CDDpress

www.nature.com/cddis

ARTICLE OPEN

Decoding the transcriptome of Duchenne muscular dystrophy to the single nuclei level reveals clinical-genetic correlations

Xavier Suárez-Calvet[1,6], Esther Fernández-Simón[2,6], Daniel Natera[3], Cristina Jou[4], Patricia Pinol-Jurado[2], Elisa Villalobos[2], Carlos Ortez[3], Alexandra Monceau[2], Marianela Schiava[2], Anna Codina[4], José Verdu-Díaz[2], James Clark[2], Zoe Laidler[2], Priyanka Mehra[2], Rasya Gokul-Nath[2], Jorge Alonso-Perez[5], Chiara Marini-Bettolo[2], Giorgio Tasca[2], Volker Straub[2], Michela Guglieri[2], Andrés Nascimento[3] and Jordi Diaz-Manera[1,2]✉

Duchenne muscular dystrophy is a genetic disease produced by mutations in the dystrophin gene characterized by early onset muscle weakness leading to severe and irreversible disability. The cellular and molecular consequences of the lack of dystrophin in humans are only partially known, which is crucial for the development of new therapies aiming to slow or stop the progression of the disease. Here we have analyzed quadriceps muscle biopsies of seven DMD patients aged 2 to 4 years old and five age and gender matched controls using single nuclei RNA sequencing (snRNAseq) and correlated the results obtained with clinical data. SnRNAseq identified significant differences in the proportion of cell population present in the muscle samples, including an increase in the number of regenerative fibers, satellite cells, and fibro-adipogenic progenitor cells (FAPs) and a decrease in the number of slow fibers and smooth muscle cells. Muscle samples from the younger patients with stable mild weakness were characterized by an increase in regenerative fibers, while older patients with moderate and progressive weakness were characterized by loss of muscle fibers and an increase in FAPs. An analysis of the gene expression profile in muscle fibers identified a strong regenerative signature in DMD samples characterized by the upregulation of genes involved in myogenesis and muscle hypertrophy. In the case of FAPs, we observed upregulation of genes involved in the extracellular matrix regeneration but also several signaling pathways. Indeed, further analysis of the potential intercellular communication profile showed a dysregulation of the communication profile in DMD samples identifying FAPs as a key regulator of cell signaling in DMD muscle samples. In conclusion, our study has identified significant differences at the cellular and molecular levels in the different cell populations present in skeletal muscle samples of patients with DMD compared to controls.

Cell Death and Disease (2023)14:596 ; https://doi.org/10.1038/s41419-023-06103-5

INTRODUCTION

Duchenne muscular dystrophy (DMD) is a genetic disorder produced by mutations in the dystrophin gene and characterized by early onset progressive muscle weakness leading to irreversible severe disability [1, 2]. Treatment with corticosteroids is part of the standard for care as they delay disease progression, although unfortunately do not change substantially the natural history of the disease [1, 3, 4]. Several new therapies have been tested or are still under research in clinical trials using different strategies, including but not limited to cell therapy, gene therapy, anti-inflammatory, pro-regenerative, and, antioxidant drugs [5].

Dystrophin is localized in the cytoplasmic face of the muscle membrane linking the sarcomeric proteins to the sarcolemma and the extracellular matrix (ECM) [6]. Muscle fibers lacking dystrophin are injured during muscle contraction leading to continuous cycles of muscle regeneration that ultimately fail resulting in the loss of muscle fibers and expansion of fat and fibrotic tissue. The process of muscle degeneration in DMD involves a complex interplay between muscle fibers, muscle resident cells such as satellite cells and fibroadipogenic progenitor cells (FAPs), and, circulating cells infiltrating the muscle such as macrophages and lymphocytes [7, 8]. Despite considerable progress in the understanding of the degenerative process in DMD, there is still a considerable lack of knowledge of what are the cellular and molecular consequences of the absence of dystrophin in humans [9]. Most of the studies performed in humans have analyzed muscle samples using bulk proteomics or RNA identifying dysregulated molecular pathways in DMD, but these studies are not able to characterize what cells are responsible for each pathway, or how cells interplay with each other during the process of muscle degeneration [10–13].

[1]Neuromuscular Diseases Unit, Department of Neurology, Hospital de la Santa Creu i Sant Pau, Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), 08041 Barcelona, Spain. [2]John Walton Muscular Dystrophy Research Center, Newcastle University Translational and Clinical Research Institute, Newcastle Upon Tyne, UK. [3]Neuromuscular Disorders Unit. Neurology department, Hospital Sant Joan de Deu, Esplugues de Llobregat, Spain. [4]Pathology department, Hospital Sant Joan de Deu, Esplugues de Llobregat, Spain. [5]Neuromuscular Disease Unit. Neurology Department. Hospital Nuestra Señora de Candelaria, Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Tenerife, Spain. [6]These authors contributed equally: Xavier Suárez-Calvet, Esther Fernández-Simón. ✉email: Jordi.diaz-manera@newcastle.ac.uk

Edited by Anastasis Stephanou

Received: 22 March 2023 Revised: 15 August 2023 Accepted: 22 August 2023
Published online: 07 September 2023

Single cell RNA sequencing (scRNAseq) and single nuclei RNA sequencing (snRNAseq) allows the analysis of gene expression to the single cell/nuclei levels using small fragments of tissue [14]. This technology enables the identification of the cell populations present in the tissue of interest in healthy and disease conditions, the study of differentially expressed genes in each cell population compared to controls, and, the identification of potential communications between cells easing the understanding of how the disease process is orchestrated and its dynamics along disease progression [15, 16]. In the case of DMD, both scRNAseq and snRNAseq offer a unique opportunity to study the changes in gene expression profiles in different muscle cell populations using biopsies that were obtained for diagnosis and stored in biobanks. SnRNAseq offers some advantages in the study of skeletal muscle biopsies compared to scRNAseq. First, snRNAseq allows to study of the gene expression of nuclei myofibers, which represent more than 80% of the nuclei present in muscle. Nuclei in myofiber are lost if scRNAseq is performed, as this technology requires viable cell suspension for sequencing [16–19]. Second, adipocyte fragility difficult their inclusion in scRNAseq studies with the risk of losing an important contributor to transcriptomic variability in the case of muscular dystrophies, such as DMD, while this difficulty is reduced if snRNAseq is used for the analysis [15]. Finally, scRNAseq is not convenient for frozen tissues, requiring fresh muscle, which is a clear limitation due to the reduced availability of muscle tissue from already diagnosed patients [20]. There have been some snRNAseq studies published so far using murine models of DMD that have provided valuable clues about the nature of the process of muscle degeneration, but the information coming from human samples is very limited [21, 22]. Here, we have applied snRNAseq to muscle samples obtained from DMD patients that were biopsied between the age of two and four years old, at the early stages of the disease before steroids were started. We have implemented a protocol that has allowed us to use a minimum amount of tissue, around 25 mg of frozen muscle, and obtain approximately between 10 to 20 thousand nuclei for the analysis [16, 23]. We are comparing the gene expression profile to the single nuclei level of these samples with age and gender matched controls to understand what are the cellular and molecular consequences of the lack of dystrophin that influence the process of muscle degeneration in humans.

RESULTS

Patients and samples included

We performed snRNAseq on 7 muscle samples of patients with DMD and 5 muscle samples of age and gender matched controls. Table 1 summarizes the main demographic, genetic, and clinical features of the patients included. Supplementary Fig. 1 shows representative areas of the haematoxylin-eosin (HE) staining of the muscle samples included in this study.

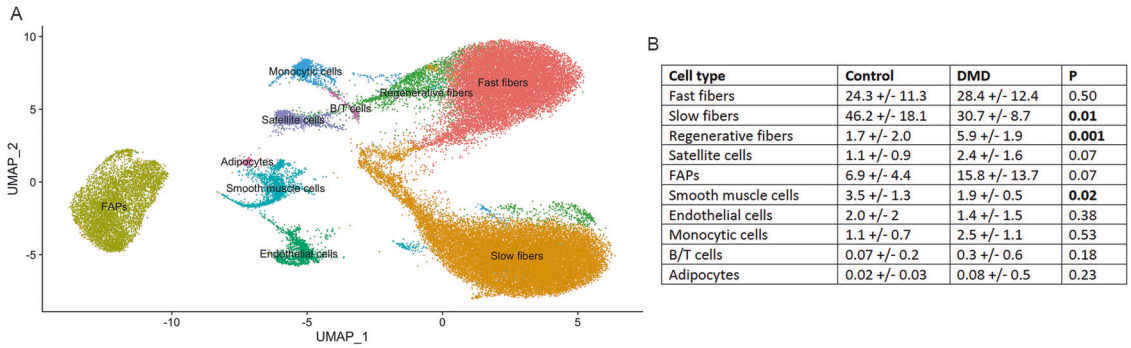
Characterization of cell populations identified in the skeletal muscle samples

A total of 30857 nuclei from controls and 25817 nuclei from DMD were included in the analysis. Unsupervised clustering using the Seurat package identified 19 different nuclei clusters (Supplementary Fig. 2) [24]. An analysis of the differentially expressed gene signatures allows us to attribute clusters to 10 putative identities (Fig. 1A and Supplementary Fig. 3). The largest number of nuclei in the samples were from myofibers expressing *Ckm*, a marker of mature myonuclei. As expected, we identified fast and slow type myofibers characterized by the expression of *Myh1* and *Myh2* and, *Myh7b* respectively, but also regenerative fibers characterized by the expression of *Myh3* and *Myh8* (Fig. 1 and Supplementary Fig. 4). Fast myofibers included type IIa fibers expressing *Myh2* and type IIx myofibers expressing *Myh1*, but we just identified a minority of nuclei expressing *Myh4* which is typical of type IIb

Table 1. Clinical and genetic data of DMD patients and control included in the study.

Code	Age	G.	Muscle	Mutation	NSAA Baseline	6MWT baseline	10MWT baseline	TSUF baseline	NSAA 4 y.	6MWT 4 y.	10MWT 4 y.	TSUF 4 y.
DMD-1	4 y 0 m	M	Quad.	c.6651_6652del; p.Asp2221Phefs*3	29	NA	NA	4.2 s	21	443 m	4.5 s	5.6 s
DMD-2	4 y 7 m	M	Quad.	Duplication Exon 3-7	28	NA	5.3 s	3.4 s	24	473 m	4.8 s	4.7 s
DMD-3	3 y 1 m	M	Quad.	Deletion Exon 48-55	32	NA	3.5 s	NA	32	515 m	3.2 s	1.9 s
DMD-4	2 y 2 m	M	Quad.	Deletion Exon 45-50	NA*	NA*	NA*	NA*	NA	NA	NA	NA
DMD-5	3 y 2 m	M	Quad.	c.4918_4919delinsTG	30	NA	5.4 s	3.2 s	31	NA	3.4 s	2.9 s
DMD-6	3 y 2 m	M	Quad.	c.5899 C > T; p.Arg1967X	32	NA	NA	4.0 s	34	NA	3.5 s	2 s
DMD-7	4 y 8 m	M	Quad.	Deletion Exon 46-52	21	NA	5.8 s	3.3 s	29	512 m	3.1	2.5 s
Control-1	8 y	M	Quad.									
Control-2	5 y	M	Quad.									
Control-3	9 y	M	Quad.									
Control-4	10 y	M	Quad.									
Control-5	10 y	M	Quad.									

G gender. *Quad.* quadriceps. *NSAA* North Star Ambulatory Assessment, *6MWT* Six minutes walking test, *10MWT* 10 min walking test, *TSUF* Time to stand up from floor, *S* seconds, *M* meters, *Y* years.



B

Cell type	Control	DMD	P
Fast fibers	24.3 +/- 11.3	28.4 +/- 12.4	0.50
Slow fibers	46.2 +/- 18.1	30.7 +/- 8.7	**0.01**
Regenerative fibers	1.7 +/- 2.0	5.9 +/- 1.9	**0.001**
Satellite cells	1.1 +/- 0.9	2.4 +/- 1.6	0.07
FAPs	6.9 +/- 4.4	15.8 +/- 13.7	0.07
Smooth muscle cells	3.5 +/- 1.3	1.9 +/- 0.5	**0.02**
Endothelial cells	2.0 +/- 2	1.4 +/- 1.5	0.38
Monocytic cells	1.1 +/- 0.7	2.5 +/- 1.1	0.53
B/T cells	0.07 +/- 0.2	0.3 +/- 0.6	0.18
Adipocytes	0.02 +/- 0.03	0.08 +/- 0.5	0.23

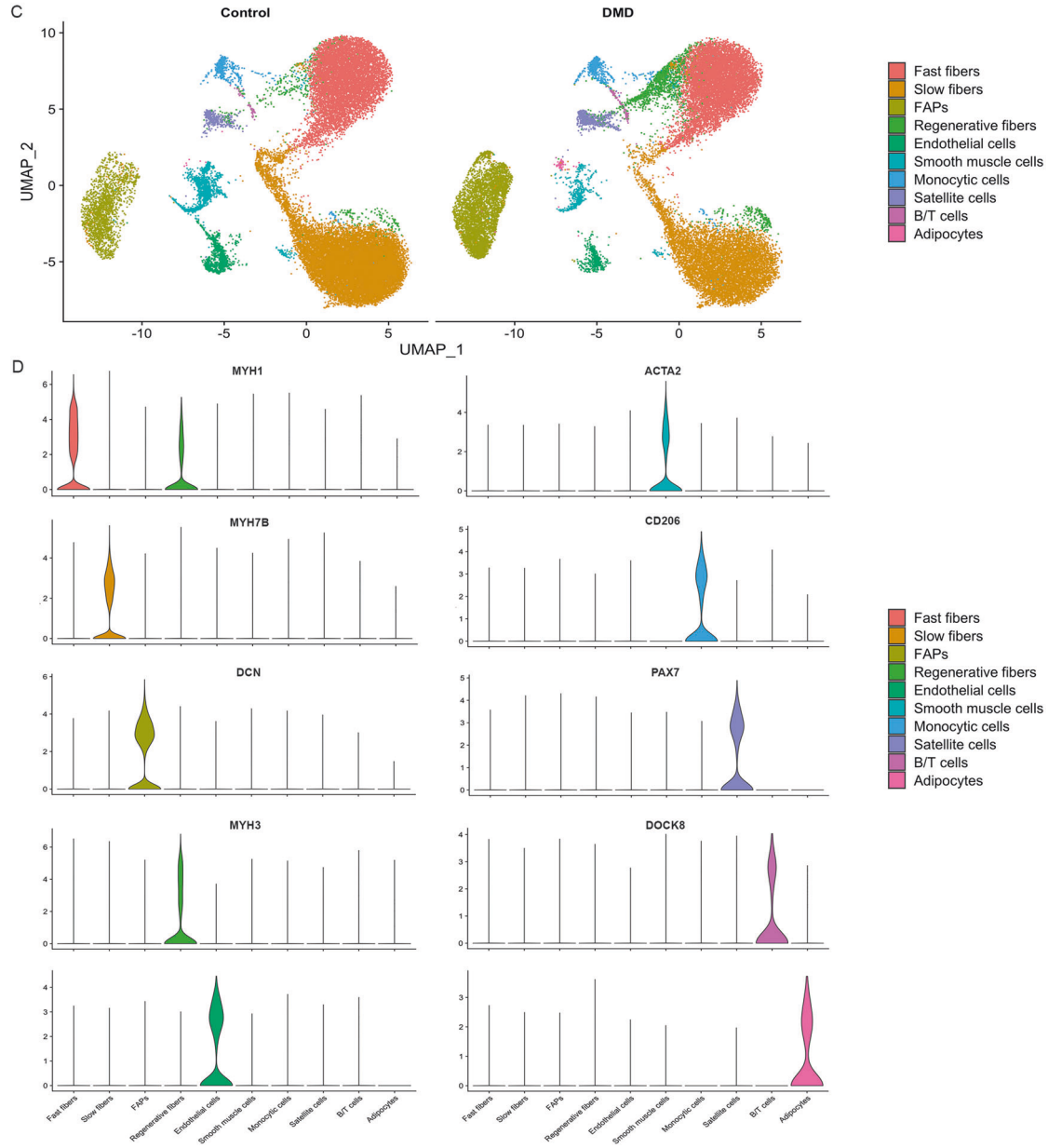


Fig. 1 Classification of nuclei/cell types in normal and DMD muscle samples. A UMAP visualization of all the nuclei from control and DMD samples colored by cluster identity. **B** Table comparing the proportion of cell population between control and DMD samples. **C** UMAP showing clusters identified in control (left) and DMD (samples). **D** Violin plots showing the expression of selected marker genes for each cluster of nuclei. FAPs fibroadipogenic progenitor cells.

myofibers. Fast myofibers cluster displayed high levels of genes encoding the fast isoforms of troponins (*Tnnt3*, *Tpm1*, and *Tnni2)*, the sarcoplasmic reticulum-calcium ATPase 1 (SERCA1/ *Atp2a1*), and glycolytic enzymes (*Eno3, Pfkm, Pkm* and *Pygm)* [25]. The slow myofibers cluster displayed high levels of genes encoding the slow isoforms of troponins (*Tnnt1*, *Tpm3, and Tnni1)*, slow myosin light chains *(Myl2* and *Myl3)* and the sarcoplasmic reticulum-calcium ATPase 2 (SERCA2/ *Atp2a2*). Regenerative fibers were identified by the expression of *Myh3* and *Myh8* that encode the embryonic and perinatal MyHC isoforms respectively, also *Tnnt2* that encodes for an isoform of Troponin-T expressed by cardiac muscle but also embryonic skeletal muscles [26]. As expected, regenerative fibers expressed high levels of *Ncam1*, involved in the reinnervation process of new fibers, and *Myog*, which encodes myogenin and is expressed by both fusing myoblast and newly regenerated fibers [27]. Closely located to the regenerating fibers in the UMAP, we identified a cluster of cells expressing *Pax7* that was identified as satellite cells and shared many genes with regenerative fibers in our samples. This expression profile similitudes illustrated by UMAP is compatible with the origin of regenerative fibers from satellite cells and leads us to study potential gene expression profiles driving cellular transitions using pseudotime trajectories as shown in Supplementary Fig. 5. As observed cells in the earliest stages of myoblast differentiation showed expression of genes related with ribosomal and mitochondrial function such as *Rpl30*, *Rpl37*, *Mt-co2* or *Mt-co3*, reorganization of the cytoskeleton such as *Myl2*, *Myl1* or *Acta1* and, genes promoting the differentiation of satellite cells such as *Meg3* or *Rassf4* [28, 29]. These were followed by myonuclei expressing genes expressed in fast fibers such as *Myh1*, *Tpm1*, *Tnni2*, and, finally, genes expressed in slow fibers such as *Myh7*, *Tnnt1 orTpm3*. Interestingly, we did not identify any population expressing genes related to the myotendinous junction, such as *Col22a1* or *Ankrd1*, as reported in snRNAseq studies done with murine samples that use the whole muscle for the analysis [16, 30, 31]. Moreover, we did not observe a specific cluster expressing genes of the neuromuscular junction (NMJ) as has also been reported in mice, although the expression of *Chrna1* or *Chrng*, was enriched in the regenerative fibers suggesting that there is an active process of remodeling of the NMJ in this cluster as has already been suggested (Supplementary Fig. 4) [11].

We observed six other non-myofiber clusters of nuclei that included endothelial cells expressing *Pecam1* and *Ptprb*, smooth muscle cells (SMC) expressing *Acta2*, *Pdgfrb* or *Myh11* or adipocytes expressing *Adipoq* (Fig. 1). Inflammatory cells characterized by the expression of *Ptprc/Cd45*, were further divided in macrophages expressing *Mrc1/Cd206* and B/T cells expressing *Dok8* [32, 33]. Within the macrophages, we identified nuclei expressing M2 markers such as *F13a1, Fcer2/Cd23*, and *Cd209* and nuclei expressing M1 markers such as *Cd44*, *Cd86*, *Tlr2*, and *Fcgr3a* [34]. Profibrotic genes, such as *Tgfb1* and *Spp1* were also expressed by M2 macrophages. A large cluster of nuclei was characterized by the expression of *Pdgfra* and *Dcn* which are well-known markers of fibroadipogenic progenitor cells being labeled as fibroadipogenic precursor cells (FAPs) and will be described later in detail.

Distinctive signature profile in skeletal muscle of patients with DMD

We identified significant differences in the proportion of each cell population when comparing the samples from control and DMD patients (Fig. 1B and C). Specifically, there was a significant decrease in the percentage of slow type I fibers (50.9% in controls vs 31.2% in DMD, $p = 0.01$, Mann–Whitney test) and SMC (3.3% vs 1.6%, $p = 0.02$, Mann–Whitney test) and an increase in the percentage of regenerative fibers (1.3 vs 5.7, $p = 0.001$, Mann–Whitney test) in DMD samples. There was also a trend for an increase in the number of satellite cells (0.9% vs 2.4%, $p = 0.07$) and of FAPs (5.6% vs 14.1%, $p = 0.07$). We validated these results using immunohistochemistry and immunofluorescence as it shown in supplementary Figs. 1 and 6. When assessing these differences between samples at the individual level, we observed that the proportion of cells in the muscle was similar among all controls, while there was greater variability in the DMD samples, which is compatible with an active process of muscle degeneration going through different stages (Fig. 2A and B). PCA analysis differentiated between DMD and controls by assessing the proportion of each cell population in the biopsy (Fig. 2C). The control patients were all closely located in the 2D dimension PCA map. Interestingly, the distribution of DMD patients on the PCA graph moved according to their clinical phenotype revealing two different subgroups, one that could be earlier in the process of muscle degeneration and included five samples from younger patients (2 to 3 years) with mildly affected muscle function and characterized by an increase in the number of nuclei from regenerative fibers (DMD-3 to DMD-7) and, another group of two samples from patients slightly older (4 years) who were already showing clear signs of muscles weakness and were more advanced in the process of muscle degeneration with a reduced number nuclei from slow and fast fibers and an increase in the nuclei from FAPs (DMD-1 and DMD-2).

Differential expression of genes in muscle fibers of patients with DMD

To gain insight in myofibers transcriptome in each cluster, we compared the gene expression profile of fast and slow muscle fibers nuclei between healthy controls and DMD patients (Fig. 3). Considering those genes with a higher $\log_2FC > 0.5$, we found 292 genes significantly upregulated and 85 downregulated in fast myofibers while in slow fibers we found 238 upregulated and 89 downregulated in DMD compared to controls. Fast and slow myofibers in DMD shared several genes in the top ten upregulated genes, such as *Myh3*, a characteristic marker of regenerative myofibers ($\log_2FC = 3.1$ and 2.1 respectively), *Meg3*, a LncRNA involved in myoblast plasticity and differentiation ($\log_2FC = 2.3$ and 2.9 respectively) and *Ldb3* which acts as an adapter in skeletal muscle to couple protein kinase C-mediated signaling via its LIM domains to the cytoskeleton ($\log_2FC = 1.6$ and 1.3 respectively) (Fig. 3A–D) [35, 36]. Interestingly, we observed an upregulation of genes involved in the transport of calcium (*Cacnas1*, *Ryr1*) and also, proteases such as *Capn3 and Capn2*, two pathways already known to be involved in the process of muscle degeneration in DMD [37]. Gene set enrichment analysis (GSEA) revealed several dysregulated molecular pathways (Fig. 3H). We observed an enrichment in both slow and fast fibers of the expression of genes involved in myogenesis, and muscle growth compatible with an active process of muscle regeneration, but also genes involved in axon guidance suggesting that new and regenerated muscle fibers release signals for the growth of terminal axons needed for reinnervation or, genes involved in adherens junction probably due to the need of new fibers to link again to the ECM. Interestingly, DMD fibers had a reduction in several metabolic pathways, when compared with control individuals especially oxidative phosphorylation but also glycolysis and lipid transport confirming previous observations [11]. Validation of these results using staining is displayed in supplementary Fig. 7. We observed a large number of fibers expressing embryonic myosin heavy chain and TNNT2 which are markers of regenerative fibers and, NCAM, a marker expressed by denervated fibers. Moreover, ITGB1 expression was increased in DMD samples. Additionally, we observed abnormalities in SDH/COX staining in many muscle fibers of the DMD patients, supporting an abnormal mitochondrial function (Supplementary Fig. 8).

In the case of regenerative myofibers, as they were just a minority in control patients (1.7%) and much more abundant in DMD (5.9%), we decided to analyze the genes increased in this

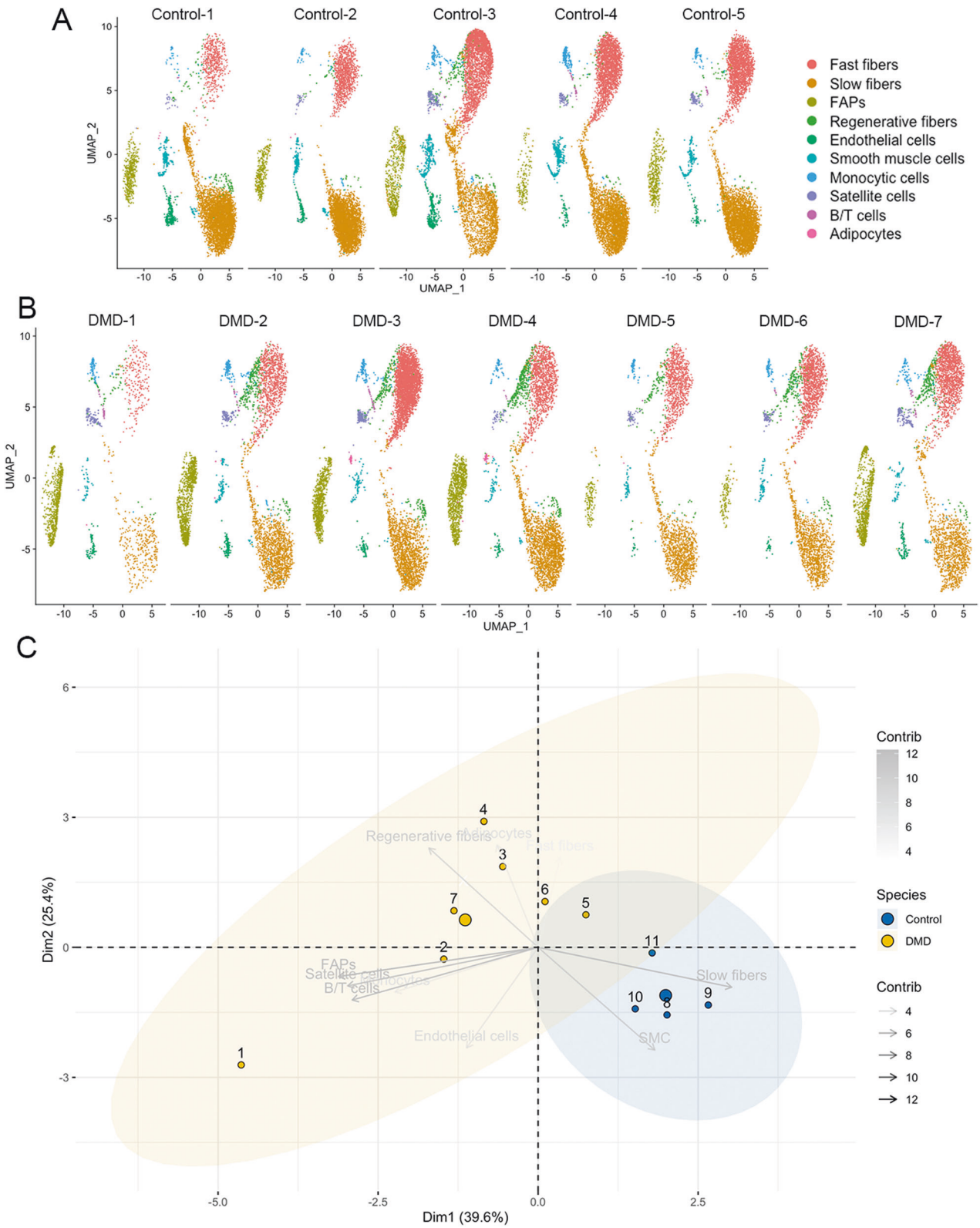
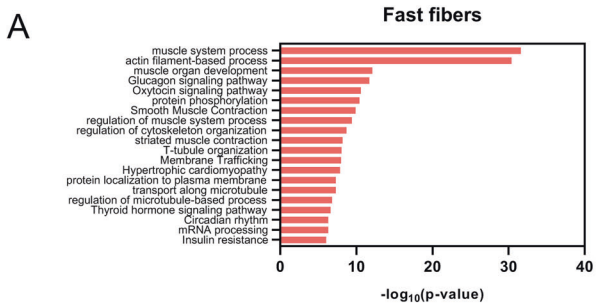


Fig. 2 Differences in cell population to the individual level. A UMAP visualization of nuclei from control individuals colored by cluster identity. **B** UMAP visualization of nuclei from DMD individuals colored by cluster identity. **C** PCA analysis showing the distribution of individuals based on each cell population proportion.

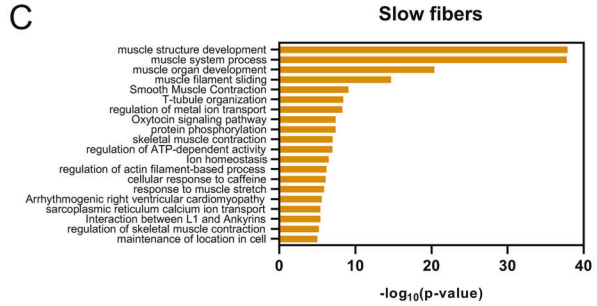
A



B

Gene	Protein	Potential function in muscle fibers	Fold increase
MYH3	Myosin Heavy Chain 3	Muscle embryonic myosin	3.07
MEG3	Maternally Expressed 3 (lncRNA)	Promotes myoblast differentiation	2.30
MYH8	Myosin Heavy Chain 8	Muscle fetal myosin	2.10
HNF1A-AS1	Hepatocyte nuclear factor 1 homeobox A antisense RNA 1 (lncRNA)	Unknown, NF-κβ-associated lncRNA	2.05
MYH1	Myosin Heavy Chain 1	Muscle contraction	2.01
ANKRD1	Ankyrin Repeat Domain 1	Myofibrillar stress and transcriptional regulation	1.98
EMC10	ER Membrane Protein Complex Subunit 10	Unknown, potentially involved in ER protein trafficking	1.99
ACTN3	Actinin Alpha 3	Myofibrillar protein, band Z	1.74
LDB3	LIM Domain Binding 3	Cytoskeletal assembly	1.68

C



D

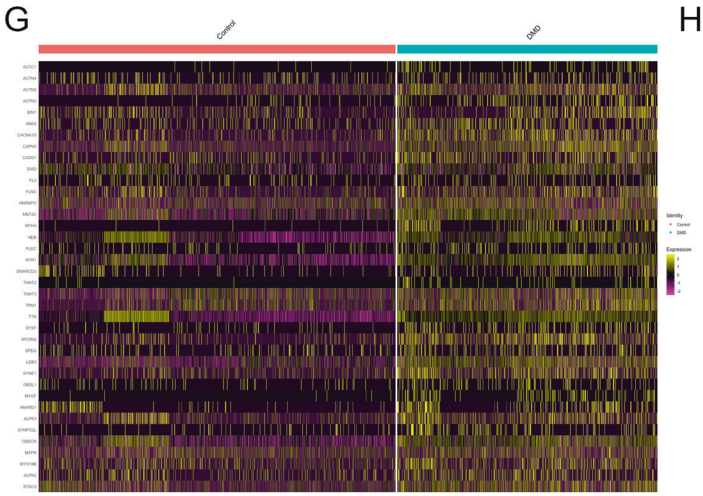
Gene	Protein	Potential function in muscle fibers	Fold increase
MEG3	Maternally Expressed 3 (lncRNA)	Promotes myoblast differentiation	2.95
CA3	Carbonic Anhydrase 3	Catalyzes carbon dioxide	2.05
HNF1A-AS1	Hepatocyte nuclear factor 1 homeobox A antisense RNA 1 (lncRNA)	Unknown, NF-κβ-associated lncRNA	1.97
MYH3	Myosin Heavy Chain 3	Muscle embryonic myosin	1.97
EMC10	ER Membrane Protein Complex Subunit 10	Unknown, potentially involved in ER protein trafficking	1.80
MEG8	Maternally Expressed 8 (lncRNA)	Unknown, involved in proliferation of smooth muscle cells	1.72
ATP1B4	ATPase Na+/K+ Transporting Family Member Beta 4	Hydrolyses ATP	1.41
LDB3	LIM Domain Binding 3	Cytoskeletal assembly	1.34
MYL1	Myosin Light Chain 1	Structural myofibrillar protein	1.33
MYH6	Myosin Heavy Chain 6	Structural myofibrillar protein	1.29

E

F

Gene	Protein	Potential function in muscle fibers	Fold increase
MYH8	Myosin Heavy Chain 8	Muscle fetal myosin	3.05
MYH3	Myosin Heavy Chain 3	Muscle embryonic myosin	2.88
TNNT2	Troponin T2	Cardiac and embryonic troponin-T	2.56
CALD1	Caldesmon 1	Actin linked regulatory protein	2.29
MEG3	Maternally Expressed 3 (lncRNA)	Promotes myoblast differentiation	1.99
ANKRD1	Ankyrin Repeat Domain 1	Myofibrillar stress and transcriptional regulation	1.86
RASSF4	Ras Association Domain Family Member 4	Regulator of the Hippo developmental pathway	1.76
CHRNG	Cholinergic receptor nicotinic gamma subunit	Neuromuscular junction component	1.73
MYOF	Myoferlin	Membrane fusion and repair	1.59
MACF1	Microtubule actin crosslinking factor 1	Microtubule, involved in synapsis differentiation	1.51

G



H

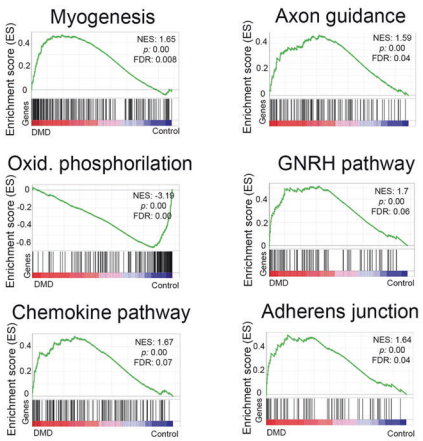


Fig. 3 Analysis of gene expression changes in DMD myonuclei compared to control myonuclei. A Top molecular pathways upregulated in fast fibers. **B** List of the top ten genes upregulated in myonuclei of fast fibers of DMD samples. **C** Top molecular pathways upregulated in slow fibers. **D** List of the top ten genes upregulated in myonuclei of slow fibers of DMD samples. **E** Top molecular pathways upregulated in regenerative myofibers. **F** List of the top ten genes upregulated in myonuclei of regenerative myofibers. **G** Heatmap showing expression of genes involved in muscle growth in Control and DMD fast and slow myonuclei. **H** GSEA plots showing enrichment score (ES) of the significantly enriched hallmark gene sets in fast and slow myonuclei. A positive value of ES indicates enriched in DMD and a negative value indicates enriched in normal conditions and down-regulated in DMD. GSEA gene set enrichment analysis, NES normalized enrichment score, FDR false discovery rate. Oxid Oxidative. GNRH Gonadotropin hormone-releasing hormone.

population compared to slow and fast muscle fibers of patients with DMD. We found 69 upregulated and 63 downregulated genes in regenerative fibers. As expected, among the top ten upregulated genes we identified many involved in the process of muscle regeneration such as *Myh3*, *Myh8*, *Tnnt2* or *Cald1* involved in the regeneration of the myofibrillar system, but also genes involved in membrane fusion (*Myof*), neuromuscular junction development (*Chrng* or *Macf1*) or Hippo pathway (*Rassf4*) involved in myoblast activation and muscle growth [38, 39].

As myogenesis and cell growth were one of the main pathways upregulated in muscle fibers, we further investigated the mechanism controlling these processes [40]. We analyzed several pathways and as shown in Supplementary Fig. 9, we observed an upregulation of the MEF2 family of transcription factors that have been classically involved in the myogenic program, especially *Mef2a* and *Mef2c*, while *Myf6* expression, an inhibitor of MEF2 was reduced in the muscle fibers [41]. Additionally, expression of *Tead1* and *Tead4*, the downstream effectors of the Hippo pathway involved in muscle hypertrophy, were also upregulated in DMD muscle fibers, associated with an increase in *Yap* and *Taz* expression two cofactors of this pathway [42]. In concordance with these findings, pro-atrophic genes such as *Murf1 (Trim63)* and *Atrogin-1* were downregulated in slow/fast muscle fibers. We did not observe an increase in the expression of genes belonging to the myostatin or insulin growth factor pathway in slow or fast myofibers.

FAPs from patients with DMD express genes related with cell proliferation and extracellular matrix remodeling

We compared the transcriptional profile of FAPs from healthy controls and DMD patients. DMD FAPs had a significant upregulation of 249 genes and a downregulation of 68 genes ($log_2FC > 0.5$) compared to controls. Among the top upregulated genes, we found genes encoding different types of collagens (*Col1a1, Col1a2, Col6a6, Col3a1*, or *Col21a1* among others), but also other components of the extracellular matrix such as elastin (*Eln*) and several fibulins. Genes encoding for proteins involved in extracellular matrix assembly such as *Sned1*, matrix remodeling (*Adamtsl1*), or interaction between cells and matrix such as laminins (*Lamb1* and *Lama4*) were also upregulated [43]. As many of the genes produced by DMD FAPs were components of the extracellular matrix, we compared the expression levels of matrix components between control and DMD samples and observed significant differences, not only in the expression levels but also in the components identified as shown in Supplementary Fig. 7. Apart of the extracellular matrix genes, we observed an upregulation of genes involved in relevant signaling pathways such as PDGF and NCAM signaling, tyrosine kinase activation and, Rho-GTPase cycle suggesting that FAPs are not a mere producer of extracellular matrix but they could also play a role as a potential regulator of the activity of other muscle resident cells (Fig. 4B and D). Several markers of fibrosis were increased in the muscle samples of patients with DMD compared to controls as shown in Supplemental Fig. 10, including collagen I and VI, CTGF and PDGF-AA expression as well as TE7, a marker of fibroblast.

Based on the myriad of biological processes upregulated in DMD samples, we decided to explore if there could be subpopulations of FAPs at different stages of differentiation that could be distinguished based on their gene expression profile. We re-clustered the FAP subgroup and identified seven different subclusters of cells as shown in Fig. 4C. These clusters shared the expression of many genes such as *Dcn*, *Pdgfra*, *Col1a1* or *Col3a1*, however some genes that were preferentially expressed in some of the clusters (Fig. 4F). For example, cluster 0 that was predominant in control samples expressed higher levels of the antiproliferative genes *Fos*(*p55*) and *Itih5*, but also *Col4a2* and *Lama2*. Cluster 4 was characterized by the expression of *Fbn1* while cluster 5 was characterized by the expression of *Lum* and could confirm the existence of this population of FAPs in DMD patients recently described by Rubinstein et al. [44]. Lum+ FAPs were the ones expressing the highest levels of collagen related genes such as *Col1a1* or *Col3a1*. Two of the clusters identified, cluster 3 and 6, were almost exclusively present in DMD samples and were characterized by the expression of genes involved in cell proliferation such as *Ahnak*, *Ccdc102b*, and *Podn* or *Parp14*, *Bod1l1* or *Smg1* respectively. Interestingly, cluster 6 was distinctively present in the patient that had the greatest decline in muscle function during follow up. Monocle analysis identified potential trajectories in the differentiation process of FAPs over time that started in Cluster 0, majority of controls, and end in Cluster 6 (Fig. 4D). Moreover we also identified genes differently expressed through the differentiation process (Fig. 4E).

As expected, the population of adipocytes, a type of cell known to derive from FAPs in the skeletal muscles, was higher in DMD samples than in controls. Further, we studied the molecular pathways activated in adipocytes based on their gene expression profile and observed that apart of pathways involved in lipid metabolism, adipocytes had an increased expression of genes of the Rho pathway and genes encoding for components of the basal lamina such as *Col4a1*, *Col4a2*, *Lama4* and, *Lamb1* (Supplementary Fig. 11).

Gene expression in different stages of disease progression

We were interested in investigating differences in the gene expression profile between DMD patients at different stages of disease progression. To do so, we reviewed the clinical and muscle function information present in the clinical notes and observed that there were consistent differences in clinical function at baseline and disease progression over the first four years after the muscle biopsy was obtained. As shown in Table 1, patients DMD 3, 5, 6, and 7 had mild muscle impairment at baseline and showed either stabilization or improvement in muscle function during follow-up period. On the other hand, patients DMD 1 and 2, showed a worse baseline performance and a decline in muscle function during the follow up period [45]. We explored if there were differences between control samples (Group A), stable patients (Group B) and declining patients (Group C) in the gene expression profile of muscle fibers and FAPs (Fig. 5). Muscle fibers from controls were enriched in the expression of genes such as *Pdk4 and Txnip* involved in the metabolism of glucose and lipids, *Linc-Pint* and *Btg2* inhibiting cell division and, *Trim63* (*Murf1*) involved in protein ubiquitination [46–48]. In the case of DMD patients, we did not observe many differences between Group B and C in the upregulated genes of, that were predominantly involved in muscle regeneration, either on satellite cell activation, membrane fusion, or sarcomere assembly (*Meg3*, *Meg8*, *Myh3, Cald1, Igfn1, Myof or Myo18B*). However, when we analyzed gene expression of FAPs among the three groups we did observe interesting results. As previously mentioned, control FAPs had a statically significant upregulation in the expression of antimitotic genes, such as *Fos* or *Btg2* compared to DMD FAPS. FAPs from Group B (DMD stable patients) had a statically significant upregulated expression of the proapoptotic gene *Xaf1*, interferon induced genes such as *Ifi44l* or *Mx2* or the profibrotic differentiation transcription factor *Spry1* while, FAPs from Group C (DMD declining patients) had the highest expression of collagen genes (*Col1a1*, *Col1a2*, *Col3a1*) but also high expression of genes actively involved cell division (*Ccdc80* or *Ccdc102b*), indicating that in the declining patients FAPs actively proliferate and express EXM components replacing the muscle fibers lost.

Communication between cells populations is dysregulated in DMD muscles

We studied the predicted intercellular communications of each cell population and compared the communication network between the control and DMD using CellChat package [49, 50].

A

Gene	Protein	Potential function	Fold increase
COL1A1	Collagen-I	Extracellular matrix protein	2.00
KCNQ1OT1	KCNQ1 overlapping transcript 1 (lncRNA)	Unknown, involved in tissue growth and development	1.87
COL6A6	Collagen-VI	Extracellular matrix protein	1.83
H19	H19 Imprinted Maternally Expressed Transcript (lncRNA)	Unknown, involved in cell differentiation	1.78
C3	Complement C3	Inflammation and adipogenic differentiation	1.74
HIF3A	Hypoxia Inducible Factor 3 Subunit Alpha	Potential role in angionenesis and adipogenesis	1.64
CCDC80	Coiled-Coil Domain Containing 80	Extracellular matrix organization	1.61
IFI44L	Interferon Induced Protein 44 Like	Unknown, involved in IFN inflammatory response	1.37
AHNAK	AHNAK nucleoprotein	Cell structure and migration	1.27
COL3A1	Collagen-III	Extracellular matrix protein	1.27

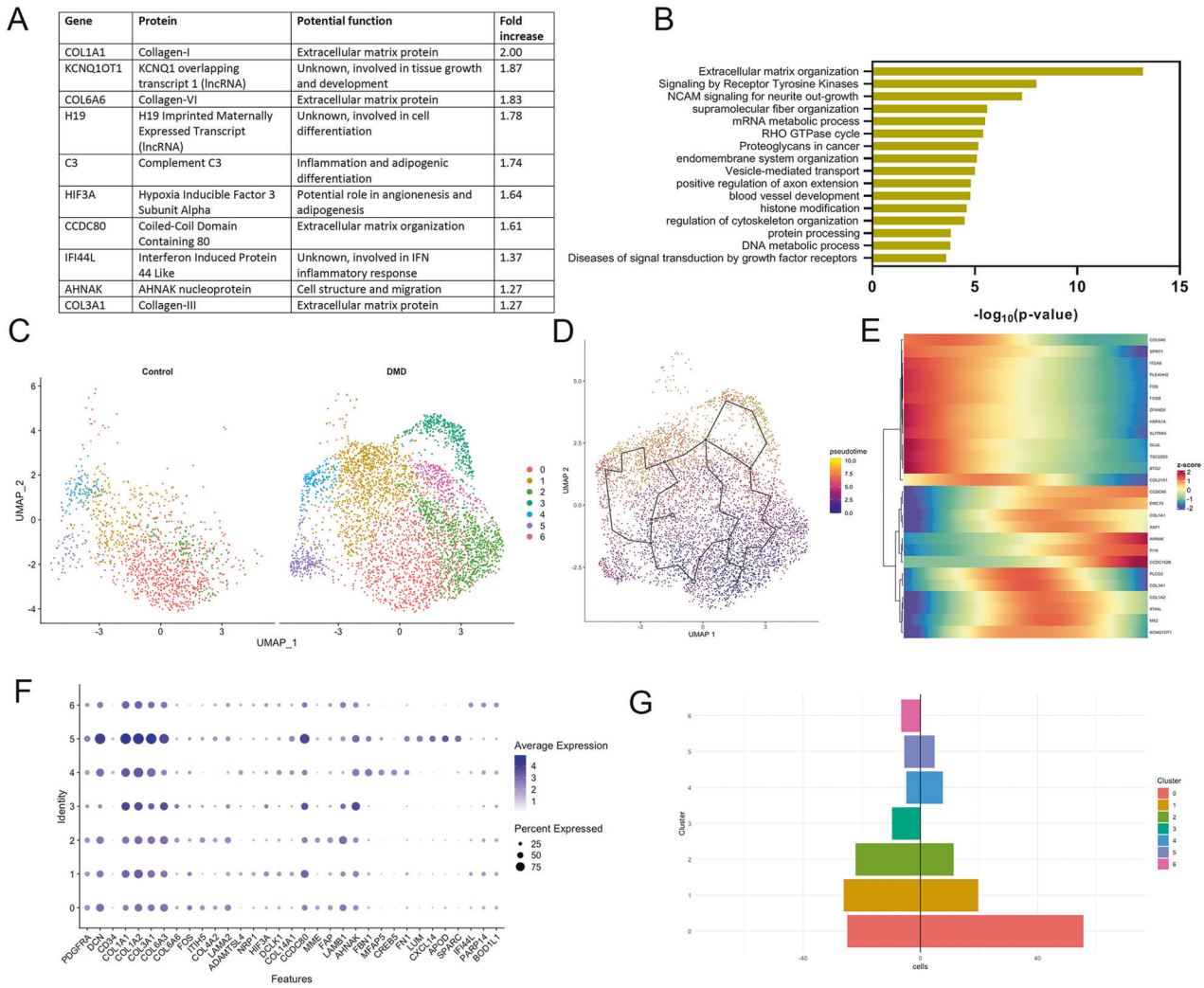


Fig. 4 Analysis of gene expression changes in DMD FAPs compared to control nuclei. A List of the top ten genes upregulated in FAPs of DMD samples. **B** Top molecular pathways upregulated in DMD FAPs. **C** UMAP visualization of nuclei from FAPs of control and DMD individuals coloured by subpopulation identity. **D** Monocle analysis showing pseudotime trajectories of the re-clustered FAPs. **E** Heatmap showing selected gene expression across pseudotime trajectories. **F** Selected genes expressed in each FAP subcluster. **G** Population of subclusters of FAPs in Control and DMD samples.

The analysis revealed significant differences in the number of cell interactions. As shown in Fig. 6A, B, FAPs and satellite cells became the most important source of ligands in DMD, potentially interacting with all other cell populations. Adipocytes, which were mainly present in DMD samples, played also an important role in cell-to-cell communication in DMD samples. CellChat detected 55 significant ligand-receptor pairs in the Control dataset and 61 in the DMD samples among the 11 nuclei clusters (Fig. 6C, D). A number of molecular pathways were identified exclusively in DMD samples such as cadherins (CDH), NCAM, major histocompatibility class- I, and neuroregulin, while others were exclusively identified in Control samples including CD40, CD80 or IL-2 among others. Signaling pathways upregulated in DMD samples were involved in several processes such as cell migration and remodeling of extracellular matrix (FGF, Collagen, Laminin), nerve growth and reinnervation (NCAM, NGF, and NPR2), and inflammation (MHC-I, CXCL, THBS). A detailed analysis of the expression levels of those ligand-receptor pairs that showed more significant differences between DMD and control samples can be found in Supplemental Fig. 12. As FAPs were identified as the main producer of ligands outgoing to other cell populations both in control and DMD (Fig. 6E, F) we decided to investigate further the main molecular signals released by these cells (Fig. 6G). Collagens and laminins were the most upregulated molecules signals produced by FAPs in DMD, followed by others such as members of the PDGF and FGF family, but also tenascin, thrombospondin, and fibronectin. As collagen and laminin were components of the ECM and could potentially influence all muscle cell behavior, we decided to study more precisely their potential communication network. Network centrality analysis of the inferred collagen identified FAPs as the most prominent sources of collagen either in control and in DMD samples, acting onto endothelial and smooth muscle cells in control, but also onto adipocytes in DMD (Fig. 7). Notably, among all known ligand-receptor pairs, DMD collagen signaling was mainly dominated by collagen I, IV and VI and its receptor *Itga1/Itga2* + *Itgb1*. FAPs were also the most prominent source of laminin either in control and DMD samples, although adipocytes became an important source as well in DMD (Fig. 8). In control samples, the laminin pathway was dominated by the *Lama2* and *Lamb1* ligands and its *Itga1/Itgb1* and *Itga7/Itgb1* receptors on endothelial and smooth muscles cells. In DMD, *Lama4* and *Lamb1* predominated acting through multiple *Itga/Itgb* receptors on adipocytes, regenerative fibers, and adipocytes in addition to smooth and endothelial cells.

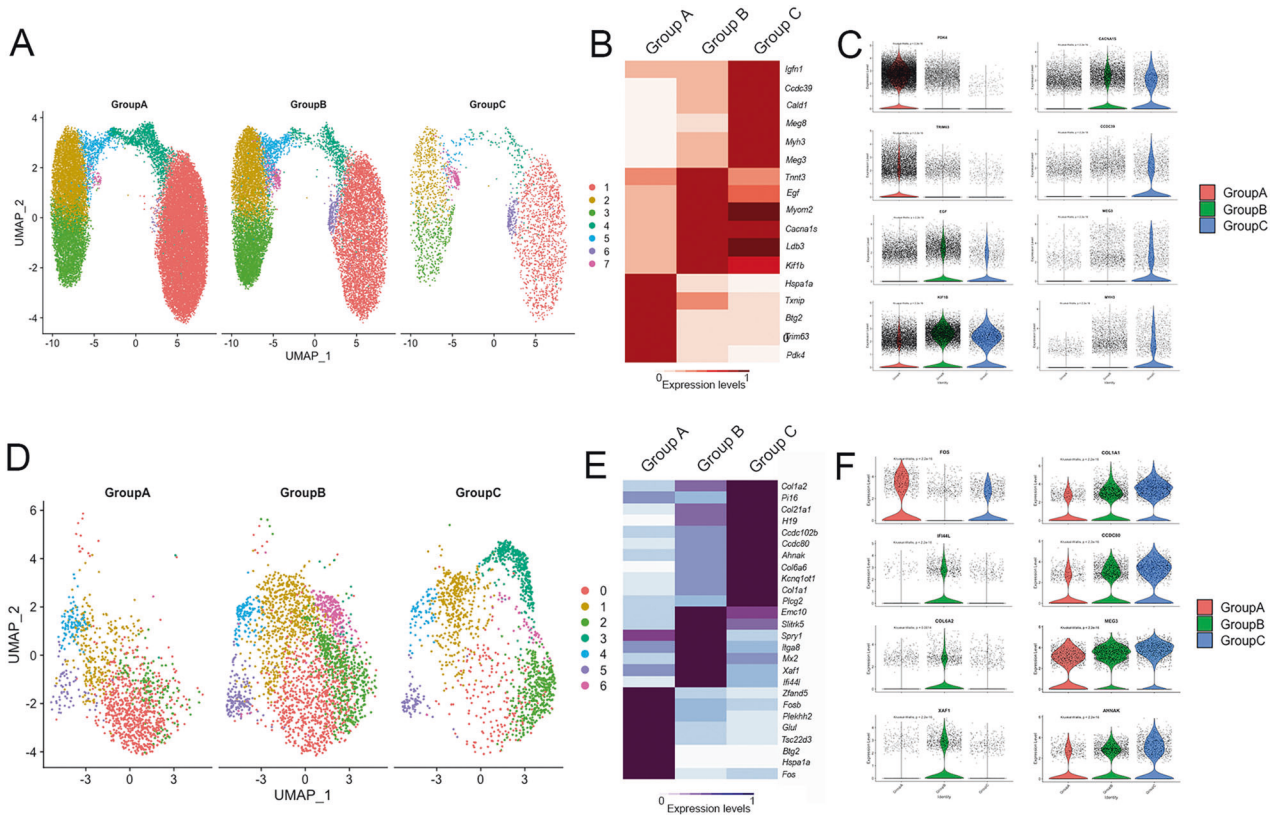


Fig. 5 Differences in cell population and gene expression profile between stable and progressing DMD patients. A UMAP of the subpopulations of myonuclei in control, stable, and declining DMD patients. **B** Heatmap showing the top upregulated genes in myonuclei from controls, stable and declining DMD patients samples. **C** Violin plot showing the expression of selected markers genes for myonuclei of controls, stable and declining DMD patients. **C** UMAP of the subpopulations of FAPs in control, stable, and declining DMD patients. **B** Heatmap showing the top upregulated genes in FAPs from controls, stable and declining DMD patients samples. **C** Violin plot showing the expression of selected markers genes for FAPs of controls, stable and declining DMD patients.

DISCUSSION

We have investigated the gene expression of cells present in muscle samples of DMD patients and age/gender matched controls. Our study has revealed novel and significant differences in the population of cells present in DMD muscles, their gene expression profile identifying several dysregulated molecular pathways, and changes in the intercellular signaling network. Our data show that these cellular and genetic modifications are dynamic through the disease's natural history impacting patient's muscle function and clinical progression.

One of the most relevant findings is the change in the proportion of each cell population in the muscle samples of DMD patients happening from the early stages of the disease. The most consistent change is the increase in the number of regenerative fibers that probably contribute to maintaining muscle function when there is not yet a massive loss of mature fibers, as observed in our patients who had mild muscle function impairment and who were clinically stable over four years after the biopsy was obtained. However, despite a regenerative gene signature was also observed in the muscle samples of the declining patients, they had reduced number of mature muscle fibers, an increase in the number of FAPs, and the presence of adipocytes. These data linking the proportion of cell types and muscle function suggest that the cellular changes observed in skeletal muscles of patients with DMD are identified since early ages and are dynamic over time influencing disease progression [51, 52]. Further studies analysing muscle samples from patients at later stages of disease progression are required to better understand the complexity of how the cell population changes over time and which molecular pathways are progressively activated. However, the availability of muscle biopsies at these later stages of disease progression is usually very restricted or absent because once the diagnosis is reached, muscle biopsies are not requested.

In concordance with an increased number of nuclei corresponding to regenerative fibers, we observed an upregulation of genes involved in myogenesis and muscle repair in myonuclei of slow and fast muscle fibers suggesting that at these early stages of the disease, injured muscle fibers are efficiently activating the regenerative machinery expressing genes coding for several developmental isoforms of sarcomeric proteins, molecules involved in linking new fibers to the extracellular matrix or molecules involved in the reorganization of the T-tubule system. Consequently, genes involved in muscle fiber growth and hypertrophy were upregulated in slow/fast myonuclei, while proatrophic factors were downregulated [53]. Interestingly, genes codifying components of the neuromuscular junction were also upregulated, including the fetal nAChR gamma subunit (*Chrng*) which expression is stopped at birth and substituted by postnatal epsilon subunit (*Chrne*), reinforcing the idea that neuromuscular junction is remodeled during muscle regeneration [54]. Moreover, genes involved in axon guidance such as *Ncam1*, *Gdnf*, and members of the semaforin family (*Sema3a* or *Sema4d*) were also increased suggesting that reinnervation is partially driven by signals released from the newly formed myofibers [53]. Although the regenerative signature was predominant in the myofibers, we also identified upregulation of genes involved in the process of protein degradation, such as overexpression of the proteases *Capn3* and *Capn2* and calcium channels *Ryr1* and *Cacnas1s*, two pathways suggested to be involved in the process of muscle degeneration in DMD [55, 56]. However, and in concordance with

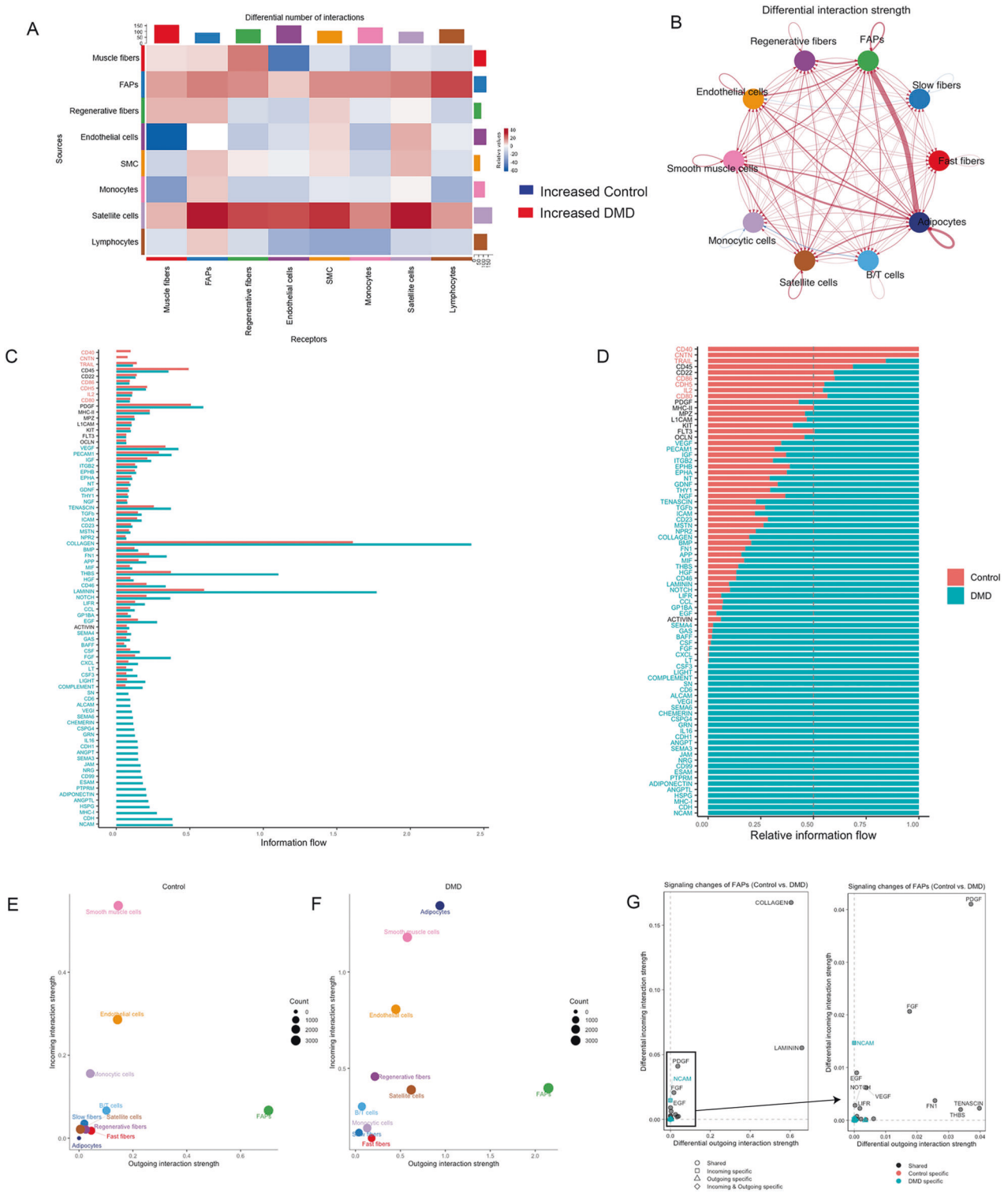


Fig. 6 Analysis of intercellular communications in control and DMD muscle samples. A Heatmap showing differential number of interactions between clusters in DMD samples compared to controls. Red: increased interactions in DMD. Blue: Increased interactions in controls. **B** Chord plot displaying intercellular ligand-receptor interaction strength comparing DMD and control samples. Red: increased interactions in DMD. Blue: Increased interactions in controls. **C** Bar graph showing relative information flow per each signaling path in control(red) and DMD (green) samples. **D** Bar graph showing the weight of each signaling path in control (red) and DMD (green) samples. **E** Dot plot showing the weight of each cell cluster in outgoing-incoming signaling in control samples. **F** Dot plot showing the weight of each cell cluster in outgoing-incoming signaling in control samples. **G** Dot plot showing the main molecules released and received by FAPs in DMD and control samples.

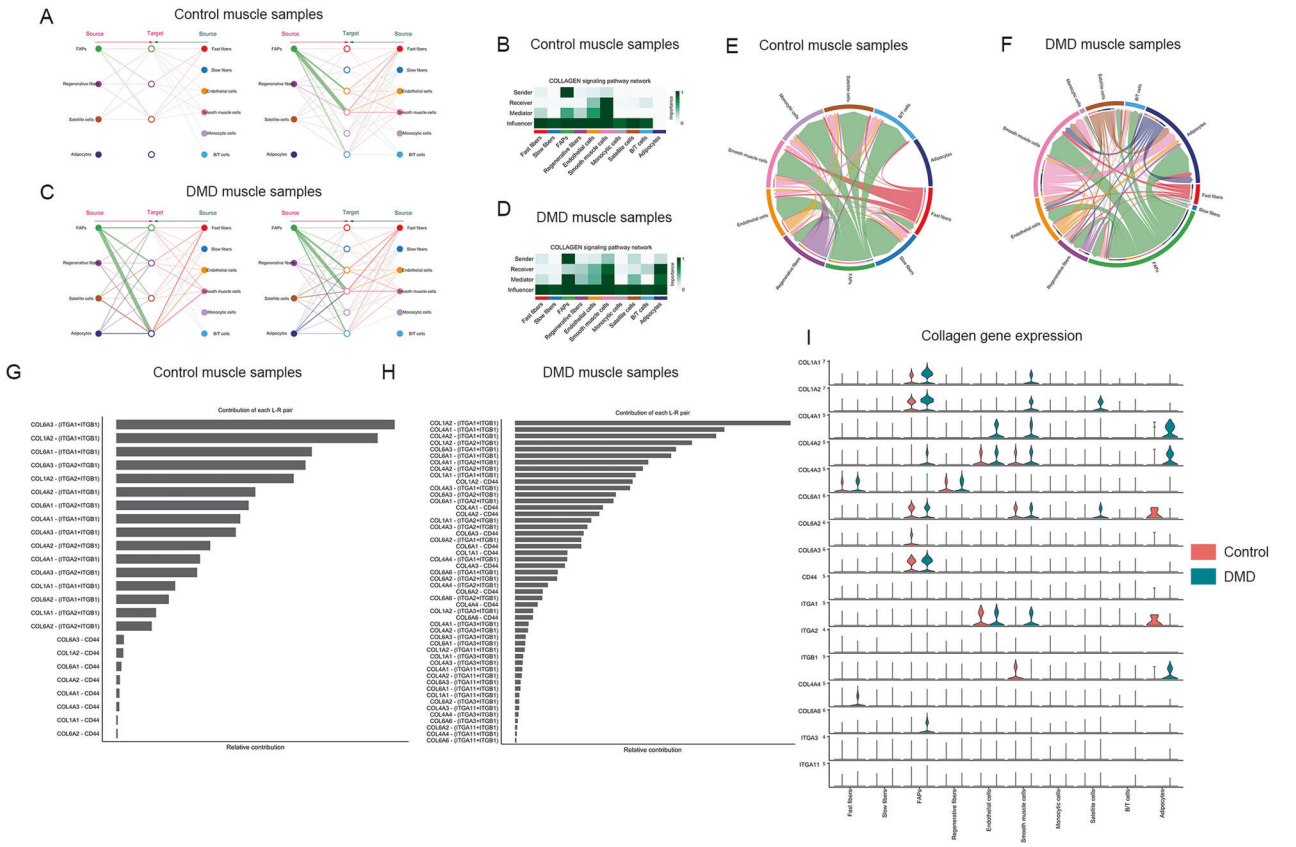


Fig. 7 Collagen signaling pathway in control and DMD muscle samples. A Hierarchical plot shows the inferred intercellular communication network for collagen signaling. This plot consists of two parts: Left and right portions highlight the autocrine and paracrine signaling to FAPs, regenerative fibers, satellite cells, and adipocytes and to the other clusters identified, respectively. Solid and open circles represent the source and target, respectively. Circle sizes are proportional to the number of cells in each cell group and edge width represents the communication probability. Edge colors are consistent with the signaling source. **B** Heatmap shows the relative importance of each cell group based on the computed network centrality measures of the collagen signaling network in control samples. **C** Hierarchical plot shows the inferred intercellular communication network for collagen signaling in DMD samples. **D** Heatmap shows the relative importance of each cell group based on the computed network centrality measures of the collagen signaling network in control samples. **E** Chord plot displaying intercellular communication network for collagen signaling in controls. **F** Chord plot displaying intercellular communication network for collagen signaling in DMD. **G** Relative contribution of each ligand-receptor pair to the overall communication network of a collagen signaling pathway in control samples, which is the ratio of the total communication probability of the inferred network of each ligand-receptor pair to that of the collagen signaling pathway. **H** Relative contribution of each ligand-receptor pair to the overall communication network of a collagen signaling pathway in DMD samples. **I** Violin plot showing the expression distribution of signaling genes involved in the inferred collagen signaling.

a cell that is actively growing, genes coding for enzymes involved in protein ubiquitination or atrophy were repressed, even though they have been reported to be increased in muscle fibers of the mdx mice [21]. These results agree with previously published data using bulk RNA analysis or proteomics of muscle biopsies of DMD patients at early stages of disease progression, showing a strong muscle regeneration signature and validating the results of our snRNAseq analysis [11, 12].

An increase in the number of FAPS was another prominent change observed in DMD patients. DMD-FAPs were characterized by an upregulation of the genes involved in the expansion and remodeling of the ECM including but not limited to collagen and laminin genes (*Col1a1*, *col6a6*, and *col3a1*), metalloproteinases (*Adamtsl1*) and molecules involved in the assembly of the extracellular matrix (*Sned1, Pcolce, Dcn*). Interestingly, we observed significant differences in the expression of genes codifying ECM components between DMD and control samples, suggesting that there are differences not only in the quantity of some of the components but also in the composition of the matrix with a substantial increase in collagen I, III and VI which are part of the interstitial matrix, while collagen IV, one of the components of the basal lamina remained stable [57, 58]. A complete understating of the impact that these changes have on muscle cells' behavior is lacking, but it is known that ECM apart from providing structural support to cells, also facilitates communication, regulates cell growth, promotes or restrict cell movement, and transmits mechanical signals [59]. We have identified different subpopulation of FAPs present in both control and DMD muscles, including the already described Lum+ and Fbn1+ cells, but not other FAPs populations described in murine models, such as the DPP4+ FAPs [44, 60]. Our analysis revealed a change in the predominant FAP subpopulations present in DMD and the existence of subpopulations that are not present in controls characterized by upregulation of genes involved in cell division. These subpopulations were mainly identified in patients with declining muscle function during disease progression suggesting that FAPs cell expansion could be a hallmark of muscle degeneration in DMD. However, a complete understanding of the potential role of these subpopulations of FAPs requires further characterization of the cells. Interestingly, FAPs were identified as the main messenger of signals either in control and DMD muscles by CellChat suggesting that FAPs are not a simple producer of ECM. The paracrine signals generated by FAPs targeted mostly endothelial and smooth muscle cells in control samples but also satellite cells, regenerative fibers, and adipocytes in DMD. The interaction between these cells is driven by many different molecules, but we have observed that collagens

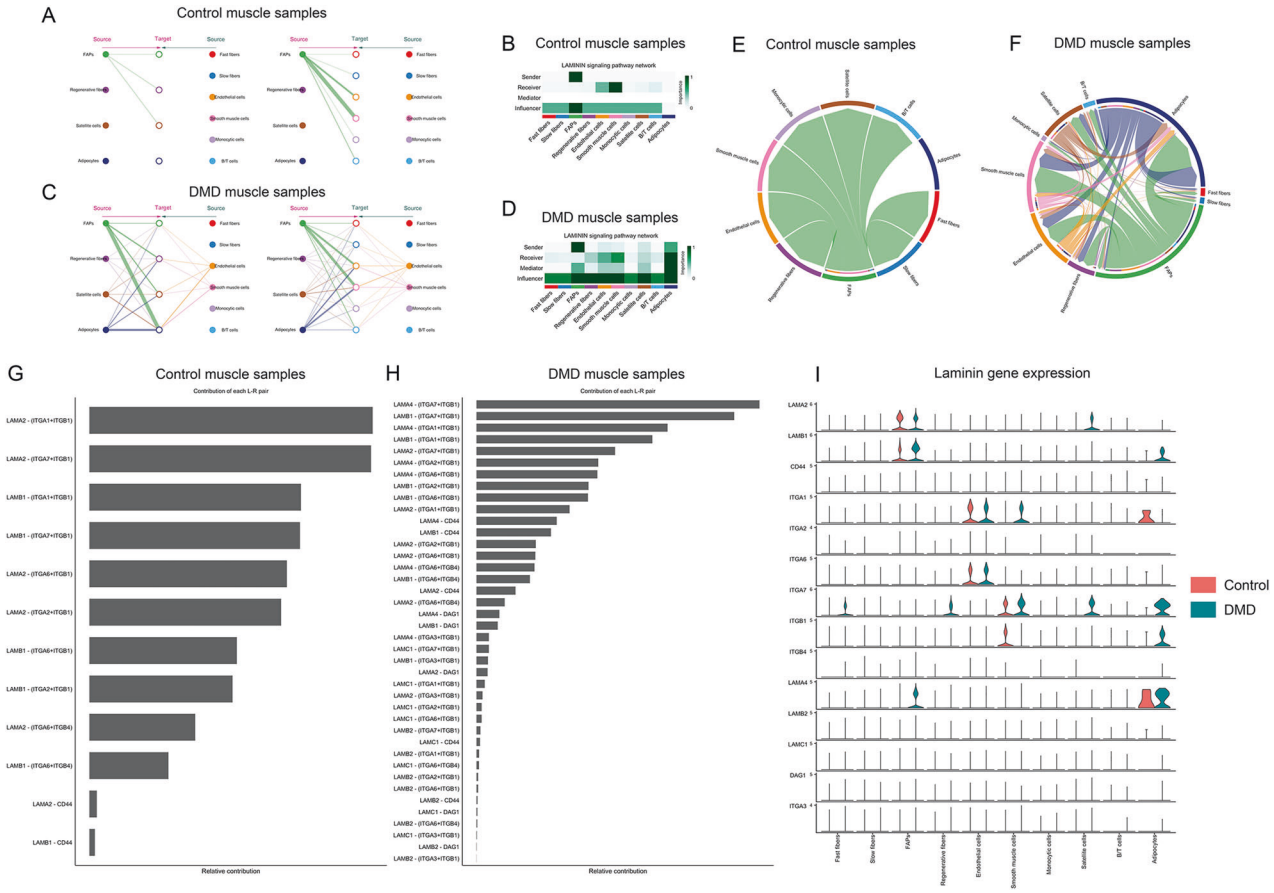


Fig. 8 Laminin signaling pathway in control and DMD muscle samples. A Hierarchical plot shows the inferred intercellular communication network for laminin signaling. This plot consists of two parts: Left and right portions highlight the autocrine and paracrine signaling to FAPs, regenerative fibers, satellite cells, and adipocytes and to the other clusters identified, respectively. Solid and open circles represent the source and target, respectively. Circle sizes are proportional to the number of cells in each cell group and edge width represents the communication probability. Edge colors are consistent with the signaling source. **B** Heatmap shows the relative importance of each cell group based on the computed network centrality measures of the laminin signaling network in control samples. **C** The hierarchical plot shows the inferred intercellular communication network for laminin signaling in DMD samples. **D** Heatmap shows the relative importance of each cell group based on the computed network centrality measures of the laminin signaling network in control samples. **E** Chord plot displaying intercellular communication network for laminin signaling in controls. **F** Chord plot displaying intercellular communication network for laminin signaling in DMD. **G** Relative contribution of each ligand-receptor pair to the overall communication network of a laminin signalling pathway in control samples, which is the ratio of the total communication probability of the inferred network of each ligand-receptor pair to that of the laminin pathway. **H** Relative contribution of each ligand-receptor pair to the overall communication network of a laminin signaling pathway in DMD samples. **I** Violin plot showing the expression distribution of signaling genes involved in the inferred laminin signaling.

and laminins could play an important role in intercellular communication driven by FAPs in the muscle. Adipocytes, which derive from muscle resident FAPS, irrupt in DMD muscle as an important player regulating cell signaling through the production of laminins contributing to the modified ECM. We have identified a substantial number of signaling pathways predominant or even only observed in DMD samples compared to control, somehow drawing a kind of DMD signaling fingerprint that helps to summarize the events that are taking place in these early stages of muscle degeneration. These events include ECM remodeling but also cell adhesion, migration, chemotaxis of cells, proliferation, differentiation, and inflammation.

The clinical features of the patients included in this study were homogeneous, as expected for patients at two, three, or four years of age when the muscle biopsy was obtained. However, muscle function revealed subtle differences between patients, and two groups were distinguishable, one characterized by mild impaired muscle performance at baseline and stability over time and another one with worse muscle function and deterioration after muscle biopsy. It is important to remark that the muscle biopsied was the quadriceps which is essential for the muscle function tests performed. When we compared differences in the gene expression profile between stable and progressive patients we did not observe major significant differences between myofibers that were characterized by a strong regenerative signature. However, we did observe that FAPs of the declining group were characterized by expression of genes involved in cell division and have an upregulation of genes encoding components of the ECM, compared to stable patients. This suggests that the existence of a population of active proliferative FAPs that could produce higher levels of collagen is key in the process of active muscle degeneration reinforcing the idea that treating patients with drugs inhibiting FAP proliferation or activation could be beneficial for DMD patients [61, 62].

To date, our understanding of the process of muscle degeneration in muscular dystrophies is mainly based on studies performed in murine models of the disease [63, 64]. These studies have provided valuable knowledge, although the results obtained have not been always validated in humans, mainly because of the lack of good animal models mimicking the process of muscle degeneration observed in patients. This is especially true in the case of DMD, where the existing murine models develop a milder

disease characterized by loss of muscle fibers and its replacement by fibrotic tissue only at the late stages of the disease, while there is almost no fat present in the muscles [65, 66]. The study of disease mechanisms in DMD using human samples can be complex because of the lack of available tissue especially since the popularization of molecular studies for diagnosis purposes [67].

Our work has some limitations. First, the number of samples from DMD patients was limited and we were not able to find muscle samples from the quadriceps of healthy controls of a similar age to the patients included in the study. Although there is not too much information about what are the factors that could influence muscle cells' gene expression, it is probable that age is one of these factors and therefore snRNAseq studies should try to have control samples from patients of similar ages. However, we performed a PCA analysis using average gene expression and could see that the samples of four out of five controls were closely located in the PCA map, suggesting that their gene expression is comparable (Supplementary Fig. 13). Second, we were not able to find enough muscle biopsies of DMD patients at different stages of disease progression to study how the gene expression profile is modified along disease progression, which should be investigated in further studies.

In summary, we have studied the gene expression profile to the single nuclei level in muscle samples of controls and DMD patients at an early stage of disease progression. We have focused our analysis on changes happening in muscle fibers and FAPs, as the two populations of cells show more changes in their number between controls and DMD. We have observed a substantial number of genes whose expression is dysregulated in both types of cells pointing towards an enhanced regenerative activity in DMD patients at this stage, associated with an increased proliferative activity of FAPS, which produce high levels of extracellular matrix components.

MATERIAL AND METHODS

Muscle specimens from healthy controls and patients with DMD

Muscle samples of boys with genetically confirmed DMD were obtained for diagnosis purposes or for research from patients seen at the Newcastle Hospital NHS Foundation Trust or at the Hospital Sant Joan de Deu Hospital in Barcelona. Muscle samples from controls were obtained from healthy children undergoing orthopedic surgery at Hospital Sant Joan de Deu in Barcelona. Patients' and controls' parents or legal representatives signed a consent form for the biopsy. Muscle samples were stored at the biobanks of each institution in liquid nitrogen tanks. The obtention of the biopsy and storage in the biobank was approved by the local Ethics Committee at both Institutions. The research study performed here with the samples was approved by the Ethics Committee of Newcastle University (reference 13866/2020).

Nuclei purification from human muscle biopsies

Frozen muscle biopsies (25 to 40 mg) were placed in a homogenization buffer (0.25 M sucrose and 1%BSA in Mg2+/Ca2+ -free, RNase-free PBS). Tissue was homogenized using a Tissue Ruptor II (Qiagen) and incubated for 5 min with 2.5% Triton-X100 (added at 1:6 ratio). The resulting homogenates were filtered through 100 µm and 70 µm strainers (Miltenyi Biotec), pelleted by centrifugation (3000 xg, 10 min at 4 °C), resuspended in sorting buffer (2% BSA/RNase-free PBS and 0.2 U/µl Protector RNase inhibitor (Roche)) and re-filtered through a 40 µm strainer. Then, the nuclei suspension was labelled with 10 µg/ml DAPI (Merck) and sorted (12-14,000 nuclei per sample) in a 96 well plate directly into 10X RT master mix (without RT Enzyme C) using a FACSAria™ Fusion Flow Cytometer (BD Biosciences). Then 8.3 µl of RT Enzyme C was added to each well.

10X Single nuclei RNA-seq

Samples were loaded into the Chromium controller (10X Genomics) for nuclei partition into Gel Bead-In-Emulsions (GEMs). cDNA sequencing libraries were prepared using the Next GEM Single Cell 3' Reagent Kits v3.1 (10X Genomics, 1000268), following manufacturer's instructions. Briefly, after GEM-RT clean up, cDNA was amplified during 12 cycles, and cDNA QC and quantification were performed on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies). cDNA libraries were indexed by PCR using the PN-1000215 Dual Index Kit TT, Set A plate. Size distribution and concentration of 3' cDNA libraries were verified on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies). Finally, sequencing of cDNA libraries was carried out on an Illumina NovaSeq 6000 using the following sequencing conditions: 28 bp (Read 1) + 8 bp (i7 index) + 0 bp (i5 index) + 89 bp (Read 2), to obtain approximately 20-30.000 reads per nucleus.

Tissue staining

Slices of the muscle samples from DMD and controls were studied using conventional staining protocol in the Department of Pathology of Hospital Sant Joan de Deu and Royal Victoria Infirmary Hospital in Newcastle including Hematoxilin-Eosin, Gomori Thrichromic, Fast and Slow myosis and Fetal myosin, Succinate dehydrogenase (SDH) and Cytochrome C-oxidase (COX) staining. We validated some of the results of the snRNAseq using immunofluorescence as previously described for the markers reported in Table 2 [62].

Bioinformatic analysis

Various R packages and software were used for the analysis of the samples. Seurat package (4.1.0) was used for the integration of the samples and unsupervised clustering [24]. Monocle-3 was used for trajectory analysis [68]. JavaGSEA was used for gene set enrichment analysis [69]. CellChat was used to study ligand-receptor characterization for cell-cell

Table 2. Antibodies used for validating the results.

Target	Clone	Host	Work dilution	Cell target
TE-7	TE-7	Mouse	1/80	Fibroblast
Collagen I	Polyclonal	Mouse	1/50	Extracellular matrix
Collagen VI	EPR17072	Rabbit	1/30	Extracellular matrix
Collagen III	EPR17673	Rabbit	1/15	Extracellular matrix
Laminin	4H8-2	Rat	1/5	Muscle fiber
CD31	PECAM1	Mouse	1/50	Endothelial cell
PDGFR alpha	Polyclonal	Goat	1/15	FAPs
CTGF	Polyclonal	Rabbit	1/10	Profibrotic cytokine
PDGF-AA	Polyclonal	Rabbit	1/10	Profibrotic cytokine
CD56	NCAM16.2	Mouse	1/50	Satellite cells / Regenerative fibers
MYH3	Polyclonal	Rabbit	1/100	Regenerative fibers
CD206	15-2	Mouse	1/50	M2 macrophages
TNNT2	1C11	Mouse	1/50	Regenerative fibers
ITGB1	12G10	Mouse	1/50	Muscle fiber

PDGFR platelet derived growth factor, *CTGF* connective tissue growth factor, *MYH3* Myosin heavy chain 3, *TNNT2* Troponin T2, *ITGB1* Integrin subunit beta 1.

communication prediction [49]. Raw and processed sequencing data are available under request to the corresponding author. For other detailed methods please refer to the supporting information.

Functional enrichment analysis

To reveal the precise biological properties of each cluster and in normal or pathological conditions, we used Metascape (http://metascape.org) to perform enrichment analysis including KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways, CORUM, WikiPathways and PANTER Pathway. Genes with a $\log_2$FC > 0.5 were analysed for each cluster and condition and all genes in the genome were used as the enrichment background.

Statistics

We confirmed that data on cell population did not follow normal distribution using Shapiro–Wilk test and therefore used nonparametric studies, specifically Mann–Whitney *U* test, to identify significant differences in cell population between samples. Comparison in gene expression between two or more groups was performed using Kruskal–Wallis test. The level of significance was set at p value < 0.05.

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Birnkrant DJ, Bushby K, Bann CM, Apkon SD, Blackwell A, Brumbaugh D, et al. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management. Lancet Neurol. 2018;17:251–67.
2. Mercuri E, Bonnemann CG, Muntoni F. Muscular dystrophies. Lancet .2019;394:2025–38.
3. Matthews E, Brassington R, Kuntzer T, Jichi F, Manzur AY. Corticosteroids for the treatment of Duchenne muscular dystrophy. Cochrane Database Syst Rev. 2016;2016:CD003725.
4. Bello L, Gordish-Dressman H, Morgenroth LP, Henricson EK, Duong T, Hoffman EP, et al. Prednisone/prednisolone and deflazacort regimens in the CINRG Duchenne Natural History Study. Neurology .2015;85:1048–55.
5. Verhaart IEC, Aartsma-Rus A. Therapeutic developments for Duchenne muscular dystrophy. Nat Rev Neurol. 2019;15:373–86.
6. Morin A, Stantzou A, Petrova ON, Hildyard J, Tensorer T, Matouk M, et al. Dystrophin myonuclear domain restoration governs treatment efficacy in dystrophic muscle. Proc Natl Acad Sci USA. 2023;120:e2206324120.
7. Morgan J, Partridge T. Skeletal muscle in health and disease. Dis Model Mech. 2020;13:dmm042192.
8. Cappellari O, Mantuano P, De Luca A. "The Social Network" and muscular dystrophies: the lesson learnt about the niche environment as a target for therapeutic strategies. Cells. 2020;9:1659.
9. Turk R, Sterrenburg E, de Meijer EJ, van Ommen GJ, den Dunnen JT, t Hoen PA. Muscle regeneration in dystrophin-deficient mdx mice studied by gene expression profiling. BMC Genom. 2005;6:98.
10. Szigyarto CA, Spitali P. Biomarkers of Duchenne muscular dystrophy: current findings. Degener Neurol Neuromuscul Dis. 2018;8:1–13.
11. Pescatori M, Broccolini A, Minetti C, Bertini E, Bruno C, D'Amico A, et al. Gene expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression. FASEB J. 2007;21:1210–26.
12. Haslett JN, Sanoudou D, Kho AT, Bennett RR, Greenberg SA, Kohane IS, et al. Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle. Proc Natl Acad Sci USA. 2002;99:15000–5.
13. Haslett JN, Sanoudou D, Kho AT, Han M, Bennett RR, Kohane IS, et al. Gene expression profiling of Duchenne muscular dystrophy skeletal muscle. Neurogenetics .2003;4:163–71.
14. Wolfien M, David R, Galow AM. Single-Cell RNA sequencing procedures and data analysis. In: Helder IN, editor. Bioinformatics. Brisbane (AU) 2021.
15. Deutsch A, Feng D, Pessin JE, Shinoda K. The impact of single-cell genomics on adipose tissue research. Int J Mol Sci. 2020;21:4773.
16. Petrany MJ, Swoboda CO, Sun C, Chetal K, Chen X, Weirauch MT, et al. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. Nat Commun. 2020;11:6374.
17. Giordani L, He GJ, Negroni E, Sakai H, Law JYC, Siu MM, et al. High-dimensional single-cell cartography reveals novel skeletal muscle-resident cell populations. Mol Cell. 2019;74:609–21.e6.
18. De Micheli AJ, Laurilliard EJ, Heinke CL, Ravichandran H, Fraczek P, Soueid-Baumgarten S, et al. Single-cell analysis of the muscle stem cell hierarchy identifies heterotypic communication signals involved in skeletal muscle regeneration. Cell Rep. 2020;30:3583–95.e5.
19. Dell'Orso S, Juan AH, Ko KD, Naz F, Perovanovic J, Gutierrez-Cruz G, et al. Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and regenerative conditions. Development .2019;146:dev174177.
20. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat Med. 2020;26:792–802.
21. Chemello F, Wang Z, Li H, McAnally JR, Liu N, Bassel-Duby R, et al. Degenerative and regenerative pathways underlying Duchenne muscular dystrophy revealed by single-nucleus RNA sequencing. Proc Natl Acad Sci USA. 2020;117:29691–701.
22. Scripture-Adams DD, Chesmore KN, Barthelemy F, Wang RT, Nieves-Rodriguez S, Wang DW, et al. Single nuclei transcriptomics of muscle reveals intra-muscular cell dynamics linked to dystrophin loss and rescue. Commun Biol. 2022;5:989.
23. Lin H, Ma X, Sun Y, Peng H, Wang Y, Thomas SS, et al. Decoding the transcriptome of denervated muscle at single-nucleus resolution. J Cachexia Sarcopenia Muscle. 2022;13:2102–17.
24. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell .2021;184:3573–87.e29.
25. Rasmussen M, Jin JP. Troponin variants as markers of skeletal muscle health and diseases. Front Physiol. 2021;12:747214.
26. Wei B, Jin JP. TNNT1, TNNT2, and TNNT3: Isoform genes, regulation, and structure-function relationships. Gene .2016;582:1–13.
27. Fuchtbauer EM, Westphal H. MyoD and myogenin are coexpressed in regenerating skeletal muscle of the mouse. Dev Dyn. 1992;193:34–9.
28. Cheng X, Li L, Shi G, Chen L, Fang C, Li M, et al. MEG3 promotes differentiation of porcine satellite cells by sponging miR-423-5p to relieve inhibiting effect on SRF. Cells. 2020;9:449.
29. Lin YT, Deel MD, Linardic CM. RASSF4 is required for skeletal muscle differentiation. Cell Biol Int. 2020;44:381–90.
30. Charvet B, Guiraud A, Malbouyres M, Zwolanek D, Guillon E, Bretaud S, et al. Knockdown of col22a1 gene in zebrafish induces a muscular dystrophy by disruption of the myotendinous junction. Development .2013;140:4602–13.
31. Baumeister A, Arber S, Caroni P. Accumulation of muscle ankyrin repeat protein transcript reveals local activation of primary myotube endcompartments during muscle morphogenesis. J Cell Biol. 1997;139:1231–42.
32. Pillay BA, Avery DT, Smart JM, Cole T, Choo S, Chan D, et al. Hematopoietic stem cell transplant effectively rescues lymphocyte differentiation and function in DOCK8-deficient patients. JCI Insight. 2019;5:e127527.
33. Torlakovic EE, Naresh K, Kremer M, van der Walt J, Hyjek E, Porwit A. Call for a European programme in external quality assurance for bone marrow immunohistochemistry; report of a European Bone Marrow Working Group pilot study. J Clin Pathol. 2009;62:547–51.
34. Orecchioni M, Ghosheh Y, Pramod AB, Ley K. Macrophage polarization: different gene signatures in M1(LPS+) vs. classically and M2(LPS-) vs. alternatively activated macrophages. Front Immunol. 2019;10:1084.
35. Dill TL, Carroll A, Pinheiro A, Gao J, Naya FJ. The long noncoding RNA Meg3 regulates myoblast plasticity and muscle regeneration through epithelial-mesenchymal transition. Development .2021;148:dev194027.
36. Yamashita Y, Matsuura T, Kurosaki T, Amakusa Y, Kinoshita M, Ibi T, et al. LDB3 splicing abnormalities are specific to skeletal muscles of patients with myotonic dystrophy type 1 and alter its PKC binding affinity. Neurobiol Dis. 2014;69:200–5.
37. Wallace GQ, McNally EM. Mechanisms of muscle degeneration, regeneration, and repair in the muscular dystrophies. Annu Rev Physiol. 2009;71:37–57.
38. Posey AD Jr, Demonbreun A, McNally EM. Ferlin proteins in myoblast fusion and muscle growth. Curr Top Dev Biol. 2011;96:203–30.
39. Vita GL, Polito F, Oteri R, Arrigo R, Ciranni AM, Musumeci O, et al. Hippo signaling pathway is altered in Duchenne muscular dystrophy. PLoS One. 2018;13:e0205514.
40. Schiaffino S, Reggiani C, Akimoto T, Blaauw B. Molecular Mechanisms of Skeletal Muscle Hypertrophy. J Neuromuscul Dis. 2021;8:169–83.
41. Liu N, Nelson BR, Bezprozvannaya S, Shelton JM, Richardson JA, Bassel-Duby R, et al. Requirement of MEF2A, C, and D for skeletal muscle regeneration. Proc Natl Acad Sci USA. 2014;111:4109–14.
42. Watt KI, Turner BJ, Hagg A, Zhang X, Davey JR, Qian H, et al. The Hippo pathway effector YAP is a critical regulator of skeletal muscle fibre size. Nat Commun. 2015;6:6048.
43. Vallet SD, Davis MN, Barque A, Thahab AH, Ricard-Blum S, Naba A. Computational and experimental characterization of the novel ECM glycoprotein SNED1 and prediction of its interactome. Biochem J. 2021;478:1413–34.

44. Rubenstein AB, Smith GR, Raue U, Begue G, Minchev K, Ruf-Zamojski F, et al. Single-cell transcriptional profiles in human skeletal muscle. Sci Rep. 2020;10:229.
45. Mazzone ES, Pane M, Sormani MP, Scalise R, Berardinelli A, Messina S, et al. 24 month longitudinal data in ambulant boys with Duchenne muscular dystrophy. PLoS One. 2013;8:e52512.
46. Zhang S, Hulver MW, McMillan RP, Cline MA, Gilbert ER. The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility. Nutr Metab (Lond). 2014;11:10.
47. He T, Yuan C, Zhao C. Long intragenic non-coding RNA p53-induced transcript (LINC-PINT) as a novel prognosis indicator and therapeutic target in cancer. Biomed Pharmacother. 2021;143:112127.
48. Kim SH, Jung IR, Hwang SS. Emerging role of anti-proliferative protein BTG1 and BTG2. BMB Rep. 2022;55:380–8.
49. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12:1088.
50. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17:159–62.
51. Wangkaew S, Suwansirikul S, Aroonrungwichian K, Kasitanon N, Louthrenoo W. The correlation of muscle biopsy scores with the clinical variables in idiopathic inflammatory myopathies. Open Rheumatol J. 2016;10:141–9.
52. Kinali M, Arechavala-Gomeza V, Cirak S, Glover A, Guglieri M, Feng L, et al. Muscle histology vs MRI in Duchenne muscular dystrophy. Neurology .2011;76:346–53.
53. Pratt SJP, Valencia AP, Le GK, Shah SB, Lovering RM. Pre- and postsynaptic changes in the neuromuscular junction in dystrophic mice. Front Physiol. 2015;6:252.
54. Mishina M, Takai T, Imoto K, Noda M, Takahashi T, Numa S, et al. Molecular distinction between fetal and adult forms of muscle acetylcholine receptor. Nature .1986;321:406–11.
55. Badalamente MA, Stracher A. Delay of muscle degeneration and necrosis in mdx mice by calpain inhibition. Muscle Nerve. 2000;23:106–11.
56. Zablocka B, Gorecki DC, Zablocki K. Disrupted calcium homeostasis in Duchenne muscular dystrophy: a common mechanism behind diverse consequences. Int J Mol Sci. 2021;22:11040.
57. Capitanio D, Moriggi M, Torretta E, Barbacini P, De Palma S, Vigano A, et al. Comparative proteomic analyses of Duchenne muscular dystrophy and Becker muscular dystrophy muscles: changes contributing to preserve muscle function in Becker muscular dystrophy patients. J Cachexia Sarcopenia Muscle. 2020;11:547–63.
58. Schuler SC, Liu Y, Dumontier S, Grandbois M, Le Moal E, Cornelison D, et al. Extracellular matrix: Brick and mortar in the skeletal muscle stem cell niche. Front Cell Dev Biol. 2022;10:1056523.
59. Thomas K, Engler AJ, Meyer GA. Extracellular matrix regulation in the muscle satellite cell niche. Connect Tissue Res. 2015;56:1–8.
60. Kim JH, Kang JS, Yoo K, Jeong J, Park I, Park JH, et al. Bap1/SMN axis in Dpp4+ skeletal muscle mesenchymal cells regulates the neuromuscular system. JCI Insight. 2022;7:e158380.
61. Alonso-Perez J, Carrasco-Rozas A, Borrell-Pages M, Fernandez-Simon E, Pinol-Jurado P, Badimon L, et al. Nintedanib reduces muscle fibrosis and improves muscle function of the alpha-sarcoglycan-deficient mice. Biomedicines .2022;10:2629.
62. Fernandez-Simon E, Suarez-Calvet X, Carrasco-Rozas A, Pinol-Jurado P, Lopez-Fernandez S, Pons G, et al. RhoA/ROCK2 signalling is enhanced by PDGF-AA in fibro-adipogenic progenitor cells: implications for Duchenne muscular dystrophy. J Cachexia Sarcopenia Muscle. 2022;13:1373–84.
63. Hogarth MW, Defour A, Lazarski C, Gallardo E, Diaz Manera J, Partridge TA, et al. Fibroadipogenic progenitors are responsible for muscle loss in limb girdle muscular dystrophy 2B. Nat Commun. 2019;10:2430.
64. Mazala DA, Novak JS, Hogarth MW, Nearing M, Adusumalli P, Tully CB, et al. TGF-beta-driven muscle degeneration and failed regeneration underlie disease onset in a DMD mouse model. JCI Insight. 2020;5:e135703.
65. Pessina P, Cabrera D, Morales MG, Riquelme CA, Gutierrez J, Serrano AL, et al. Novel and optimized strategies for inducing fibrosis in vivo: focus on Duchenne Muscular Dystrophy. Skelet Muscle. 2014;4:7.
66. van Putten M, Putker K, Overzier M, Adamzek WA, Pasteuning-Vuhman S, Plomp JJ, et al. Natural disease history of the D2-mdx mouse model for Duchenne muscular dystrophy. FASEB J. 2019;33:8110–24.
67. Salto-Tellez M, Gonzalez de Castro D. Next-generation sequencing: a change of paradigm in molecular diagnostic validation. J Pathol. 2014;234:5–10.
68. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566:496–502.
69. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.

ACKNOWLEDGEMENTS

We thank the Single Cell Genomics Group at the National Center of Genomic Analysis (CNAG, Barcelona, Spain), specially to Holger Hein, Ginevra Caratu and Domenica Marchese for their technical assistance. We also thank Rachel Queen and Adrienne Unsworth for their help in the analysis of the bioinformatic data.

AUTHOR CONTRIBUTIONS

XS-C designed and performed experiments, obtained funding, wrote and reviewed the paper. EF-S designed and performed experiments, wrote and reviewed the paper. DN collected clinical data, wrote and reviewed the paper. CJ provided and reviewed muscle biopsies, performed staining on the muscle biopsies, wrote and reviewed the paper. PP-J performed staining on the muscles, reviewed the paper. EV performed staining on muscle samples, reviewed the paper. CO collected clinical data, wrote and reviewed the paper. AM performed bioinformatic analysis and wrote paper. MS collected clinical data, performed statistical analys, reviewed the paper. AC collected muscle samples, processed muscle samples and isolated nuclei, reviewed the paper. JV-D performed bioinformatic analysis, reviewed the paper. JC analysis of the data, extracellular matrix analysis in muscle samples, reviewed the paper. ZL, PM collected muscle biopsies, processed muscle biopsies, reviewed the paper. RG-N supervised bioinformatic analysis, designed pipelines for the analysis, wrote and reviewed the paper. JA-P collected and analysed clinical data. CM-B, GT collected clinical data, provided muscle samples, reviewed the paper. VS, MG collected clinical data, provided muscle samples, reviewed the paper. AN designed experiments, collected clinical data, provided muscle samples, critical analysis of the data, wrote and reviewed the paper. JD-M designed the experiments, collected clinical and pathology data, reviewed muscle biopsy data, analyzed results, obtained funding, wrote and reviewed the paper.

FUNDING

This work has been funded by grants from Academy of Medical Sciences Professorship Scheme (APR4/1007), Medical Research Council (MR/W019086/1) and Fundación Isabel Gemio to Jordi Díaz-Manera and a grant from La Marató de TV3 (#202034-10) to Xavier Suárez-Calvet.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41419-023-06103-5.

Correspondence and requests for materials should be addressed to Jordi Diaz-Manera.

Reprints and permission information is available at http://www.nature.com/reprints

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