were fractionated
506 by SDS-PAGE, transferred into nitrocellulose membrane, blocked with 5% milk and 0.1%
507 Tween-20 in PBS, and incubated overnight in 4°C with primary antibodies diluted in 5% BSA
508 and 0.1% Tween-20 in PBS. After three times washing with PBST, membranes were incubated
509 with secondary antibodies diluted in blocking solution for one hour at room temperature. After
510 PBST washing, membranes were developed with prime ECL (Amarsham). The antibodies used
511 for this study were β-actin (Cell Signaling, 1:1000), ColVI (Abcam, 1:2000) and Ebf1 (1:1000).

512

513 **RT-qPCR**

Cell pellets were resuspended in 1ml Trizol (Thermofisher). RNA was isolated according to manufacturer's guideline. 1µg of isolated RNA and random hexamer primer (Thermofisher) were used for reverse transcription using ProtoSciprt II RT (NEB). Synthesized cDNA was diluted five times in water, and 1µl was used per one qPCR reaction. qPCR was performed using 2X Sybergreen mix (Thermofisher) and CFX96 machine (Biorad). We used β-actin for normalization. Primers were designed using the 'Primer bank' website. Following primers were used for this study. β-actin For; 5'-GGCTGTATTCCCCTCCATCG-3', β-actin Rev; 5'-CCAGTTGGTAACAATGCCATGT-3', Pdgfrb For; 5'-TTCCAGGAGTGATACCAGCTT-3', Pdgfrb Rev; 5'-AGGGGGCGTGATGACTAGG-3', Col1a1 For; 5'-GCTCCTCTTAGGGGCCACT-3', Col1a1 Rev; 5'-CCACGTCTCACCATTGGGG-3', Col5a3 For; 5'-CGGGGTACTCCTGGTCCTAC-3', Col5a3 Rev; 5'-GCATCCCTACTTCCCCCTTG-3', Col6a1 For; 5'-CTGCTGCTACAAGCCTGCT-3', Col6a1 Rev; 5'-CCCCATAAGGTTTCAGCCTCA-3', Col6a3 For; 5'-GCTGCGGAATCACTTTGTGC-3', Col6a3 Rev; 5'-CACCTTGACACCTTTCTGGGT-3', Lrp1 For; 5'-ACTATGGATGCCCCTAAAACTTG-3', Lrp1 Rev; 5'-GCAATCTCTTTCACCGTCACA-3', Mgp For; 5'-GGCAACCCTGTGCTACGAAT-3', Mgp Rev; 5'-CCTGGACTCTCTTTTGGGCTTTA-3', Fstl1 For; 5'-CACGGCGAGGAGGAACCTA-3', Fst1l Rev; 5'-TCTTGCCATTACTGCCACACA-3', Cilp For; 5'-ATGGCAGCAATCAAGACTTGG-3' and Cilp Rev; 5'-AGGCTGGACTCTTCTCACTGA-3'.

Human biopsies

Human muscle biopsy specimens were obtained from M. vastus lateralis. The tissue was snap frozen under cryoprotection. Research use of the material was approved by the regulatory agencies (EA1/203/08, EA2/051/10, EA2/175/17, Charité Universitätsmedizin Berlin, Germany). Informed consent was obtained from the donors.

References

1 Stoeger, T., Battich, N. & Pelkmans, L. Passive Noise Filtering by Cellular Compartmentalization. *Cell* **164**, 1151-1161, doi:10.1016/j.cell.2016.02.005 (2016).

2 Sanes, J. R. & Lichtman, J. W. Development of the vertebrate neuromuscular junction. *Annual review of neuroscience* **22**, 389-442, doi:10.1146/annurev.neuro.22.1.389 (1999).

3 Hall, Z. W. & Sanes, J. R. Synaptic structure and development: the neuromuscular junction. *Cell* **72 Suppl**, 99-121, doi:10.1016/s0092-8674(05)80031-5 (1993).

4 Jasmin, B. J., Lee, R. K. & Rotundo, R. L. Compartmentalization of acetylcholinesterase mRNA and enzyme at the vertebrate neuromuscular junction. *Neuron* **11**, 467-477, doi:10.1016/0896-6273(93)90151-g (1993).

5 Schaeffer, L., de Kerchove d'Exaerde, A. & Changeux, J. P. Targeting transcription to the neuromuscular synapse. *Neuron* **31**, 15-22, doi:10.1016/s0896-6273(01)00353-1 (2001).

6 Charvet, B., Ruggiero, F. & Le Guellec, D. The development of the myotendinous junction. A review. *Muscles, ligaments and tendons journal* **2**, 53-63 (2012).

7 Maartens, A. P. & Brown, N. H. The many faces of cell adhesion during Drosophila muscle development. *Developmental biology* **401**, 62-74, doi:10.1016/j.ydbio.2014.12.038 (2015).

8 Schweitzer, R., Zelzer, E. & Volk, T. Connecting muscles to tendons: tendons and musculoskeletal development in flies and vertebrates. *Development* **137**, 2807-2817, doi:10.1242/dev.047498 (2010).

9 Giordani, L. *et al.* High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations. *Molecular cell* **74**, 609-621 e606, doi:10.1016/j.molcel.2019.02.026 (2019).

10 Dell'Orso, S. *et al.* Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and regenerative conditions. *Development* **146**, doi:10.1242/dev.174177 (2019).

11 Rubenstein, A. B. *et al.* Single-cell transcriptional profiles in human skeletal muscle. *Scientific reports* **10**, 229, doi:10.1038/s41598-019-57110-6 (2020).

12 Zeng, W. *et al.* Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity. *Nucleic acids research* **44**, e158, doi:10.1093/nar/gkw739 (2016).

13 van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature methods* **14**, 935-936, doi:10.1038/nmeth.4437 (2017).

14 Machado, L. *et al.* In Situ Fixation Redefines Quiescence and Early Activation of Skeletal Muscle Stem Cells. *Cell reports* **21**, 1982-1993, doi:10.1016/j.celrep.2017.10.080 (2017).

15 Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome biology* **17**, 77, doi:10.1186/s13059-016-0938-8 (2016).

16 Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular cell* **65**, 631-643 e634, doi:10.1016/j.molcel.2017.01.023 (2017).

17 Augusto, V. *et al.* Skeletal muscle fiber types in C57BL6J mice. *Braz. J. morphol. Sci.* **21**, 2, 89-94 (2004).

18 Schiaffino, S. & Mammucari, C. Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: insights from genetic models. *Skeletal muscle* **1**, 4, doi:10.1186/2044-5040-1-4 (2011).

19 Wiederstein, J. L. *et al.* Skeletal Muscle-Specific Methyltransferase METTL21C Trimethylates p97 and Regulates Autophagy-Associated Protein Breakdown. *Cell reports* **23**, 1342-1356, doi:10.1016/j.celrep.2018.03.136 (2018).

20 Tintignac, L. A., Brenner, H. R. & Ruegg, M. A. Mechanisms Regulating Neuromuscular Junction Development and Function and Causes of Muscle Wasting. *Physiological reviews* **95**, 809-852, doi:10.1152/physrev.00033.2014 (2015).

21 Bao, Z. Z., Lakonishok, M., Kaufman, S. & Horwitz, A. F. Alpha 7 beta 1 integrin is a component of the myotendinous junction on skeletal muscle. *Journal of cell science* **106 (Pt 2)**, 579-589 (1993).

22 Charvet, B. *et al.* Knockdown of col22a1 gene in zebrafish induces a muscular dystrophy by disruption of the myotendinous junction. *Development* **140**, 4602-4613, doi:10.1242/dev.096024 (2013).

23 Can, T. *et al.* Proteomic analysis of laser capture microscopy purified myotendinous junction regions from muscle sections. *Proteome science* **12**, 25, doi:10.1186/1477-5956-12-25 (2014).

24 Jobsis, G. J. *et al.* Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nature genetics* **14**, 113-115, doi:10.1038/ng0996-113 (1996).

25 Baghdadi, M. B. *et al.* Reciprocal signalling by Notch-Collagen V-CALCR retains muscle stem cells in their niche. *Nature* **557**, 714-718, doi:10.1038/s41586-018-0144-9 (2018).

26 Labialle, S. *et al.* The miR-379/miR-410 cluster at the imprinted Dlk1-Dio3 domain controls neonatal metabolic adaptation. *The EMBO journal* **33**, 2216-2230, doi:10.15252/embj.201387038 (2014).

27 Wust, S. *et al.* Metabolic Maturation during Muscle Stem Cell Differentiation Is Achieved by miR-1/133a-Mediated Inhibition of the Dlk1-Dio3 Mega Gene Cluster. *Cell metabolism* **27**, 1026-1039 e1026, doi:10.1016/j.cmet.2018.02.022 (2018).

28 Christodoulou, A., Santarella-Mellwig, R., Santama, N. & Mattaj, I. W. Transmembrane protein TMEM170A is a newly discovered regulator of ER and nuclear envelope morphogenesis in human cells. *Journal of cell science* **129**, 1552-1565, doi:10.1242/jcs.175273 (2016).

29 Jacob, A. *et al.* Rab40b regulates trafficking of MMP2 and MMP9 during invadopodia formation and invasion of breast cancer cells. *Journal of cell science* **126**, 4647-4658, doi:10.1242/jcs.126573 (2013).

30 Hainzl, T., Huang, S. & Sauer-Eriksson, A. E. Structural insights into SRP RNA: an induced fit mechanism for SRP assembly. *Rna* **11**, 1043-1050, doi:10.1261/rna.2080205 (2005).

31 Passerieux, E. *et al.* Structural organization of the perimysium in bovine skeletal muscle: Junctional plates and associated intracellular subdomains. *Journal of structural biology* **154**, 206-216, doi:10.1016/j.jsb.2006.01.002 (2006).

32 Haddix, S. G., Lee, Y. I., Kornegay, J. N. & Thompson, W. J. Cycles of myofiber degeneration and regeneration lead to remodeling of the neuromuscular junction in two mammalian models of Duchenne muscular dystrophy. *PloS one* **13**, e0205926, doi:10.1371/journal.pone.0205926 (2018).

33 Pratt, S. J. P., Valencia, A. P., Le, G. K., Shah, S. B. & Lovering, R. M. Pre- and postsynaptic changes in the neuromuscular junction in dystrophic mice. *Frontiers in physiology* **6**, 252, doi:10.3389/fphys.2015.00252 (2015).

34 Law, D. J. & Tidball, J. G. Dystrophin Deficiency Is Associated with Myotendinous Junction Defects in Prenecrotic and Fully Regenerated Skeletal-Muscle. *American Journal of Pathology* **142**, 1513-1523 (1993).
35 Banks, G. B., Judge, L. M., Allen, J. M. & Chamberlain, J. S. The Polyproline Site in Hinge 2 Influences the Functional Capacity of Truncated Dystrophins. *PLoS genetics* **6**, doi:ARTN e1000958

10.1371/journal.pgen.1000958 (2010).
36 Bell, C. D. & Conen, P. E. Histopathological Changes in Duchenne Muscular Dystrophy. *J Neurol Sci* **7**, 529-&, doi:Doi 10.1016/0022-510x(68)90058-0 (1968).
37 Ravenscroft, G. *et al.* Mutations in KLHL40 are a frequent cause of severe autosomal-recessive nemaline myopathy. *American journal of human genetics* **93**, 6-18, doi:10.1016/j.ajhg.2013.05.004 (2013).
38 Vorgerd, M. *et al.* A mutation in the dimerization domain of filamin c causes a novel type of autosomal dominant myofibrillar myopathy. *American journal of human genetics* **77**, 297-304, doi:10.1086/431959 (2005).
39 Windpassinger, C. *et al.* An X-linked myopathy with postural muscle atrophy and generalized hypertrophy, termed XMPMA, is caused by mutations in FHL1. *American journal of human genetics* **82**, 88-99, doi:10.1016/j.ajhg.2007.09.004 (2008).
40 Huang, Y. *et al.* AHNAK, a novel component of the dysferlin protein complex, redistributes to the cytoplasm with dysferlin during skeletal muscle regeneration. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **21**, 732-742, doi:10.1096/fj.06-6628com (2007).
41 Molt, S. *et al.* Aciculin interacts with filamin C and Xin and is essential for myofibril assembly, remodeling and maintenance. *Journal of cell science* **127**, 3578-3592, doi:10.1242/jcs.152157 (2014).
42 Juo, L. Y. *et al.* HSPB7 interacts with dimerized FLNC and its absence results in progressive myopathy in skeletal muscles. *Journal of cell science* **129**, 1661-1670, doi:10.1242/jcs.179887 (2016).
43 Leber, Y. *et al.* Filamin C is a highly dynamic protein associated with fast repair of myofibrillar microdamage. *Human molecular genetics* **25**, 2776-2788, doi:10.1093/hmg/ddw135 (2016).
44 Otten, C. *et al.* Xirp proteins mark injured skeletal muscle in zebrafish. *PloS one* **7**, e31041, doi:10.1371/journal.pone.0031041 (2012).
45 Bansal, D. *et al.* Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature* **423**, 168-172, doi:10.1038/nature01573 (2003).
46 Albert, Y. *et al.* Transcriptional regulation of myotube fate specification and intrafusal muscle fiber morphogenesis. *The Journal of cell biology* **169**, 257-268, doi:10.1083/jcb.200501156 (2005).
47 Cheret, C. *et al.* Bace1 and Neuregulin-1 cooperate to control formation and maintenance of muscle spindles. *The EMBO journal* **32**, 2015-2028, doi:10.1038/emboj.2013.146 (2013).
48 Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiological reviews* **91**, 1447-1531, doi:10.1152/physrev.00031.2010 (2011).
49 Rossi, A. C., Mammucari, C., Argentini, C., Reggiani, C. & Schiaffino, S. Two novel/ancient myosins in mammalian skeletal muscles: MYH14/7b and MYH15 are

expressed in extraocular muscles and muscle spindles. *The Journal of physiology* **588**, 353-364, doi:10.1113/jphysiol.2009.181008 (2010).

50 Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55-60, doi:10.1126/science.1193270 (2010).

51 Hippenmeyer, S., Huber, R. M., Ladle, D. R., Murphy, K. & Arber, S. ETS transcription factor Erm controls subsynaptic gene expression in skeletal muscles. *Neuron* **55**, 726-740, doi:10.1016/j.neuron.2007.07.028 (2007).

52 Li, Y. *et al.* Molecular layer perforant path-associated cells contribute to feed-forward inhibition in the adult dentate gyrus. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9106-9111, doi:10.1073/pnas.1306912110 (2013).

53 Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263-267, doi:10.1038/376263a0 (1995).

54 Malcher, J. *et al.* Exon Skipping in a Dysf-Missense Mutant Mouse Model. *Molecular therapy. Nucleic acids* **13**, 198-207, doi:10.1016/j.omtn.2018.08.013 (2018).

55 Wurmus, R. *et al.* PiGx: reproducible genomics analysis pipelines with GNU Guix. *GigaScience* **7**, doi:10.1093/gigascience/giy123 (2018).

56 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

57 Alles, J. *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC biology* **15**, 44, doi:10.1186/s12915-017-0383-5 (2017).

58 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

59 Muller, T. *et al.* The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron* **34**, 551-562, doi:10.1016/s0896-6273(02)00689-x (2002).

60 Roessler, S. *et al.* Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Molecular and cellular biology* **27**, 579-594, doi:10.1128/MCB.01192-06 (2007).

61 Vogler, T. O., Gadek, K. E., Cadwallader, A. B., Elston, T. L. & Olwin, B. B. Isolation, Culture, Functional Assays, and Immunofluorescence of Myofiber-Associated Satellite Cells. *Methods in molecular biology* **1460**, 141-162, doi:10.1007/978-1-4939-3810-0_11 (2016).

Acknowledgements

We thank Dr. K. Song for advice and protocols on snRNAseq, E. Lowenstein and Dr. T. Muller for critical reading of the manuscript, C. Paeseler and P. Stallerow for help with animal care (all MDC), as well as the MDC core facilities for flow cytometry (led by Dr. Hans-Peter Rahn) and next generation sequencing (led by Dr. Sascha Sauer). We are grateful to Dr. Martyn Goulding (Salk institute) for providing the H2B-GFP reporter mouse line, and Drs. M. Derecka and R. Grosschedl (MPI for Immunobiology and Epigenetics) for providing Ebf1 tissues and reagents.

This work was supported by the Helmholtz association (C.B) and AVH postdoctoral fellowship (M.K).

Author contributions

M.K and C.B conceived the work and designed the project. M.K. led the project, performed the experiments and analyzed the data. V.F and A.A performed the bioinformatics analysis. B.B contributed to the experiments, especially generated the sequencing library. S.S collected and prepared human patient biopsies and *Dysferlin* mutant mouse samples. M.K and C.B wrote the manuscript with comments by all authors.

Competing interests: We have no conflicting interests.

Materials & correspondence: Requests for reagents should be addressed to cbirch@mdc-berlin.de

Data availability: All the raw sequencing data have been deposited in Array Express (Accession number: E-MTAB-8623). The data will be made public after publication.

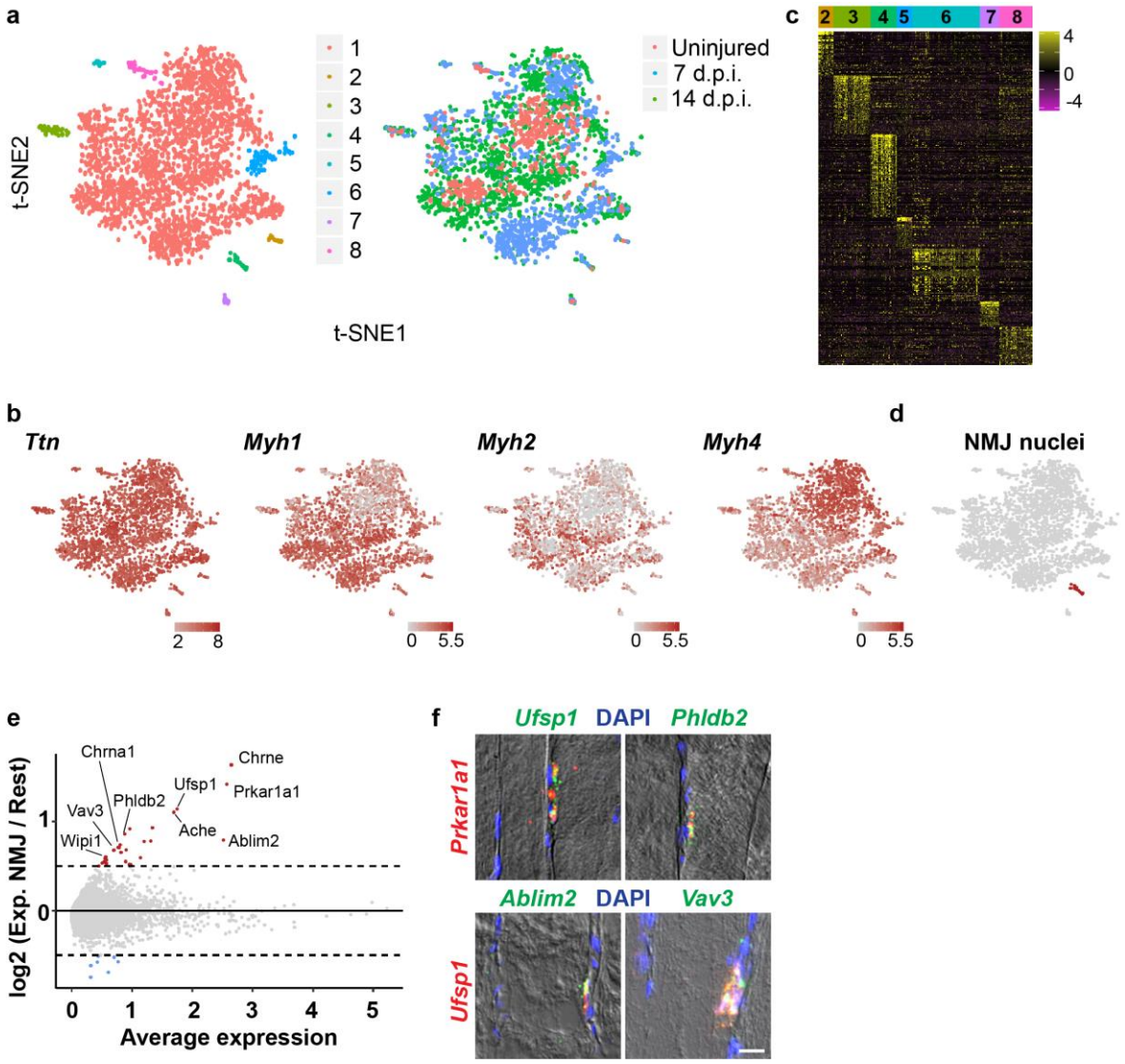

Figure 1. Nuclear heterogeneity in uninjured and regenerating muscles. a, t-SNE map of transcripts detected in nuclei of uninjured and regenerating muscles; the colors identify different nuclear populations (left) or nuclei in uninjured or regenerating muscle (right). **b**, Expression of *Ttn* and *Myosin* genes identifies myonuclei. **c**, Heat-map of specific genes enriched in clusters 2-8. **d**, Detection of NMJ nuclei. **e**, Volcano plot of NMJ nuclei enriched genes; Exp., expression. **f**, Upper row - double FISH against a known NMJ marker (*Prkar1a1*) and newly identified NMJ genes (green). Bottom row – double single molecule FISH (smFISH; RNAscope) against *Ufsp1* (red) and other newly identified NMJ genes (green). Expression patterns were validated in 2 or more individuals. Scale bar, 10µm.

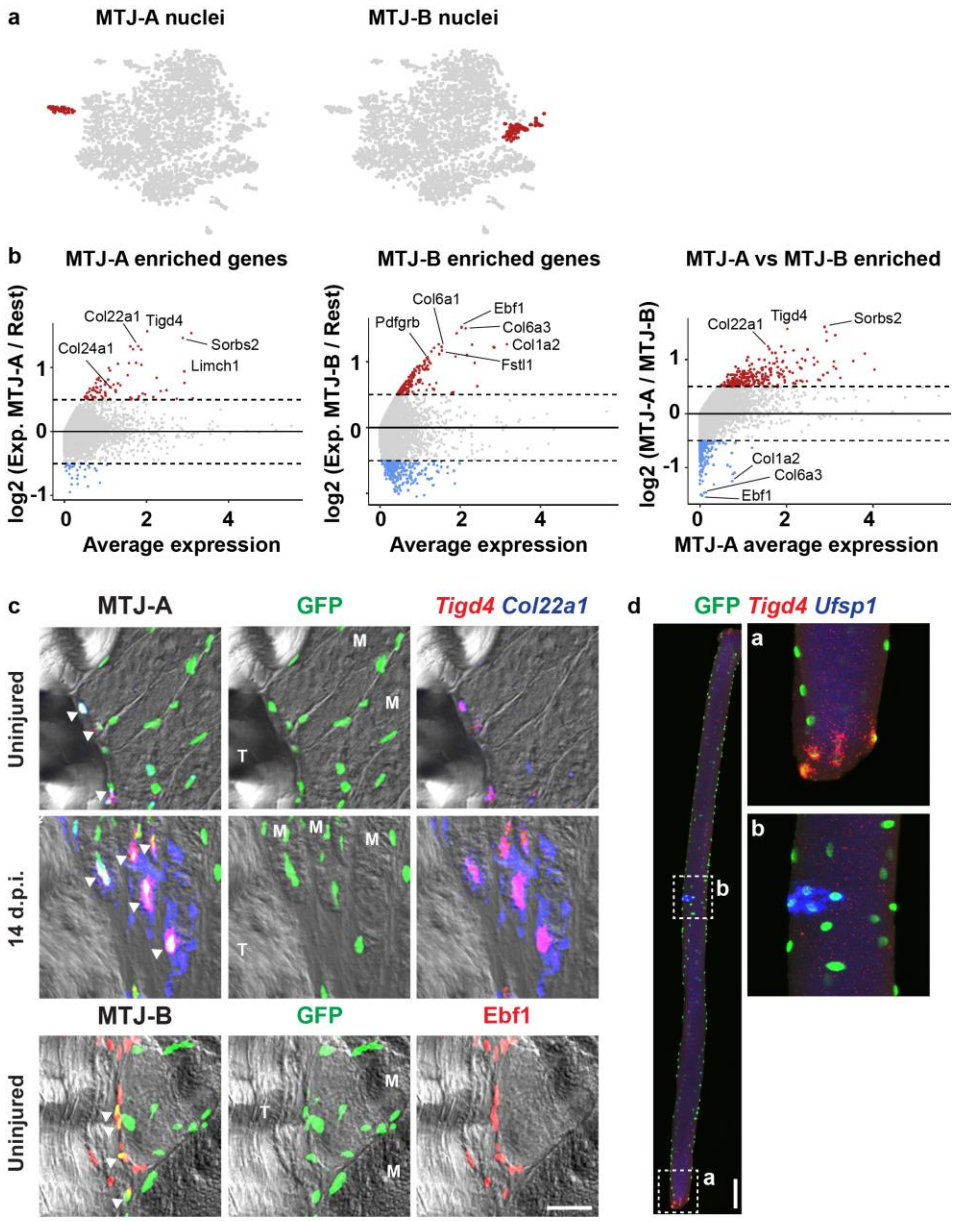

Figure 2. Two distinct nuclear populations at the myotendinous junction. a, Detection of MTJ-A and MTJ-B clusters. **b**, Enriched genes expressed in MTJ-A and MTJ-B nuclei, and comparison of genes enriched in MTJ-A and -B clusters. **c**, Upper two rows – smFISH of two MTJ-A markers (*Tigd4* and *Col22a1*) in uninjured or 14 d.p.i TA muscle expressing H2B-GFP in myonuclei. Bottom row – Ebf1 (MTJ-B marker) immunohistochemistry in uninjured TA muscle expressing H2B-GFP in myonuclei. Shown are MTJ regions; T, tendon and M, myofiber. Arrowheads indicate co-localization of MTJ marker genes and GFP. Scale bar, 30µm. **d**, Double FISH experiment in isolated single EDL fiber. Insets show magnification of MTJ (a) and NMJ (b) regions. Scale bar, 100µm. Expression patterns were validated in 2 or more individuals.

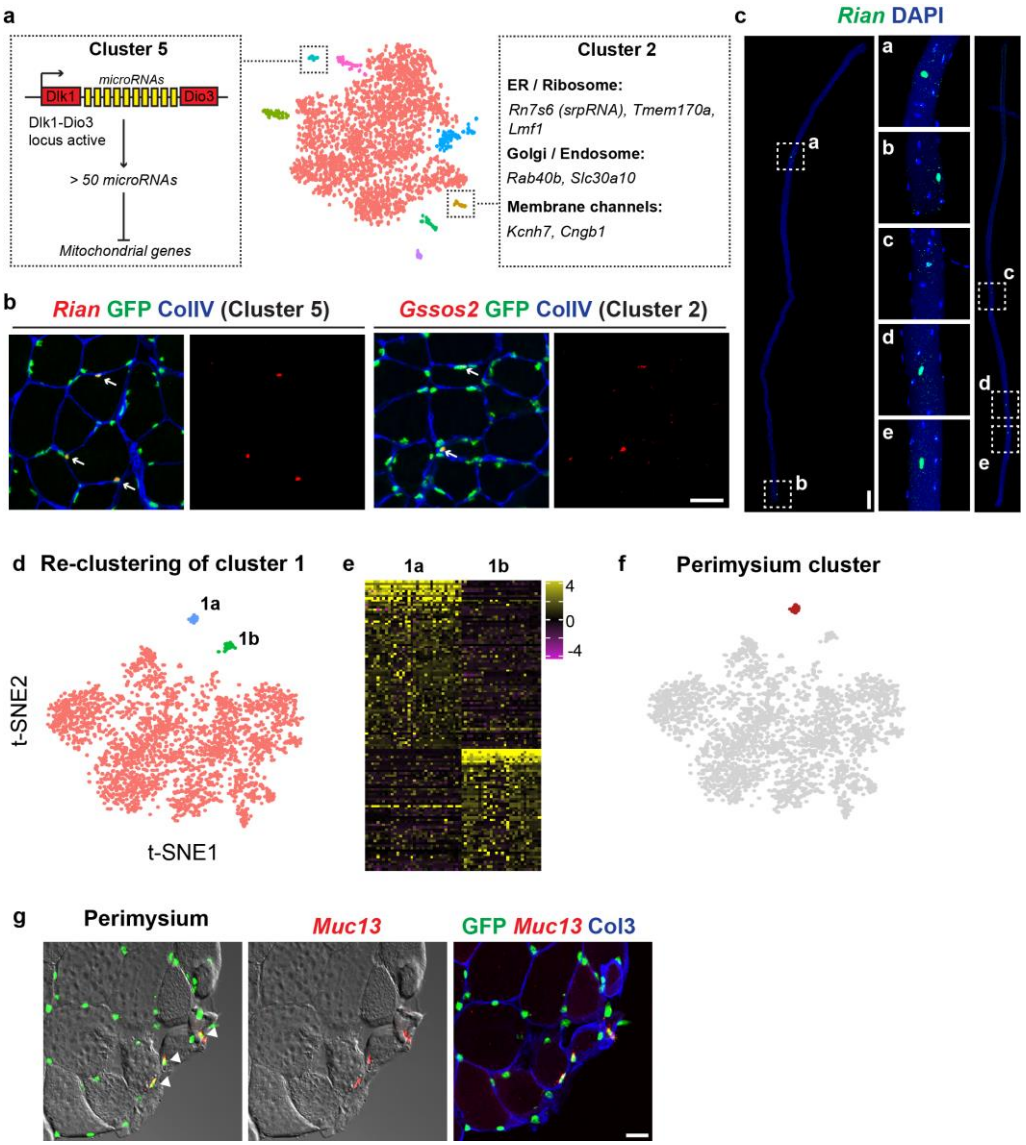

Figure 3. Identification of novel nuclear subtypes. **a**, Illustrations of potential functions of clusters 5 and 2. Cluster 5 might regulate local mitochondrial metabolism through microRNAs embedded in Dlk1-Dio3 locus, whereas cluster 2 potentially regulates local protein synthesis and entry into the secretory pathway. **b**, Validation of the top markers of clusters 5 and 2 by smFISH. Note their strong expression in a subset of myonuclei. Scale bar, 30µm. **c**, Expression of *Rian* in isolated EDL fibers. Insets show magnifications of indicated areas. Scale bar, 100 µm. **d**, t-SNE map of re-clustering of the major population identified in Figure 1a. Note the emergence of two more distinct sub-populations, 1a and 1b. **e**, Heat-map showing differentially expressed genes in clusters 1a and 1b. **f**, Location of the perimysial sub-population (1a) in the t-SNE map. **g**, Validation of Muc13 in myonuclei adjacent to the perimysium by smFISH. Scale bar, 30µm. Expression patterns were validated in 2 or more individuals.

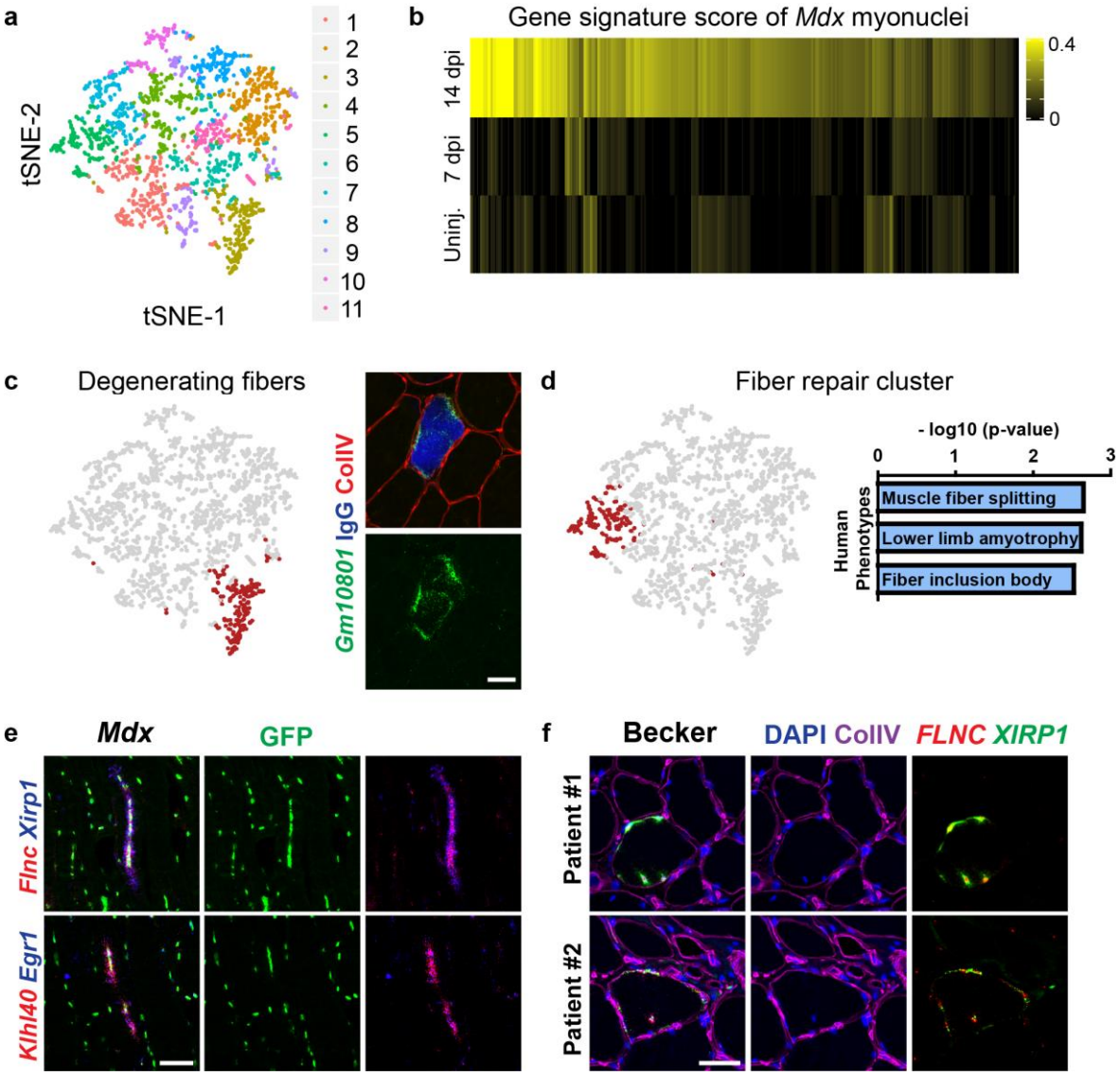

Figure 4. snRNAseq analysis of *Mdx* muscle. a, t-SNE map of Mdx myonuclei (1,939 nuclei). **b**, Each Mdx myonucleus was assigned a gene signature score, i.e. a gene expression score indicating similarity with uninjured (uninj.), 7 or 14 d.p.i. myonuclei. Each column represents an individual nucleus. **c**, Location of the cluster representing degenerating (necrotic) fibers in the t-SNE map (left). An ncRNA identified in this cluster (*Gm10801*) is expressed in IgG-positive fibers, indicating that it defines nuclei of necrotic fibers. **d**, Nuclear population expressing high levels of various genes implicated in fiber repair that are mutated in myopathies. **e**, Indicated marker genes of the cluster in d are co-expressed in longitudinal sections of muscle of *Mdx* mice. **f**, Co-expression of *FLNC* and *XIRP1* in the muscle from Becker's syndrome patients. Control images for e and f are shown in Extended Data Fig. 8a. All scale bars, 50µm.

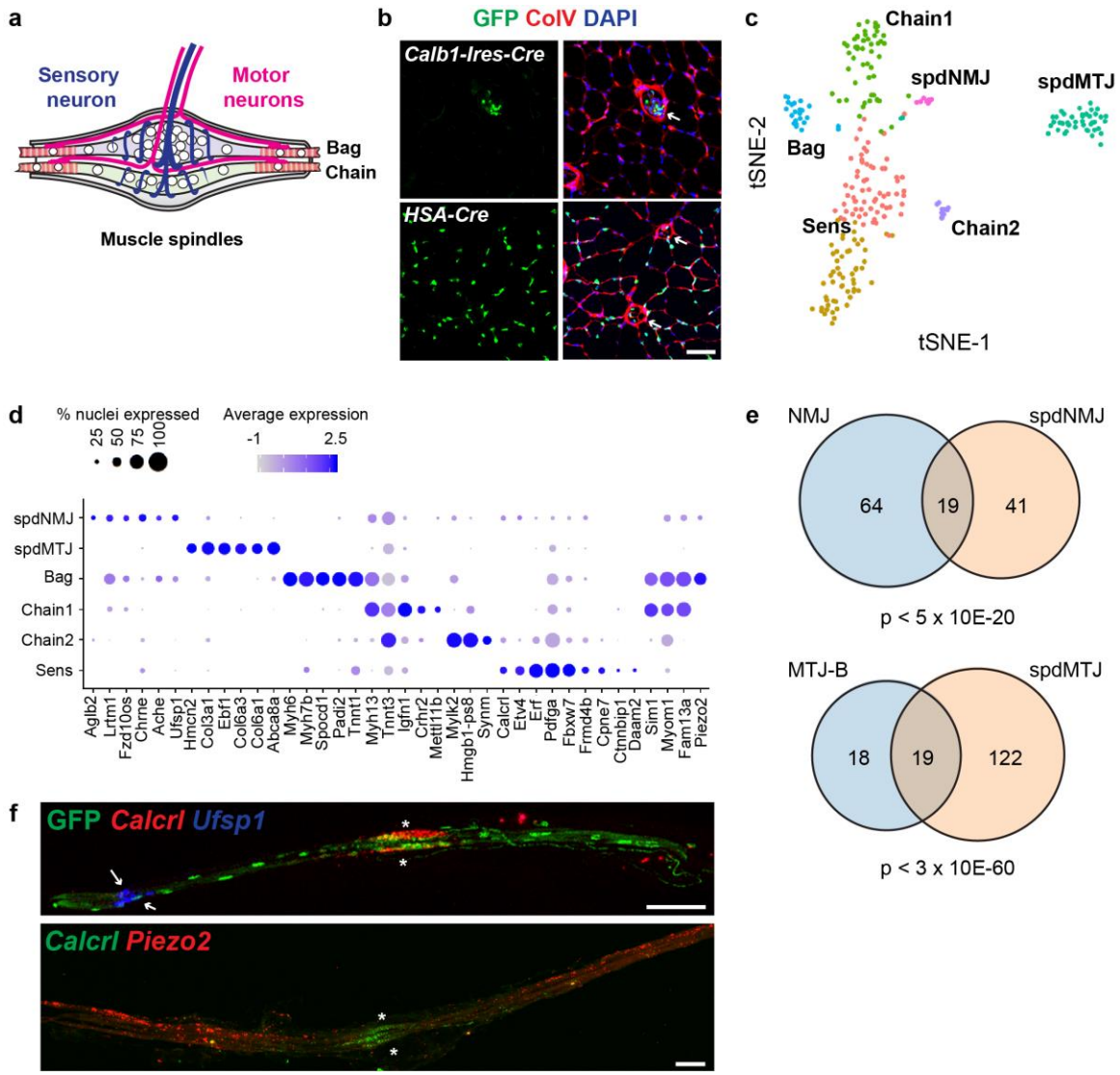

Figure 5. Functional compartments inside muscle spindle fibers. a, Cartoon showing the structure of the muscle spindle. **b**, Specific labeling of spindle myonuclei using *Calb1-Ires-Cre.* Arrows indicate muscle spindles**. c,** t-SNE map of muscle spindle myonuclei (260 nuclei). **d,** Expression map of nuclear populations identified in c. **e,** Venn diagram comparing 'spdNMJ vs extrafusal fiber NMJ' and 'spdMTJ vs extrafusal fiber MTJ-B'. Genes enriched in each population (average logFC > 0.7) were used to generate the diagrams. Statistical analysis was performed using hypergeometric test using all genes detected in uninjured/regenerating and spindle datasets as background. **f,** Double smFISH experiments of isolated muscle spindle fibers. Arrows indicate spdNMJ, and asterisks the central non-contractile parts of spindle fibers. All scale bars, 50µm.