Elucidating DUX4-mediated molecular mechanisms underlying FSHD pathophysiology using multiomics approaches

Zheng, D.

Chapter 2

snRNA-seq analysis in multinucleated myogenic FSHD cells identifies heterogeneous FSHD transcriptome signatures associated with embryonic-like program activation and oxidative stress-induced apoptosis

Dongxu Zheng[1], Annelot Wondergem[1], Susan Kloet[1], Iris Willemsen[1], Judit Balog[1], Stephen J. Tapscott[2], Ahmed Mahfouz[1,3], Anita van den Heuvel[1] & Silvère M. van der Maarel[1]

[1]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

[2]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, USA

[3]Delft Bioinformatics Lab, Delft University of Technology, Delft, Netherlands

Human Molecular Genetics. 2024 Feb. 1; 33(3):284-298.

https://doi.org/10.1093/hmg/ddad186

Abstract

The sporadic nature of DUX4 expression in FSHD muscle challenges comparative transcriptome analyses between FSHD and control samples. A variety of DUX4 and FSHD-associated transcriptional changes have been identified, but bulk RNA-seq strategies prohibit comprehensive analysis of their spatiotemporal relation, interdependence and role in the disease process. In this study, we used single-nucleus RNA-sequencing of nuclei isolated from patient- and control-derived multinucleated primary myotubes to investigate the cellular heterogeneity in FSHD. Taking advantage of the increased resolution in snRNA-sequencing of fully differentiated myotubes, two distinct populations of DUX4-affected nuclei could be defined by their transcriptional profiles. Our data provides insights into the differences between these two populations and suggests heterogeneity in two well-known FSHD-associated transcriptional aberrations: increased oxidative stress and inhibition of myogenic differentiation. Additionally, we provide evidence that DUX4-affected nuclei share transcriptome features with early embryonic cells beyond the well-described cleavage stage, progressing into the 8-cell and blastocyst stages. Altogether, our data suggests that the FSHD transcriptional profile is defined by a mixture of individual and sometimes mutually exclusive DUX4-induced responses and cellular state-dependent downstream effects.

Introduction

With an estimated prevalence of 1:8,500-1:15,000 facioscapulohumeral dystrophy (FSHD) (OMIM: 158900) is one of the most common inherited forms of muscular dystrophy. The disease is characterized by progressive and often asymmetric muscle atrophy, dystrophy and wasting of the skeletal muscles of the face, shoulders, and upper limbs. With disease progression, other muscles can also be affected. The molecular pathology of FSHD is defined by misexpression of the retrogene *DUX4* (OMIM: 606009) in skeletal muscle. *DUX4* encodes a germline and cleavage-stage transcription factor of which a copy of its open reading frame is located within every unit of the D4Z4 macrosatellite repeat array at the distal end of chromosome 4q (4q35) (1, 2). This macrosatellite repeat array is normally epigenetically silenced in most somatic tissues, while in FSHD this silencing is incomplete resulting in the de-repression of the *DUX4* locus in skeletal muscle (2-5). Misexpression of DUX4 in skeletal muscle leads to a wide range of downstream events (6-18), but the exact mechanism behind DUX4-induced toxicity (i.e., the spatiotemporal relationship between these events, the order of the events, their interdependency or mutual exclusivity) is still largely unknown. Besides, DUX4 expression in FSHD muscle is likely sporadic. In tissue culture, only 1/1,000 myoblasts or 1/200 myotube nuclei in patient-derived primary myogenic cell cultures show DUX4 protein by immunofluorescence (19, 20). *In vivo*, MRI-based imaging analyses of muscle have shown intramuscular heterogeneity with focal signs of disease activity (21). RT-qPCR analysis identifies DUX4 transcripts in approximately 50% of FSHD muscle biopsies (19). Bulk RNA sequencing analyses on FSHD muscle biopsies, however, have only been able to detect DUX4 in a few biopsies, possibly because of mapping issues. Therefore, for RNA-seq studies, DUX4 target gene expression is often used as signature of DUX4 activity, which is detected in approximately 60% of all FSHD biopsies (22, 23). Signs of DUX4 activity (i.e., DUX4 target gene expression) are more likely to be detected in regions of active disease defined by increased fat fraction and STIR-positivity, which is considered a marker for inflammation (22, 24). Besides, a recent proximity ligation study in a single FSHD muscle biopsy supports the interpretation of sporadic DUX4 expression (25). Collectively, these observations provide evidence for sporadic bursts of DUX4 expression in FSHD muscle. This sporadic nature of DUX4 generates a characteristic large heterogeneity in the FSHD-associated transcriptome signature, which is largely masked in bulk RNA-sequencing by the majority of DUX4-negative muscle and non-muscle cells in the tissue.

We and others have previously reported on single-cell RNA-sequencing (scRNA-seq) studies in mononuclear myocytes and myonuclei derived from myotube cultures of FSHD patients, demonstrating the cellular heterogeneity and describing the first modeling of transcriptional dynamics during FSHD cellular progression (26, 27). Though, with the limited number of myogenic cells expressing DUX4 (and DUX4 target genes), our insights into the transcriptional dynamics of FSHD-associated pathways remain limited. As DUX4 expression is known to significantly increase during myogenic differentiation (*in vitro*) (28), enriching for late-myogenic multinucleated myotubes most prone to DUX4 activation, and purifying them from the majority of contaminating mononuclear early myoblasts and myocytes, should significantly increase the resolution of FSHD-associated transcriptome changes (27). Where size restrictions exclude multinucleated myotubes from most single-cell RNA-sequencing technologies, single-

nucleus RNA-sequencing (snRNA-seq) overcomes this limitation by analyzing all individual nuclei within the myotube.

A recent single-nucleus RNA-sequencing study in FSHD2 myotube cell lines revealed the presence of two distinct FSHD-associated nuclei populations defined by a high or low level of DUX4 target gene expression (27). Although this study increases the resolution on DUX4 and DUX4-induced heterogeneity, the interpretation of the presence of two distinct nuclei populations, e.g., whether both populations represent two different stages in FSHD cellular progression (i.e., early and late) or represent two distinct FSHD-induced responses, and how both responses would relate to each other and DUX4's role in early development, remains a challenge.

In this study, we performed snRNA-sequencing on nuclei from three or four days differentiated multinuclear myotubes derived from three FSHD patients and one healthy donor and analyzed both global and DUX4-specific transcriptional differences in FSHD and control nuclei. Focusing specifically on late-myogenic nuclei allowed us to detect DUX4 (-target) gene activity in up to 8.8% of nuclei in FSHD samples. We identify two DUX4-affected populations, both showing distinct transcriptional profiles. Our data provides insights into the differences between these two populations, suggesting heterogeneity in (and possibly mutual exclusivity for) two well-known FSHD-associated transcriptional aberrations: increased oxidative stress and inhibition of myogenic differentiation. Finally, using publicly available embryonic scRNA-seq data we provide evidence that DUX4-affected nuclei share transcriptome features with early embryonic cells beyond the well-described cleavage stage. This suggests that DUX4's natural role in early embryogenesis is re-activated in muscle, a role that may be incompatible with myogenic differentiation, causing cytotoxicity. Overall, our research improves the resolution of cellular heterogeneity in FSHD and sheds light on how an early embryonic transcription factor like DUX4 becomes toxic when re-activated in muscle cells.

Results

Single-nucleus RNA-seq on multinucleated myotubes by size exclusion filtering

We performed snRNA-seq on differentiated multinucleated myotubes of three FSHD patient-derived (FSHD-01 - FSHD-03) and one control donor-derived (CTRL-01) primary muscle cell cultures (See Table 1 for cell line information, Fig. 1A for experimental design). All four cell lines show comparable differentiation speed and myotube morphology (see Fig. S1 for representative microscopy images and fusion index calculations). For all three FSHD cell lines, DUX4-positive nuclei can be identified upon differentiation towards multinucleated myotubes. We used size exclusion to separate the multinuclear myotubes from mononuclear muscle cells and validated this size exclusion filtering by performing RT-qPCR analysis for three myogenic markers on an aliquot of the separated cell populations. The flow-through cell suspension showed enriched expression of the early myogenic marker *MYF5*, whereas cells blocked by the filter showed increased levels of the late-myogenic markers *MYOG* and *MYH3* (Fig. 1B),

indicating efficient separation of the early and late myogenic cells. Next, the cell membranes of the late myogenic cells were lysed and nuclei were purified from large cell membrane debris by a second size exclusion filtering step. The purity of the final nuclei samples was validated

in three ways. First, RT-qPCR analysis of the ratio of unspliced pre-mRNA of *RPL10a* over its spliced mRNA showed an increased ratio of the nuclear-enriched unspliced *RPL10a* mRNA in the nuclei suspensions compared to whole cells (Fig. 1C). Second, in the nuclear fractions RT-qPCR analysis of the nuclear-enriched *XIST* RNA (in the case of female samples) showed a strong enrichment over the non-nuclear-enriched mRNAs *UBC* or *GUSB* (Fig. 1C). Finally, western blot analysis showed the absence of contaminating cytoplasmic acetylated α-tubulin in the nuclei fractions, indicating an efficient purification of nuclei from cytoplasmic cell debris (Fig. 1D).

The purified nuclei fractions of all four samples were prepared for snRNA-seq using the 10X Genomics Chromium platform. After quality control and filtering (see Methods), we obtained between 1,427-4,698 nuclei per sample, detecting a median of 984-1,599 genes per nucleus with a median unique read count of 1,669-2,984 reads (Table 1).

Single-nucleus RNA-seq on multinucleated myotubes allows for increased DUX4-affected nuclei detection in FSHD samples

To further assess the quality of our snRNA-seq data, we analyzed the expression of *DUX4* and its target genes in the four samples. Due to the low and highly sporadic nature of both DUX4 and DUX4 target gene expression, we used the cumulative read count of a set of 67 previously described DUX4 target genes (23, 26) as biomarker for DUX4 activation [a signature for DUX4 activity previously used in van den Heuvel et al. (26)]. DUX4-positive (DUX4^{pos}) and DUX4 target-positive nuclei [DUX4 target^{pos}; defined as expressing ≥5 of the DUX4 target genes (23, 26)] were exclusively detected in FSHD samples (Fig. 1E, Fig. S2A-B, Table 1). In total, 41 nuclei (0.79%) from the three FSHD samples expressed detectable levels of DUX4 (read count ≥1), compared to none in the control sample. In addition, we identified 64-137 DUX4 target^{pos} nuclei in each of the three FSHD samples (totaling 311/5,175 (6%) DUX4 target^{pos} nuclei). Overall, 341 nuclei (3.8%-8.8% of all nuclei in the individual FSHD samples) were considered DUX4-affected, being either DUX4^{pos}, DUX4 target^{pos} or both. The percentage of DUX4(-target)^{pos} nuclei is higher than previously reported (26, 27), highlighting the benefit of our size exclusion-based multinucleated myotube-nuclei enrichment prior to snRNA-seq.

To enable a systematic comparison between FSHD and control (CTRL) nuclei, we combined all four snRNA-seq datasets into one single integrated dataset (Fig. 1F and Fig. S2C). Differential gene expression (DE) analysis between all 341 DUX4-affected nuclei and the remaining 9,532 non-affected nuclei, identified 914 upregulated genes in DUX4-affected nuclei (including 27/67 DUX4 target genes used to classify the nuclei, among which previously described core biomarkers of DUX4 activation (*TRIM43*, *LEUTX* and *PRAMEF2*) (23) and 703 downregulated genes (adjusted *p*-value <0.05 and absolute fold change of ≥1.5; Fig. S2D and Table S1). The DE gene set significantly overlapped with our previously identified DUX4-affected DE gene set in scRNA-seq data from primary myocytes (26) (Fig. S2E). Comparison with previously published gene sets from bulk RNA-sequencing analyses showed overlapping results [i.e., 110/570 (19.3%) genes from Rickard et al. (29), 34/213 (15%) genes from Yao et al. (23) and 185/626 (29.6%) genes from Jaganathan et al. (30), see Table S1]. Interestingly, we also identified 1,242 potential novel DUX4 target genes (593 upregulated and 649 downregulated genes), which may be interesting targets for future research. Gene set enrichment analysis (GSEA) of GO terms for biological processes

indicated that the upregulated genes in DUX4-affected nuclei were involved in RNA and protein metabolism (e.g., RNA spliceosome factors, ribosome/ribonucleoprotein complex factors) processes that have previously been shown to be affected in ectopic *DUX4* expression and reporter systems, as well as in previous scRNA-seq and snRNA-seq studies (Fig. 1G-H and Table S2) (23, 26, 27, 29, 30). The downregulated genes were active in muscle development and muscle cell differentiation (e.g., *MEF2C*, several myosins (*MYH3/6/9*), *MYOD1* and *MYOG*), which corresponds to DUX4's described inhibition of myogenic differentiation (8).

As FSHD is associated with reduced PAX7 activity (captured in the PAX7 score (24, 31, 32)), we next calculated the PAX7 score for each nucleus. We found a small but significant overall reduction in PAX7 score in FSHD versus control nuclei (Fig. S3A). This reduction no longer reached statistical significance when specifically comparing DUX4-pos versus DUX4-neg (data not shown) or DUX4-affected versus DUX4-non-affected nuclei in our data (Fig. S3B). Furthermore, we did not observe a (linear) negative correlation between DUX4 signature expression and PAX7 score (Fig. S3C), suggesting (partially) independent involvement of both factors. Though, it is important to note that the calculated PAX7 scores might be biased by the high sparsity of snRNA-seq and the limited number of PAX7 target genes being detected in the data (Fig. S3D).

Altogether, these results suggest that snRNA-seq on multinucleated myotube-enriched FSHD samples captures increased numbers of nuclei with described DUX4 and/or FSHD-associated transcriptome signatures.

Table 1 Summary metrics

Sample ID	**Repeat unit information**	**CTRL/FSHD**	**Sex**	**Nuclei count** **before filtering** *per sample*	**Thresholds for QC filtering**	**Nuclei count** **after filtering** *per sample*	**Total genes detected** *per sample*	**Median gene count** *per nucleus*	**Median UMI count** *per nucleus*	**Nuclei count: all DUX4-affected** *[% of all nuclei]*	**Nuclei count:** **DUX4-pos only** *[% of DUX4-affected]*	**Nuclei count: DUX4 & DUX4 target-pos** *[% of DUX4-affected]*	**Nuclei count: DUX4 target-pos only** *[% of DUX4-affected]*
										DUX4 expression > 0 AND / OR ≥ 5 targets expressed	**DUX4 expression > 0**	**DUX4 expression > 0 AND ≥ 5 targets expressed**	**≥ 5 targets expression**
FSHD-01	4qA (6U) / 4qB	FSHD1	Female	1,987	200 ≤ nFeatures ≤ 3,500; %MT-genes ≤ 12; %Ribo-genes ≤ 5	1,926	18,238	1,332	2,877	147 [7.6%]	10 [6.8%]	5 [3.4%]	132 [89.8%]
FSHD-02	4qA (3U) / 4qB	FSHD1	Male	1,493	200 ≤ nFeatures ≤ 3,500; %MT-genes ≤ 11; %Ribo-genes ≤ 5	1,427	17,326	1,24	2,738	125 [8.8%]	15 [12.0%]	5 [4.0%]	105 [84.0%]
FSHD-03	4qA (3U) / 4qA (21U)	FSHD1	Female	2,146	200 ≤ nFeatures ≤ 3,500; %MT-genes ≤ 12; %Ribo-genes ≤ 5	1,822	19,655	1,599	2,984	69 [3.8%]	5 [7.2%]	1 [1.4%]	63 [91.3%]
CTRL-01	4qA (32U) / 4qB	CTRL	Female	4,997	200 ≤ nFeatures ≤ 3,500; %MT-genes ≤ 10; %Ribo-genes ≤ 5	4,698	19,396	984	1,669	0 [0.0%]	0 [0.0%]	0 [0.0%]	0 [0.0%]

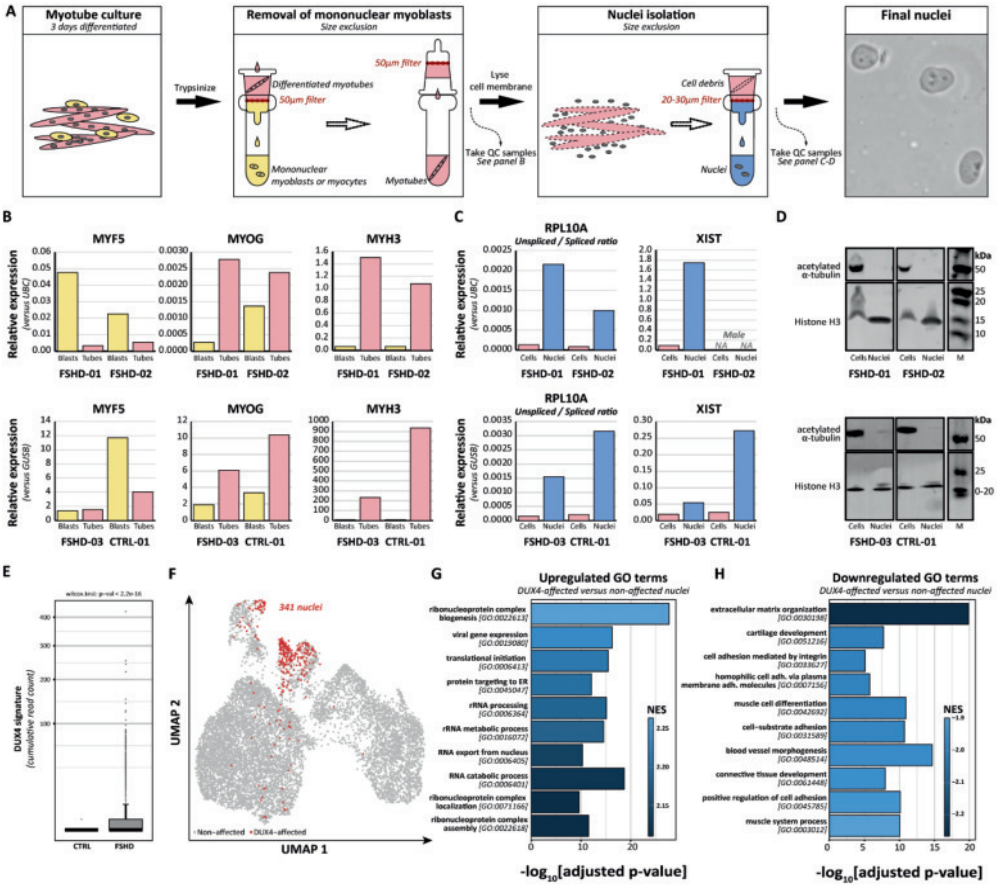

Figure 1. snRNA-seq in late myogenic cells detects increased numbers of DUX4-affected nuclei.

(A) Schematic overview of the nuclei sample preparation, including the specification of the different quality check (QC) samples collected. (B) Validation of the separation of mononuclear myoblasts/myocytes from late multinuclear myotubes, by RT-qPCR analysis of the expression of myogenic differentiation markers (i.e., *MYF5*; early myogenic, *MYOG*; intermediate myogenic, *MYH3*; late myogenic). (C) Validation of the nuclei isolation, by RT-qPCR analysis of the expression of nuclear-enriched versus non-nuclear enriched RNA. Left panels: Comparing the unspliced pre-mRNA (nuclear-enriched) versus spliced mRNA (nuclear and cytoplasmic) of *RPL10A*. Right panels: Comparing the expression of nuclear-enriched *XIST* RNA versus the mRNA of a housekeeping gene with no specific nuclear localization (i.e., *GUSB* or *UBC*). *XIST* RNA could not be detected in FSHD-02 as *XIST* is only expressed in females. (D) Validation of removal of cytoplasmic material (which would suggest whole cell instead of nuclei analysis), by western blot analysis for cytoplasmic acetylated alpha-tubulin. Nuclear histone H3 was used as control for nuclear material. (E) DUX4 signature expression in all FSHD or control nuclei. (F) UMAP plot of the integrated dataset of all four samples, color-coded for the final nuclei classification of DUX4-affected nuclei (expressing DUX4 and/or ≥ 5 DUX4 target genes) versus non-affected nuclei. (G and H) Barplots depicting the top 10 most significantly upregulated (G) or downregulated (H) gene ontology (GO) terms based on gene set enrichment analysis. The selected GO

terms have respectively the highest or lowest normalized enrichment score (NES, highlighted by color-code) and are statistically significantly affected (adjusted *P* value < 0.05).

Nuclei classification identifies two clusters of DUX4-affected nuclei

Clustering and UMAP visualization of the integrated data showed that nuclei can be separated into five major clusters (Fig. 2A). Myogenic marker gene expression analysis showed that nuclei clustering was largely based on myogenic state (Fig. 2B). Nuclei in cluster-1 were marked by high levels of the early myogenic markers *MYF5* and *FN1* (33) and were classified as Early Myonuclei, possibly originating from less differentiated myotubes. This early myogenic stage is in concordance with the closely connected cluster-2, which showed high expression of *MEGF10*, a regulator of satellite cell progression into the myogenic program (34), and *WNT5b* which becomes upregulated in the early phase of muscle regeneration (35). We, therefore, annotated the nuclei in cluster-2 as Middle Myonuclei. In contrast, nuclei in cluster-3 were classified as Late Myonuclei based on the upregulation of late myogenic markers (e.g., *MYH3* and *MEF2C*) and the upregulation of genes involved in muscle development and muscle cell differentiation (Fig. 2B, Fig. S4, and Tables S3 and S4).

Interestingly, where all three FSHD samples contributed to each of the five nuclei clusters, CTRL-01 showed only limited contribution to clusters 4 and 5 (Fig. 2C), suggesting that these clusters represented FSHD-enriched conditions. Indeed, the majority of DUX4-affected nuclei were located in clusters 4 and 5 (Fig. 1F and Fig. 2D), further validating their enrichment for FSHD-associated conditions. Cluster-4 contained the majority of DUX4-affected nuclei and GSEA results indicated increased activity in RNA metabolism, including RNA production, splicing, and localization (Fig. S4, Table S4). This cluster was closest related to the Late Myonuclei cluster (Fig. S5) and was annotated as DUX4-affected cluster I. Cluster-5 was closely associated with the Early/Middle Myonuclei cluster (Fig. S5). It had fewer DUX4-affected nuclei than Cluster-4, but the upregulated genes in this cluster were involved in RNA processing, translation, and increased protein production and energy metabolism (Fig. S4, Table S4), pathways previously described to be affected in FSHD (11, 29, 36). This cluster was annotated as DUX4-affected cluster II.

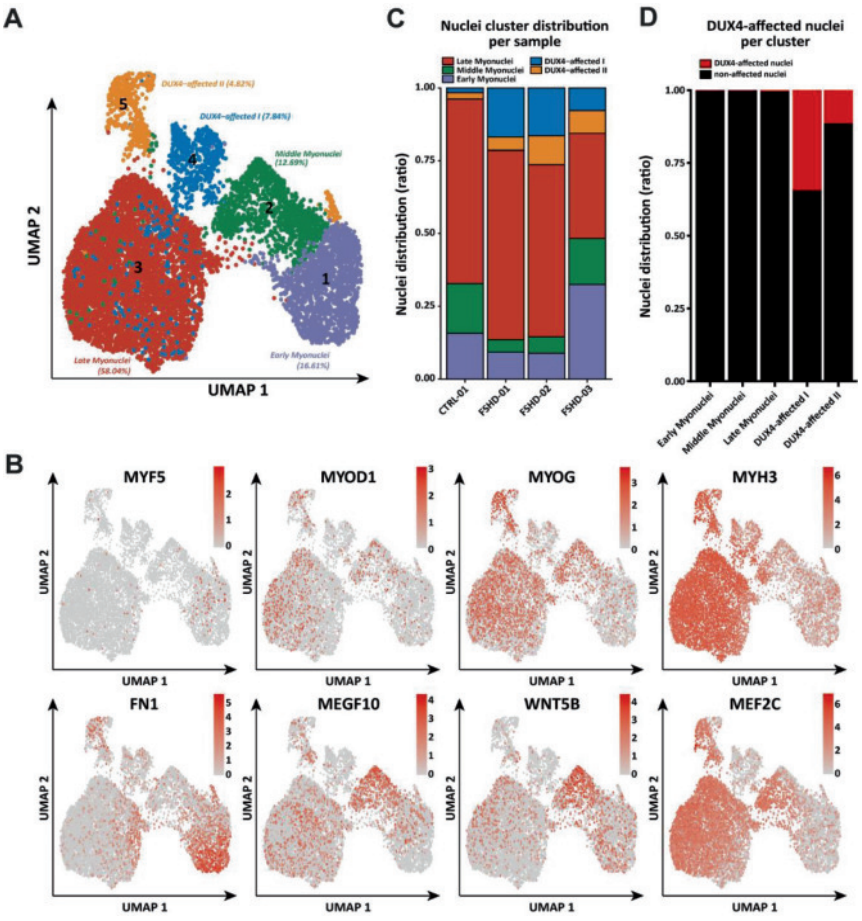

Figure 2. snRNA-seq on late myogenic cells identifies two DUX4-affected nuclei clusters.

(A) UMAP visualization of all 9873 nuclei in the integrated dataset, colored by cluster identity. (B) UMAP plots as in A) colored based on the expression level of myogenic marker genes. Color scales depict the loge-normalized expression level per 10,000 reads for each specified gene. (C) Stacked bar plot showing the proportion of nuclei in each cluster per sample. D) Stacked bar plot showing the proportion of DUX4 target^{POS} nuclei in each cluster.

DUX4-affected clusters display distinct transcriptomes

Although various DUX4-induced responses have been described, their spatiotemporal relation as well as their interdependence remains largely unclear. We therefore next focused on the cause of the separation of the DUX4-affected clusters and characterized the difference between the two DUX4-affected clusters.

FSHD nuclei in DUX4-affected cluster II showed a strong increase in oxidative phosphorylation (Molecular Signature Database (MSigDB), M5936) and energy metabolism (Gene Ontology (GO) term GO:0046034) (Fig. 3A). As FSHD has been associated with

impaired mitochondrial activity, increasing oxidative phosphorylation may lead to increased oxidative stress (11, 12, 36). Indeed, the increased expression of the oxidative phosphorylation pathway in DUX4-affected cluster II resulted in increased expression of genes involved in reactive oxygen species handling (MSigDB, M5938) in our data (Fig. 3B), genes involved in oxidative stress response (MSigDB, M3223/GO:0006979, Fig. 3C) as well as genes with antioxidant activity (MSigDB, M15021/GO:0016209, Fig. 3D), ultimately leading to the activation of oxidative stress-induced intrinsic apoptosis signaling (MSigDB, M22556/GO:0008631, Fig. 3E). DUX4-affected cluster II might thus represent a cellular state with a higher energy demand than the FSHD cells can cope with. Interestingly, the oxidative stress related pathways were already increased in FSHD nuclei without any DUX4 and/or DUX4 target activation and did not strongly correlate with increasing numbers of expressed DUX4 target genes. This is in contrast to several direct DUX4-induced pathways (e.g., spliceosome (MSigDB, M2044) and ribonucleoprotein complex biogenesis (MSigDB, M16027/GO:0022613), which showed more comparable increases in both DUX4-affected clusters and of which the activation correlated better with DUX4 signature activation (Fig. S6). These direct DUX4-induced responses seemed, therefore, independent of the cluster they were located in.

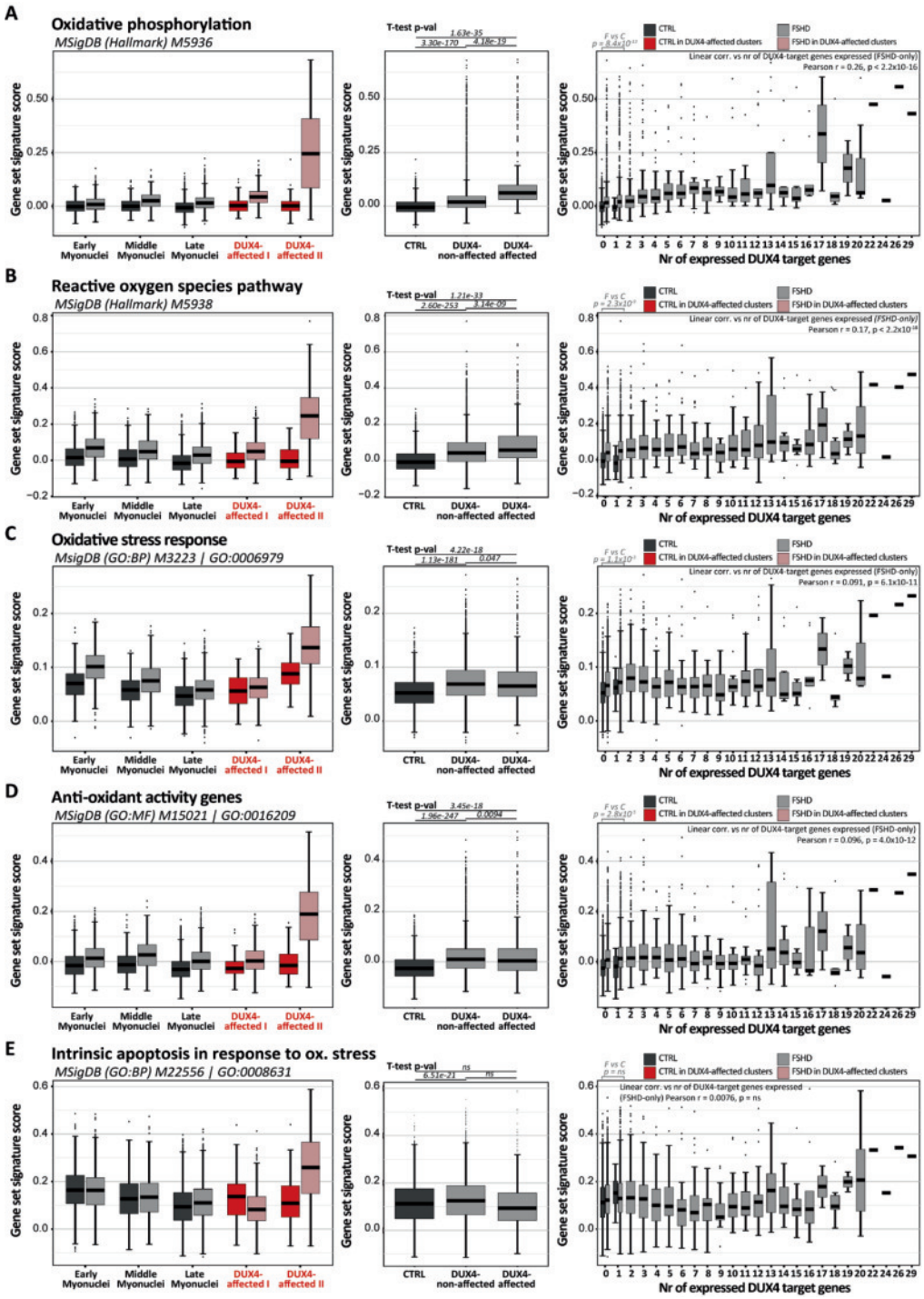
A
Oxidative phosphorylation
MSigDB (Hallmark) M5936
T-test p-val
1.63e-35
3.30e-170
4.18e-19
Gene set signature score
Early Myonuclei
Middle Myonuclei
Late Myonuclei
DUX4-affected I
DUX4-affected II
CTRL
DUX4-non-affected
DUX4-affected
CTRL
CTRL in DUX4-affected clusters
FSHD
FSHD in DUX4-affected clusters
Linear corr. vs nr of DUX4-target genes expressed (FSHD-only)
Pearson r = 0.26, p < 2.2x10-16
Nr of expressed DUX4 target genes
B
Reactive oxygen species pathway
MSigDB (Hallmark) M5938
T-test p-val
1.21e-33
2.60e-253
3.14e-09
Pearson r = 0.17, p < 2.2x10-16
C
Oxidative stress response
MsigDB (GO:BP) M3223 | GO:0006979
T-test p-val
4.22e-18
1.13e-181
0.047
Pearson r = 0.091, p = 6.1x10-11
D
Anti-oxidant activity genes
MSigDB (GO:MF) M15021 | GO:0016209
T-test p-val
3.45e-18
1.96e-247
0.0094
Pearson r = 0.096, p = 4.0x10-12
E
Intrinsic apoptosis in response to ox. stress
MSigDB (GO:BP) M22556 | GO:0008631
T-test p-val
ns
6.51e-21
ns
Linear corr. vs nr of DUX4-target genes expressed (FSHD-only) Pearson r = 0.0076, p = ns

Figure 3. DUX4-affected cluster II shows increased levels of oxidative stress and oxidative stress-related apoptosis.

(A–E) Gene set signature scores in FSHD and control nuclei, grouped per cluster (left), grouped per DUX4-affected state (middle) or nr of DUX4-target genes expressed (right), for A) Hallmark oxidative phosphorylation gene set, B) Hallmark reactive oxygen species pathway genes, C) GO (biological processes); response to oxidative stress gene set, D) GO (molecular function); genes with antioxidant activity and E) GO (biological processes); intrinsic apoptosis signaling in response to oxidative stress pathway genes.

Differences in DUX4-affected clusters might be related to myogenic differentiation stage

The cause of the strong increase in oxidative phosphorylation could be related to the switch from glycolytic to oxidative energy metabolism during myogenic differentiation (37). Interestingly, only DUX4-affected cluster II showed this increased oxidative phosphorylation, whereas DUX4-affected cluster I did not. Jiang et al. (27) have previously also identified two distinct DUX4-affected clusters which the authors speculated could be partially caused by a difference in myogenic differentiation stage, with late-stage myogenic nuclei being associated with higher levels of FSHD-associated gene expression (27). Indeed, DUX4-affected cluster I, containing the most DUX4-affected nuclei and having the highest average DUX4-target gene expression, was most related to the Late Myonuclei cluster, whereas DUX4-affected cluster II was more related to Early/Middle Myonuclei (Fig. S5). This may suggest a role for myogenic differentiation stages in FSHD-associated defects like induced oxidative stress and related apoptosis.

Although DUX4-affected cluster I was most related to the Late Myonuclei cluster (Fig. S5), it showed a strong reduction in myotube differentiation markers in FSHD nuclei in the cluster, complicating further conclusions on the true myogenic differentiation stage of the nuclei in our data (Fig. 4A). This reduction correlated with DUX4 signature activation, which corresponds with previous literature describing an inhibitory role for DUX4 in myogenic differentiation (8, 9, 12, 38). Interestingly, myogenic differentiation was inhibited only in the FSHD nuclei in DUX4-affected cluster I, whereas this was not the case in the DUX4-affected cluster II, again suggesting different FSHD-associated changes in the different clusters.

Where the DUX4 signature score is known to gradually increase from early to late myogenic stage, the FSHD-associated reduction in PAX7 score showed the opposite pattern, being most clearly reduced in early myogenic FSHD nuclei as compared to the control nuclei in this cluster (Fig. S3E-F). This reduction in PAX7 score was no longer detectable in later stages of myogenic differentiation. In concordance, the reduced PAX7 score was detectable in DUX4-affected cluster II, which has a closer relation to Early/Middle Myonuclei, whereas the reduced PAX7 score was not detected in DUX4-affected cluster I, most related to Late Myonuclei. These observations further suggest the independence of both signatures and may reflect the fact that PAX7 functions mostly in satellite cells and early stages of myogenic differentiation (39, 40) and that its expression is reduced in later stages of myogenic differentiation. Indeed also in our data PAX7 expression itself was higher in early stages of myogenic differentiation (Fig. S3G-

H), further suggesting that this is the stage at which PAX7's function can be strongest intervened in FSHD.

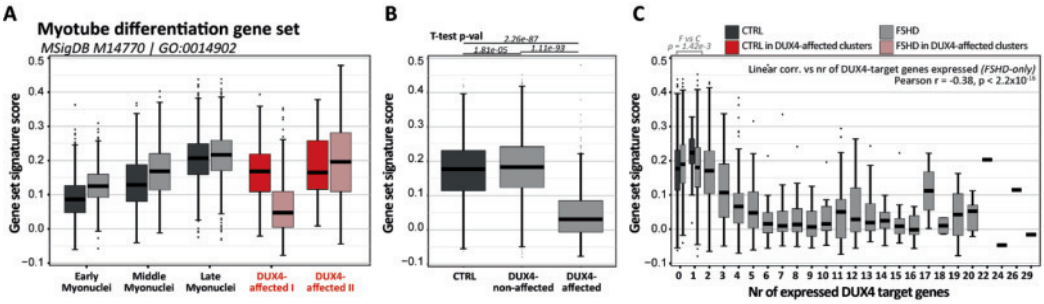

Figure 4. DUX4 expression results in inhibition of myogenic differentiation specifically in DUX4-affected cluster I.

(A-C) Myogenic differentiation gene set signature score in FSHD and control nuclei, A) grouped per cluster, B) grouped per DUX4-affected state or C) grouped per nr of DUX4-target genes expressed.

Global FSHD-associated responses

Although the DUX4-affected clusters showed the strongest FSHD-related transcriptional changes, several transcriptional changes could also be identified in the FSHD nuclei of the other three non-affected clusters (Early, Middle and Late Myonuclei) (Fig. S7A, Table S5). SMCHD1, LRIF1 and DNMT3B, three chromatin modifiers known to be repressors of DUX4 and identified as FSHD disease genes or modifiers, were reduced in FSHD nuclei compared to control nuclei in all clusters (3-5). This may provide a susceptibility for DUX4 activation in FSHD nuclei (Fig. S7B). Pathway analysis showed several pathways to be affected in FSHD versus control nuclei in the different clusters, including reduced chromosomal organization (Fig S7C highlighted in blue), increased immune-related pathways (Fig S7C highlighted in turquoise), increased extracellular matrix organization (Fig S7C highlighted in green), increased mRNA processing (Fig S7C highlighted in dark red/purple) and protein production/localization (Fig S7C highlighted in dark blue/purple) (also complete list in Table S6). This may suggest a more global effect beyond the detection limit of DUX4 signature activation. However, as also the DUX4 signature itself was already slightly increased in the FSHD nuclei of each cluster (Fig. S7D), we cannot exclude that these subtle transcriptional changes are the consequence of sporadic DUX4 expression.

FSHD-affected clusters activate an early-embryonic-like transcriptome program

DUX4 plays a critical role in embryogenesis by driving zygotic genome activation, initiating transcription of repetitive elements, and regulating chromatin structure. In addition, DUX4 has been shown to reactivate an early embryonic signature in diverse cells. For example, transient DUX4 expression in human embryonic stem cells can induce a blastomere-like expression program (41). In addition, in induced pluripotent stem cell cultures, the population of cells resembling the human embryo in early development is usually accompanied by embryonic genome activation regulated by DUX4 (42). A stem cell signature and transcription factor

network typical for early embryogenesis and pluripotent stem cells were detected upon DUX4 activation in inducible DUX4 HAP1 cells (43). And finally, also in muscle cells, DUX4 has been shown to activate several aspects of an embryonic transcriptome program (44, 45). As several of the early embryonic pathways overlap with the pathways affected in FSHD muscle, we wondered if the DUX4-affected clusters reflect stages of DUX4‘s natural role during embryonic development. To test this hypothesis, we reanalyzed a publicly available embryogenic scRNA-seq dataset (46) to define signature gene sets for the early stage of embryonic development, following DUX4 expression at the 4-cell cleavage stage (Table S7). For this, we performed DE analysis to identify the significantly upregulated genes, defining representative markers of the 4-cell, 8-cell, morulae and blastocyst stages. We next used these signature gene sets to calculate their signature scores in our snRNA-seq data. Strikingly, FSHD nuclei in DUX4-affected cluster I highly expressed the 8-cell stage signature gene set, while DUX4-affected cluster II nuclei exhibit higher expression of the blastocyst signature gene set (Fig. 5A). Vice versa, marker genes of DUX4-affected cluster I displayed a higher expression in 8-cell stage cells, and blastocysts showed elevated expression of DUX4-affected cluster II markers (Fig. 5B). To further investigate the similarity between the embryonic cells and DUX4-affected nuclei, the marker genes of embryonic cells and the DUX4-affected clusters were compared. Among them, 66 DE genes were shared between 8-cell stage cells and the DUX4-affected I cluster, while 191 DE genes were common between blastocysts and the DUX4-affected II cluster (Fig. S8). GSEA (GO Biological processes) of the most upregulated 8-cell signature genes in DUX4-affected cluster I demonstrated that these were enriched in GO terms associated with transcription initiation (e.g., GO:0006352) and RNA processing and metabolism (e.g., GO:0008380, GO:0006403 and GO:0022613), while the majority of the representative GO terms of the blastocysts markers expressed in DUX4-affected II were associated with protein production (e.g., GO:0006457) and energy metabolism (e.g., GO:0046034, GO:0032981 and GO:0006119) (Fig. S8, Table S8).

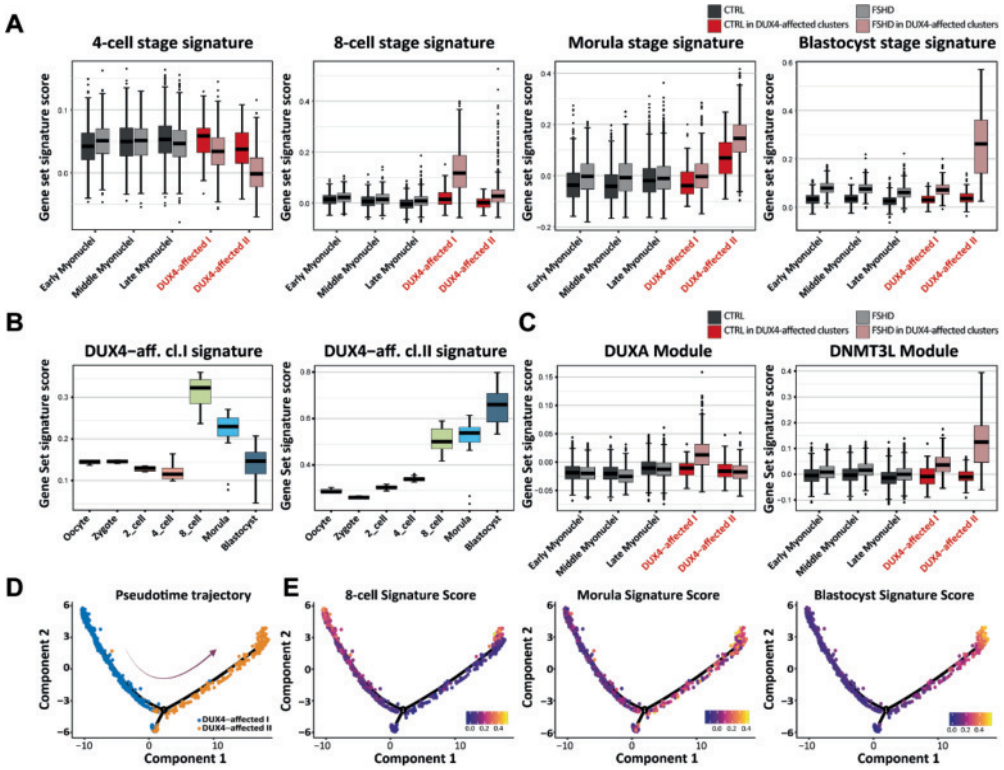

Figure 5. DUX4-affected cluster nuclei activate an embryonic-like transcriptome program.

(A) Embryonic stage signature scores in FSHD and control myotube nuclei per cluster. (B) DUX4-affected signature expression during different stages of embryonic differentiation. (C) DUXA and DNMT3L gene module score in each of the 5 nuclei clusters. (D) Pseudotime trajectory analysis of all nuclei in DUX4-affected cluster I and II, color-coded for their cluster identity, describing a possible relation between the nuclei in DUX4-affected cluster I and DUX4-affected cluster II. (Supervised pseudotiming based on their cluster marker genes.) (E) Pseudotime trajectory analysis as in D), color-coded for different embryonic stage signature score activity.

Developmental dynamics from DUX4-affected I to DUX4-affected II

Interestingly, the embryonic signature scores suggest the presence of a transition from DUX4-affected cluster I to DUX4-affected cluster II, corresponding to the development from 8-cell stage to morulae and blastocysts during embryogenesis. The transcriptomic dynamics from the 8-cell stage to late blastocysts can be defined by a few gene modules (47). We calculated the signature score of each module in the DUX4-affected clusters. Intriguingly, in DUX4-affected cluster I, the signature score of the DUXA module, which is critical for development from the 8-cell stage to morulae, was increased. In DUX4-affected cluster II nuclei, the genes in the DNMT3L module, which are activated mostly during blastocyst development, were strongly expressed (Fig. 5C). To further dissect the cellular dynamics, we performed supervised pseudotime trajectory analysis using the 8-cell and blastocysts signature gene sets on the nuclei

from both DUX4-affected clusters. Nuclei showed a clear separation, with a high signature score for the 8-cell stage genes for DUX4-affected cluster I nuclei on the left side of the pseudotime trajectory, followed by an increasing gradient of the morulae signature score to the right and eventually a high score for the blastocyst signature in the DUX4-affected cluster II nuclei on the right side of the trajectory (Fig. 5D-E). This pseudotime trajectory may thus describe a cellular transition from DUX4-affected I to DUX4-affected II, analogous to the progression from the 8-cell stage to blastocysts.

DUX4 target gene expression was most associated with the 8-cell stage-like nuclei, showing increased expression mostly towards the left side of the pseudotime trajectory, and showing expression in the few nuclei on the right side that showed high 8-cell stage signature expression (Fig. S9). Further, DUX4 expression itself was more often detected in the left branch. In contrast, both DUX4 and DUX4 signature expression was lower in the nuclei expressing the blastocyst-like signature (Fig. S9). In addition, the left branch showed the reduced myogenic differentiation previously detected in DUX4-affected cluster I, whereas the right branch showed increased oxidative stress-induced apoptosis linked to DUX4-affected cluster II. Altogether, this pseudotime analysis may thus describe a cellular transition from the activation of early DUX4-induced responses (e.g., DUX4 signature) in the left branch towards late-stage disease activation (e.g., stress-induced apoptosis) in the right branch.

Discussion

The sporadic nature of DUX4 expression in FSHD muscle has been a challenge in comparative transcriptome analyses between FSHD and control samples. A variety of DUX4 and FSHD-associated transcriptional changes have been identified, though due to bulk RNA-seq analyses their spatiotemporal relation, interdependence and role in the disease process remain unclear. In this study, we used snRNA-seq of nuclei isolated from patient- and control-derived primary multinucleated myotubes to investigate the cellular heterogeneity in FSHD. Several parameters such as the increased mRNA levels of late myogenic markers and nuclear-enriched RNAs confirmed the enrichment of late stage myotube nuclei in our snRNA-seq datasets. Taking advantage of the increased resolution in snRNA-seq of fully differentiated myotubes over our previous single myocyte study (26), two populations of DUX4-affected nuclei could be identified in three independent FSHD myotube cultures. These two populations showed distinct transcriptional profiles, suggesting separate FSHD-associated transcriptional states, possibly partially defined by mutually exclusive responses to DUX4.

The transcriptional profile of the first population of nuclei, DUX4-affected cluster I, shows high similarity with that of 8-cell stage embryos. Genes shared between these stages were associated with increased gene transcription and RNA processing and metabolism (Fig. S8), processes which are crucial in 8-cell stage embryos (46, 48, 49). In our snRNA-seq myotube data, increased activation of these pathways correlated with increased expression of the known DUX4 direct target genes, strongly suggesting that it is DUX4 reactivation that leads to the inappropriate activation of an early embryonic-like program in muscle.

Interestingly, the second population of nuclei, DUX4-affected cluster II, showed strong overlap with a later stage of embryonic development, blastocysts, when DUX4 itself is thought to be

already repressed (2). Indeed, our pseudotime trajectory analysis indicates that a cellular transition from DUX4-affected cluster I to II may continue after the dynamic burst of DUX4 (and DUX4's initial direct responses) has faded. This suggests that even in postmitotic myonuclei DUX4 initiates (part of) an embryonic program beyond its own presence, consistent with muscle cell culture studies showing that DUX4 can have perduring effects on chromatin structure and gene expression of its direct target loci (50). In addition, it may suggest that re-activation of DUX4 in muscle initiates an elaborate embryonic program beyond the immediate effects of DUX4 in the cleavage stage. Whether these two different responses (early/late embryonic) are indeed consecutive or can occur in parallel, will need to be further investigated. Alternatively, DUX4 might activate specific pathways shared between blastocyst development and FSHD pathology, thereby mimicking the blastocyst transcriptome signature. Biological processes included in this embryonic program included oxidative phosphorylation-associated energy metabolism and mitochondrial biogenesis (see also Fig S7), pathways known to be increased in blastocyst formation (51, 52).

Muscle cells may be incapable of appropriately responding to the aberrant strong increase in oxidative phosphorylation observed in DUX4-affected cluster II, resulting in oxidative stress and oxidative stress-induced apoptosis. Of note, with oxidative stress being described in a positive feedback loop with DUX4 activation (both enhancing each other) (6, 12, 53, 54), the cause of the strong increase in oxidative phosphorylation remains to be further investigated. Another explanation for the increase in oxidative phosphorylation may be related to the natural switch from glycolytic to oxidative energy metabolism during myogenic differentiation. With the impaired mitochondrial activity in FSHD (11, 12, 36), switching from glycolytic to oxidative energy metabolism may lead to impaired energy production, triggering oxidative stress and overcompensation of oxidative phosphorylation pathway gene expression. In both situations, oxidative stress and oxidative stress-induced apoptosis seem to play an important role in FSHD-associated cytotoxicity in muscle.

Both DUX4-affected clusters are populated by nuclei from the control sample, indicating that those control nuclei share transcriptional features with the FSHD nuclei in the DUX4-affected clusters, despite the fact that these control nuclei have no activated DUX4 (or DUX4 signature). One possible explanation for this is that these control nuclei share a common cellular state, characterizing the DUX4-affected clusters, independent of the DUX4-induced responses (e.g., myogenic differentiation stage, energy demand, or other). This would not only indicate the presence of a specific cellular state (or cellular states) that predispose to DUX4 activation, but it also suggests that this cellular state is also present in normal myogenic conditions (i.e., in control muscle cells) or may be stress-induced independent of DUX4 or FSHD in general. Yet, only in FSHD cells, this leads to DUX4 expression, as marked by the transcriptional differences in FSHD versus control nuclei within each cluster. This interpretation may provide guidance toward understanding what triggers DUX4 expression.

In this study, we identify two DUX4-affected nuclei populations, both showing distinct transcriptional profiles, which suggests a heterogeneous and possibly mutually exclusive mix of DUX4-induced responses in muscle cells. The cause of this heterogeneity remains only partially understood and may be dependent on both the timing of DUX4 activation (e.g., the

cellular state during which DUX4 gets activated) as well as the duration of the DUX4-induced response activation (early/late DUX4-induced responses). Our data suggests an involvement for myogenic differentiation in the FSHD-associated transcriptional heterogeneity. Similar to the study from Jiang et al. (27), nuclei from both DUX4-affected clusters originated from myotubes of different myogenic differentiation stages. In addition, only one DUX4-affected cluster showed the described DUX4-associated inhibition of myogenic differentiation. Whether these differences in myogenic differentiation stage is a cause or a consequence of DUX4 activation remains to be studied. In addition, the myogenic differentiation stage may also affect the degree of oxidative stress in the DUX4-affected cell. Inhibition of myogenic regulators has been described to inhibit mitochondrial energy metabolism and therefore oxidative phosphorylation (55). In this situation, the protection from (or at least lack of) oxidative stress in DUX4-affected cluster I is an indirect effect of DUX4's inhibition of terminal differentiation. Reversely, reduced mitochondrial activity has also been described to block myogenic differentiation (37), suggesting a reverse causal relation.

Interestingly, some of the known DUX4-direct induced responses seemed independent of the cellular state (i.e., cluster) they were in, but were more directly associated with DUX4 activation. This might provide insights into the earliest DUX4-induced responses, still independent of the cellular state of the affected cell during DUX4 activation.

It is important to note the multinuclear nature of differentiated myotubes complicating the interpretation of this data analysis. DUX4 nuclear protein spreading to neighboring nuclei within the same myotube is commonly observed in multinuclear myotubes (see also Fig. S1). In multinuclear cells, nuclear protein/signal propagation may cause non-affected nuclei to become affected by any DUX4-affected nuclei in the shared cytoplasm, even when they do not express DUX4(-target genes) themselves (20, 29, 56). In addition, DUX4 expression is dynamic and DUX4-induced cytotoxicity can be observed after DUX4 itself has faded (29). For example, even though oxidative stress-induced pathway activation did not correlate with DUX4 target gene activation directly it may be a result of DUX4 signal transduction throughout the entire affected myotube. This complicates functional validation of the presence of two distinct DUX4 induced responses and any further analysis of their coexistence (or lack thereof) within one single multinucleated myotube. This will require the identification of separatable signals, being able to analyze the earliest DUX4-induced responses before being affected by the signal transduction throughout the myotube. For this, future high-resolution (subcellular resolution) RNA analyses with retaining spatial information will be required, an analysis field which is still in development, but for which technologies are currently rapidly advancing.

Finally, previous studies have shown a role for sequential bursts of DUX4 expression in the degree of DUX4-induced responses and cytotoxicity (50). This may be an additional explanation for the two DUX4-affected clusters in our study, in which both clusters may represent different stages of bursting DUX4-expression. In this case, DUX4-affected cluster I would represent nuclei that express a non-toxic burst of DUX4, whereas multiple bursts of DUX4 would cause cells to develop into DUX4-affected cluster II where cells become apoptotic.

Altogether, our data suggests that the FSHD transcriptional profile is defined by a mixture of DUX4-induced responses and cellular state-dependent FSHD-associated (downstream) effects. In addition, combining this dataset with published embryonic scRNA-seq data revealed that the DUX4-affected clusters displayed similarities to distinct pre-implantation embryonic cell stages (i.e., 8-cell stage cells and blastocysts). This suggests that misexpression of DUX4 in the muscle environment leads to the (partial) execution of an early embryogenic program, which may ultimately be incompatible with muscle cell development.

Materials and Methods

Cell line information

Human primary myoblast cell lines originated from the University of Rochester biorepository (http://www.urmc.rochester.edu/fields-center/). Muscle samples were obtained after informed consent under a protocol approved by the Institutional Review Board at the University of Rochester. This study includes myogenic cell cultures derived from three FSHD-patients (FSHD type 1) and one control donor. Detailed information on the genetic background of the included samples is included in Table 1.

Cell culture

Myoblasts were cultured in DMEM/F-10 medium (#41550–021, Life Technologies, Waltham, Massachusetts, USA), supplemented with 20% heat-inactivated fetal bovine serum (#10270, Gibco/Life Technologies, Waltham, Massachusetts, USA), 10ng/ml rhFGF (#C-60240, Bio-Connect, Huissen, Gelderlangd, The Netherlands) and 1uM dexamethasone (#D2915, Sigma-Aldrich, St. Louis, Missouri, USA).

Cells were differentiated into multinucleated myotubes for three or four days by culturing in DMEM (#41966-029, Life Technologies, Waltham, Massachusetts, USA) supplemented with 15% knockout serum replacer (#10828-028, Life Technologies, Waltham, Massachusetts, USA).

Immunofluorescence staining

After 1, 2, 3, and 4 days of myogenic differentiation, the cells were fixed in 4% paraformaldehyde for 15 min at room temperature, permeabilized in 2% Triton for 10 min, then washed with PBS. Immunostaining was performed overnight at 4°C with a 1:2000 dilution of DUX4 E5.5 rabbit monoclonal antibody (ab124699, Abcam) in PBS, a 1:250 dilution of MF20 antibody against myosin heavy chain (MYH1E (MF20, DSHB). Cells were washed and labeled with fluorescent-conjugated Alexa 488 or Cy3 anti-rabbit or anti-mouse secondary antibodies for 1 h at room temperature, then washed again. The images were acquired on a Dragonfly 500 Confocal Microscope System using an immersion lens under 60x magnification. A fusion index was calculated for each image with CellProfiler software (v2.1.1) using a custom made analysis pipeline. In short, individual nuclei and larger nuclei clusters were segmented based on Hoechst staining and were identified based on shape and size (see Fig. S1). Fusion index was defined as the percentage of individual nuclei located in a nuclei cluster of >2 nuclei in size. A threshold of >2 nuclei was used to reduce the false positive labeling of multinucleated myotube nuclei due to a possible slight oversegmentation of the individual nuclei.

Nuclei sample preparation and RNA-sequencing

Differentiated cell cultures were trypsinized and resuspended in DMEM (#41966-029, Life Technologies, Waltham, Massachusetts, USA). Mononuclear myoblasts and myocytes were separated from multinucleated myotubes by passing the cells through a 50µm filter (CellTrics filter, #04-004-2327, Sysmex, Kobe, Hyogo, Japan). The collected flowthrough contained the myoblast/myocyte-enriched fraction. The multinucleated myotube-enriched fraction was collected by inverting the filter and flushing the filter three times with ~5ml DMEM. From this step, all steps were performed on ice. 5-10% of each fraction was used for validation of the multinuclear myotube enrichment on respectively RNA (Fig. 1B) and protein (Fig. 1D) level. The remaining cells were collected by centrifugation at 1,000rpm (table top centrifuge) for 5 min at 4°C. The cell membranes were lysed by incubating the cell pellets for a maximum of 5 min on ice in 100µl cold 1%-lysis buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM MgCl2, 1% NP40 (substitute)) with 0.2U/µl Protector RNAse-inhibitor (#3335399001, Sigma Aldrich, Saint Louis, Missouri, USA). After this, 900µl 0%-lysis buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM MgCl2) with 0.2U/µl Protector RNAse-inhibitor was added and samples were transferred to (non-stick) 1.5ml tubes, after which samples were further lysed by pipetting up/down 50x with a 1ml pipette tip. Completion of lysis was checked by visual inspection of an aliquot under the microscope (incl tryphan blue staining (#1450021, Biorad, Hercules, California, USA)). Nuclei were purified from large cell debris by passing the nuclei through a 20µm filter (CellTrics filter, #04-004-2325, Sysmex, Kobe, Hyogo, Japan). The collected flowthrough contained the nuclei fraction. The nuclei were centrifuged for 5 min at 500xg at 4°C, and washed in 1ml cold PBS-1%BSA (w/v) (#B8894-5ml, Sigma-Aldrich, Saint Louis, Missouri, USA). The nuclei were counted and split into three samples for respectively validation of nuclei isolation and purity on RNA (Fig. 1C) and protein level (Fig. 1D) and the final snRNA-seq procedure. For the final snRNA-seq procedure, samples were centrifuged once more for 5 min at 500xg at 4°C and resuspended in an appropriate volume of PBS-1%BSA to reach approximately 1,000 nuclei/µl according to 10X Genomics guidelines.

SnRNA-seq library preparation and Illumina next-generation sequencing was next performed by the Leiden Genome Technology Center (Leiden University Medical Center, Leiden, Zuid-Holland, The Netherlands). Samples were prepped for snRNA-seq according to the Chromium™ Single Cell 3′ v2 (for FSHD-01 and FSHD-02) or v3 (For FSHD-03 and CTRL-01) RNA sequencing specification. The generated cDNA was used for Illumina next-generation sequencing using a NextSeq500-v2 150 cycle kit (#FC-404-2002, Illumina, San Diego, California, USA).

Validation of multinuclear myotube enrichment

For the validation of correct multinuclear myotube separation from mononuclear myoblasts/myocytes (Fig. 1B), QC samples collected after the first size-exclusion filtering step (see above) were centrifuged for 5 min at 500xg at 4°C. RNA was isolated using the miRNeasy mini kit (#217004, Qiagen, Venlo, Limburg, The Netherlands) according to manufacturer's instructions (including an on-column 30 min DNAse I treatment (#79254, Qiagen, Venlo, Limburg, The Netherlands)). cDNA was synthesized with the RevertAID first strand cDNA synthesis kit (#K1621, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and cDNA was used for

RT-qPCR analysis using the iQ™ SYBR® Green Supermix (#1708886, Biorad, Hercules, California, USA) with a PCR program of; 95°C for 6 min, followed by 40 cycles of 95°C for 10 sec and 60°C for 30 sec in a CFX-384 or CFX-96 Real-Time PCR system (Biorad). The following primers were used:

MYH5: Fw primer; 5'-TTCTCCCCATCCCTCTCGCT-3', Rv primer; 5'-AGCCTGGTTGACCTTCTTCAG-3', *MYOG*: Fw primer; 5'-CAGCTCCCTCAACCAGGAG-3', Rv primer; 5'-GCTGTGAGAGCTGCATTCG-3' and *MYH3*: Fw primer; 5'-GATTGCAGGATCTGGTGGAT-3', Rv primer; 5'-CCTGCTGGAGGTGAAGTCTC-3' and the two housekeeping genes:

GUSB: Fw primer; 5'-CTCATTTGGAATTTTGCCGATT-3', Rv primer; 5'-CCGAGTGAAGATCCCCTTTTTA-3', *UBC*: Fw primer; 5'-gtccatcttccagctgtttccca-3', Rv primer; 5'-gcccagtgacaccatcgagaat-3'.

Validation of nuclei isolation purification

For the validation of nuclei isolation and purification based on RNA content (Fig. 1C), QC samples collected after the second size-exclusion filtering step (see above) were centrifuged for 5 min at 500xg at 4°C. RNA was isolated, cDNA was synthesized and RT-qPCR was performed as described above (for the validation of multinuclear myotube enrichment). For this specific RT-qPCR the following primers were used:

RPL10A$_{unspliced}$: Fw primer; 5'-TCTCTCGCGACACCCTGT-3', Rv primer; 5'-AGAGAGGAGGGGGGTTAAG-3'

RPL10A$_{spliced}$: Fw primer; 5'- TCTCTCGCGACACCCTGT-3', Rv primer; 5'-TTAGCCTCGTCACAGTGCTG-3'

XIST: Fw primer; 5'-tggcttcgtcattgtccttc-3', Rv primer; 5'-ctgcatttcacatcagttcacaag-3' and the two housekeeping genes as above:

GUSB: Fw primer; 5'-CTCATTTGGAATTTTGCCGATT-3', Rv primer; 5'-CCGAGTGAAGATCCCCTTTTTA-3', *UBC*: Fw primer; 5'-gtccatcttccagctgtttccca-3', Rv primer; 5'-gcccagtgacaccatcgagaat-3'.

For the validation of nuclei isolation and purification based on protein content (Fig. 1C), QC samples collected after the second size-exclusion filtering step (see above) were centrifuged for 5 min at 500xg at 4°C. The pellets were lysed for 30 min on ice in 50µl RIPA buffer (20mM Tris-HCl pH7.5, 5mM EDTA, 150mM NaCl, 0.1% (w/v) SDS, 1% (v/v) NP40, 0.5% (w/v) sodium deoxycholate) with 1x cOmplete™, EDTA-free Protease Inhibitor Cocktail (#11873580001, Roche, Basel, Switzerland). Protein concentration was determined with the Pierce™ BCA Protein Assay Kit (#23227, Thermo Fisher Scientific, Waltham, Massachusetts, USA) after which samples were supplemented 2% β-mercaptoethanol or 100mM DTT and bromophenol blue. Directly prior to loading of samples on protein gels, samples were boiled at 95°C for 5 min. Protein from each sample was separated using a 7.5% acrylamide/bisacrylamide gel (Biorad) and the proteins were next transferred to an Immobilon-FL PVDF membrane (#IPFL00010, EMB Millipore, Burlington, Massachusetts, USA). Membranes were blocked in 5% skim milk (#70166-500G, Sigma Aldrich, Saint Louis,

Missouri, USA) in PBS, followed by three 10 min PBS washes. Membranes were next probed with the following primary antibodies in 1% skim milk in PBS at 4°C overnight: Rabbit-anti-Histone H3 (#Ab1791, Abcam, Cambridge, UK, 1:5,000 dilution) and Mouse-anti-acetylated-α-tubulin (#T6199 (clone DM1A), Sigma Aldrich, Saint Louis, Missouri, USA, 1:2,000 dilution). After washing three times 10 min with PBS/0.05% Tween-20, the membranes were incubated with the secondary antibodies in 1% skim milk in PBS: Donkey-anti-Rabbit-IRDye-800CW (#926-32213, Li-cor, Miami, USA, 1:10,000 dilution) and Donkey-anti-Mouse-IRDye-680RD (#926-68072, Li-cor, Miami, USA, 1:10,000 dilution). After washing the membranes (2x 10 min with PBS/0.05% Tween-20 and 1x 10 min with PBS), the membranes were scanned on an Odyssey Classic infrared imaging system (LI-COR) using the manufacturers application software (V3.0) and further analyzed using Image Studio Lite v5.2.

Single-nucleus RNA-sequencing data processing

Sequence fastq files were processed with Cell Ranger 3.0.1 and mapped to a pre-mRNA reference generated from Genome Reference Build 38 (refdata-cellranger-GRCh38-3.0.0) by manufacturer's instructions. The reads count matrix of each sample were imported into the Seurat R package (Version 3.2.2) (*1*) for quality control analysis. Each sample was filtered according to its own thresholds including nFeatures, percentage of mitochondrial gene content and percentage of ribosome gene content (Table 1), then scaled by library size (resulting in a final 'counts per 10,000 reads') and log-transformed. Highly variable genes (HVGs) for each library were extracted using the Scater R package (Version 1.16.2) (*2*). The HVGs of each sample were used to perform a principal component analysis (PCA). A K-nearest-neighbor graph was constructed based on the euclidean distance in PCA space (k = 20 and top 20 PCs) using the "FindNeighbors" function and Louvain algorithm was applied to group nuclei together by the "FindClusters" function with the resolution parameter set to 0.4. We visualized the data using Uniform Manifold Approximation and Projection (UMAP).

To integrate nuclei from different samples we used the overlapping HVGs from Scater and Seurat to find anchors via the "FindIntegrationAnchors" and "IntegrateData" functions in Seurat. Data were normalized according to "SCTransform" method. The top 30 PCs were used for UMAP construction and unsupervised clustering (resolution = 0.4) in the integrated data. The correlation of clusters were determined by the "BuildClusterTree" function in Seurat and subclusters were merged based on the parent clusters. . The merged count matrix in "RNA assay" was normalized and used to identify specific markers in each cluster by the "FindAllMarker" function in R package Seurat (version 4.1.0). All clusters were assigned an annotation based on significant biomarkers and representative GO terms as described in the Results section.

Differential gene expression analysis and gene set enrichment analysis

Differentially expressed genes between different groups were determined by the "FindAllMarkers" function in R package Seurat. The thresholds for significant differentially expressed gene were $|avg\_log2(FC)| > log2(1.5)$ and adjusted P-value < 0.05. Gene ontology enrichment analysis was performed by using clusterProfiler R package (version 3.16.1) (*3*). Gene set enrichment analyses (GSEA) (*4*) were carried out by using the "gseGO" function in

R package clusterProfiler focusing on the biological process gene sets from MSigDB database (from msigdbr_7.4.1) (*5, 6*) and utilizing the default parameters and adjusted *P*-value < 0.05.

Pseudo-time trajectory analysis on FSHD-affected nuclei

All nuclei in FSHD-affected clusters were organized in pseudo-time using Monocle2 (Version 2.16.0) (*7*). The representative genes of FSHD-affected clusters were identified by using the "FindMarkers" function from the Seurat package. The marker genes were then used for ordering cells by DDRTree method and reverse graph embedding.

Signature scoring analysis

For gene scoring analysis, gene sets were obtained from the MSigDB database. The "AddModuleScore" function in the Seurat R package was then used to calculate the signature score of each gene set in each nucleus.

DUXA and DNMT3L module signature scores were generated based on the module gene lists from Meistermann et al. (567 genes for the DUXA module and 261 genes for the DNMT3L module) (*8*). Module score were calculated as described above.

The marker genes of all cell types during early embryogenesis were determined by the function "FindAllMarkers" in R package Seurat. Only up-regulated genes were selected as significant marker genes according to the thresholds: avg_log2(FC) > log2(1.5) and adjusted *P*-value < 0.05. Embryonic stage signatures scores were calculated based on the marker genes of each cell type during early embryogenesis as described above.

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its Supplementary material and information. Additional data can be provided by the corresponding author upon reasonable request. All snRNA-seq data along with their associated metadata will be deposited in the EGA database. The single-cell RNA-seq data of the human embryonic cells was collected from published GEO series GSE36552. All supporting codes and scripts will be deposited in GitLab (https://git.lumc.nl/avandenheuvel1/snrna-seq-in-multinucleated-myogenic-fshd-cells).

Acknowledgements

We thank all FSHD families for participating in our studies.

Conflict of Interest statement. S.v.d.M. has acted as consultant and/or is a member of the advisory board for Avidity Biosciences, Dyne Therapeutics and Fulcrum Therapeutics and is a Board member for Renogenyx.

Contribution. A.v.d.H. and S.v.d.M. designed and conceptualized this study. A.v.d.H. and A.W. designed and performed the experiments of nuclei isolation. D.Z. and I.W. performed the experiments. A.v.d.H. and S.K. performed sequencing data processing. D.Z. and A.v.d.H. analyzed the data. S.v.d.M., S.J.T., A.M., and J.B. contributed with the interpretation of results. D.Z. and A.v.d.H. drafted the manuscript. S.v.d.M., and A.M. revised the manuscript for intellectual content. All authors read and approved the final manuscript.

Funding

Our work is supported by the Prinses Beatrix Spierfonds [W.OR19-06].

References

1. Lemmers, R.J., Van der Vliet, P.J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., Jan van Ommen, G. and Padberg, G.W. (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. Science, 329, 1650-1653.
2. Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R. and Wilcox, A.L. (2017) Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nature genetics, 49, 925-934.
3. Hamanaka, K., Šikrová, D., Mitsuhashi, S., Masuda, H., Sekiguchi, Y., Sugiyama, A., Shibuya, K., Lemmers, R.J., Goossens, R. and Ogawa, M. (2020) Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. Neurology, 94, e2441-e2447.
4. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., Van Der Vliet, P.J., Almomani, R. and Straasheijm, K.R. (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics, 44, 1370-1374.
5. van den Boogaard, M.L., Lemmers, R.J., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F. and Kriek, M. (2016) Mutations in DNMT3B modify epigenetic repression of the D4Z4 repeat and the penetrance of facioscapulohumeral dystrophy. The American Journal of Human Genetics, 98, 1020-1029.
6. Bosnakovski, D., Xu, Z., Ji Gang, E., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A. and Coppée, F. (2008) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. The EMBO journal, 27, 2766-2779.
7. Bosnakovski, D., Gearhart, M.D., Toso, E.A., Ener, E.T., Choi, S.H. and Kyba, M. (2018) Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. Scientific reports, 8, 1-12.
8. Knopp, P., Krom, Y.D., Banerji, C.R., Panamarova, M., Moyle, L.A., den Hamer, B., van der Maarel, S.M. and Zammit, P.S. (2016) DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. Journal of cell science, 129, 3816-3831.
9. Bosnakovski, D., Toso, E.A., Hartweck, L.M., Magli, A., Lee, H.A., Thompson, E.R., Dandapat, A., Perlingeiro, R.C. and Kyba, M. (2017) The DUX4 homeodomains mediate inhibition of myogenesis and are functionally exchangeable with the Pax7 homeodomain. Journal of cell science, 130, 3685-3697.
10. Bosnakovski, D., Choi, S.H., Strasser, J.M., Toso, E.A., Walters, M.A. and Kyba, M. (2014) High-throughput screening identifies inhibitors of DUX4-induced myoblast toxicity. Skeletal muscle, 4, 1-11.
11. Turki, A., Hayot, M., Carnac, G., Pillard, F., Passerieux, E., Bommart, S., de Mauverger, E.R., Hugon, G., Pincemail, J. and Pietri, S. (2012) Functional muscle impairment in

facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. Free Radical Biology and Medicine, 53, 1068-1079.
12. Winokur, S.T., Barrett, K., Martin, J.H., Forrester, J.R., Simon, M., Tawil, R., Chung, S.-A., Masny, P.S. and Figlewicz, D.A. (2003) Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. Neuromuscular Disorders, 13, 322-333.
13. Dmitriev, P., Saada, Y.B., Dib, C., Ansseau, E., Barat, A., Hamade, A., Dessen, P., Robert, T., Lazar, V. and Louzada, R.A. (2016) DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. Free Radical Biology and Medicine, 99, 244-258.
14. Bou Saada, Y., Dib, C., Dmitriev, P., Hamade, A., Carnac, G., Laoudj-Chenivesse, D., Lipinski, M. and Vassetzky, Y.S. (2016) Facioscapulohumeral dystrophy myoblasts efficiently repair moderate levels of oxidative DNA damage. Histochemistry and cell biology, 145, 475-483.
15. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Mattéotti, C., Arias, C., Corona, E.D., Nuñez, N.G. and Leo, O. (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular Disorders, 17, 611-623.
16. Shadle, S.C., Zhong, J.W., Campbell, A.E., Conerly, M.L., Jagannathan, S., Wong, C.-J., Morello, T.D., van der Maarel, S.M. and Tapscott, S.J. (2017) DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. PLoS Genetics, 13, e1006658.
17. Grow, E.J., Weaver, B.D., Smith, C.M., Guo, J., Stein, P., Shadle, S.C., Hendrickson, P.G., Johnson, N.E., Butterfield, R.J. and Menafra, R. (2021) p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. Nature genetics, 53, 1207-1220.
18. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C. and Tawil, R. (2012) DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell, 22, 38-51.
19. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M. and Tapscott, S.J. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics, 6, e1001181.
20. Tassin, A., Laoudj-Chenivesse, D., Vanderplanck, C., Barro, M., Charron, S., Ansseau, E., Chen, Y.W., Mercier, J., Coppée, F. and Belayew, A. (2013) DUX 4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? Journal of cellular and molecular medicine, 17, 76-89.
21. Tasca, G., Monforte, M., Ottaviani, P., Pelliccioni, M., Frusciante, R., Laschena, F. and Ricci, E. (2016) Magnetic resonance imaging in a large cohort of facioscapulohumeral muscular dystrophy patients: pattern refinement and implications for clinical trials. Annals of neurology, 79, 854-864.
22. Wang, L.H., Friedman, S.D., Shaw, D., Snider, L., Wong, C.-J., Budech, C.B., Poliachik, S.L., Gove, N.E., Lewis, L.M. and Campbell, A.E. (2019) MRI-informed muscle biopsies

correlate MRI with pathology and DUX4 target gene expression in FSHD. Human Molecular Genetics, 28, 476-486.
23. Yao, Z., Snider, L., Balog, J., Lemmers, R.J., Van Der Maarel, S.M., Tawil, R. and Tapscott, S.J. (2014) DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Human molecular genetics, 23, 5342-5352.
24. van den Heuvel, A., Lassche, S., Mul, K., Greco, A., San León Granado, D., Heerschap, A., Küsters, B., Tapscott, S.J., Voermans, N.C. and van Engelen, B.G. (2022) Facioscapulohumeral dystrophy transcriptome signatures correlate with different stages of disease and are marked by different MRI biomarkers. Scientific Reports, 12, 1426.
25. Beermann, M.L., Homma, S. and Miller, J.B. (2022) Proximity ligation assay to detect DUX4 protein in FSHD1 muscle: a pilot study. BMC Research Notes, 15, 163.
26. van den Heuvel, A., Mahfouz, A., Kloet, S.L., Balog, J., van Engelen, B.G., Tawil, R., Tapscott, S.J. and van der Maarel, S.M. (2019) Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. Human molecular genetics, 28, 1064-1075.
27. Jiang, S., Williams, K., Kong, X., Zeng, W., Nguyen, N.V., Ma, X., Tawil, R., Yokomori, K. and Mortazavi, A. (2020) Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. PLoS genetics, 16, e1008754.
28. Balog, J., Thijssen, P.E., Shadle, S., Straasheijm, K.R., van der Vliet, P.J., Krom, Y.D., van den Boogaard, M.L., de Jong, A., F Lemmers, R.J. and Tawil, R. (2015) Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. Epigenetics, 10, 1133-1142.
29. Rickard, A.M., Petek, L.M. and Miller, D.G. (2015) Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Human molecular genetics, 24, 5901-5914.
30. Jagannathan, S., Shadle, S.C., Resnick, R., Snider, L., Tawil, R.N., van der Maarel, S.M., Bradley, R.K. and Tapscott, S.J. (2016) Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Human Molecular Genetics, 25, 4419-4431.
31. Banerji, C.R., Panamarova, M., Hebaishi, H., White, R.B., Relaix, F., Severini, S. and Zammit, P.S. (2017) PAX7 target genes are globally repressed in facioscapulohumeral muscular dystrophy skeletal muscle. Nature communications, 8, 2152.
32. Banerji, C.R. (2020) PAX7 target gene repression associates with FSHD progression and pathology over 1 year. Human Molecular Genetics, 29, 2124-2133.
33. Kim, T., Ahmad, K., Shaikh, S., Jan, A.T., Seo, M.-G., Lee, E.J. and Choi, I. (2019) Dermatopontin in skeletal muscle extracellular matrix regulates myogenesis. Cells, 8, 332.
34. Holterman, C.E., Le Grand, F., Kuang, S., Seale, P. and Rudnicki, M.A. (2007) Megf10 regulates the progression of the satellite cell myogenic program. The Journal of cell biology, 179, 911-922.
35. Polesskaya, A., Seale, P. and Rudnicki, M.A. (2003) Wnt signaling induces the myogenic specification of resident CD45+ adult stem cells during muscle regeneration. Cell, 113, 841-852.
36. Heher, P., Ganassi, M., Weidinger, A., Engquist, E.N., Pruller, J., Nguyen, T.H., Tassin, A., Decleves, A.-E., Mamchaoui, K. and Banerji, C.R. (2022) Interplay between mitochondrial

reactive oxygen species, oxidative stress and hypoxic adaptation in facioscapulohumeral muscular dystrophy: Metabolic stress as potential therapeutic target. Redox Biology, 51, 102251.
37. Wagatsuma, A. and Sakuma, K. (2013) Mitochondria as a potential regulator of myogenesis. The Scientific World Journal, 2013.
38. Celegato, B., Capitanio, D., Pescatori, M., Romualdi, C., Pacchioni, B., Cagnin, S., Viganò, A., Colantoni, L., Begum, S., Ricci, E. et al. (2006) Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. Proteomics, 6, 5303-5321.
39. Zammit, P.S., Relaix, F., Nagata, Y., Ruiz, A.P.r., Collins, C.A., Partridge, T.A. and Beauchamp, J.R. (2006) Pax7 and myogenic progression in skeletal muscle satellite cells. Journal of Cell Science, 119, 1824-1832.
40. von Maltzahn, J., Jones, A.E., Parks, R.J. and Rudnicki, M.A. (2013) Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. Proceedings of the National Academy of Sciences, 110, 16474-16479.
41. Yoshihara, M., Kirjanov, I., Nykänen, S., Sokka, J., Weltner, J., Lundin, K., Gawriyski, L., Jouhilahti, E.-M., Varjosalo, M. and Tervaniemi, M.H. (2022) Transient DUX4 expression in human embryonic stem cells induces blastomere-like expression program that is marked by SLC34A2. Stem cell reports, 17, 1743-1756.
42. Moya-Jódar, M., Ullate-Agote, A., Barlabé, P., Rodríguez-Madoz, J.R., Abizanda, G., Barreda, C., Carvajal-Vergara, X., Vilas-Zornoza, A., Romero, J.P. and Garate, L. (2023) Revealing cell populations catching the early stages of human embryo development in naive pluripotent stem cell cultures. Stem Cell Reports, 18, 64-80.
43. Ashoti, A., Alemany, A., Sage, F. and Geijsen, N. (2021) DUX4 induces a homogeneous sequence of molecular changes, culminating in the activation of a stem-cell-like transcriptional network and induction of apoptosis in somatic cells. bioRxiv, in press., 2021.2005. 2004.442407.
44. Campbell, A.E., Belleville, A.E., Resnick, R., Shadle, S.C. and Tapscott, S.J. (2018) Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle. Human molecular genetics, 27, R153-R162.
45. Mocciaro, E., Runfola, V., Ghezzi, P., Pannese, M. and Gabellini, D. (2021) DUX4 role in normal physiology and in FSHD muscular dystrophy. Cells, 10, 3322.
46. Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X. and Yan, J. (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature structural & molecular biology, 20, 1131-1139.
47. Meistermann, D., Bruneau, A., Loubersac, S., Reignier, A., Firmin, J., François-Campion, V., Kilens, S., Lelièvre, Y., Lammers, J. and Feyeux, M. (2021) Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. Cell Stem Cell, 28, 1625-1640. e1626.
48. Dang, Y., Yan, L., Hu, B., Fan, X., Ren, Y., Li, R., Lian, Y., Yan, J., Li, Q. and Zhang, Y. (2016) Tracing the expression of circular RNAs in human pre-implantation embryos. Genome biology, 17, 1-15.

49. Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.-y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L. and Sun, Y.E. (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature, 500, 593-597.
50. Resnick, R., Wong, C.-J., Hamm, D.C., Bennett, S.R., Skene, P.J., Hake, S.B., Henikoff, S., van der Maarel, S.M. and Tapscott, S.J. (2019) DUX4-induced histone variants H3. X and H3. Y mark DUX4 target genes for expression. Cell reports, 29, 1812-1820. e1815.
51. Hu, K. and Yu, Y. (2017) Metabolite availability as a window to view the early embryo microenvironment in vivo. Molecular Reproduction and Development, 84, 1027-1038.
52. Khurana, N.K. and Niemann, H. (2000) Energy metabolism in preimplantation bovine embryos derived in vitro or in vivo. Biology of reproduction, 62, 847-856.
53. Sasaki-Honda, M., Jonouchi, T., Arai, M., Hotta, A., Mitsuhashi, S., Nishino, I., Matsuda, R. and Sakurai, H. (2018) A patient-derived iPSC model revealed oxidative stress increases facioscapulohumeral muscular dystrophy-causative DUX4. Human Molecular Genetics, 27, 4024-4035.
54. Barro, M., Carnac, G., Flavier, S., Mercier, J., Vassetzky, Y. and Laoudj-Chenivesse, D. (2010) Myoblasts from affected and non-affected FSHD muscles exhibit morphological differentiation defects. Journal of cellular and molecular medicine, 14, 275-289.
55. Shintaku, J., Peterson, J.M., Talbert, E.E., Gu, J.-M., Ladner, K.J., Williams, D.R., Mousavi, K., Wang, R., Sartorelli, V. and Guttridge, D.C. (2016) MyoD regulates skeletal muscle oxidative metabolism cooperatively with alternative NF-κB. Cell reports, 17, 514-526.
56. Ferreboeuf, M., Mariot, V., Furling, D., Butler-Browne, G., Mouly, V. and Dumonceaux, J. (2014) Nuclear protein spreading: implication for pathophysiology of neuromuscular diseases. Human molecular genetics, 23, 4125-4133.
57. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019) Comprehensive integration of single-cell data. Cell, 177, 1888-1902. e1821.
58. McCarthy, D.J., Campbell, K.R., Lun, A.T. and Wills, Q.F. (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics, 33, 1179-1186.
59. Yu, G., Wang, L.-G., Han, Y. and He, Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 16, 284-287.
60. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R. and Lander, E.S. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102, 15545-15550.
61. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. and Mesirov, J.P. (2011) Molecular signatures database (MSigDB) 3.0. Bioinformatics, 27, 1739-1740.
62. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. (2015) The molecular signatures database hallmark gene set collection. Cell systems, 1, 417-425.

63. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A. and Trapnell, C. (2017) Reversed graph embedding resolves complex single-cell trajectories. Nature methods, 14, 979-982.

Supplementary figures

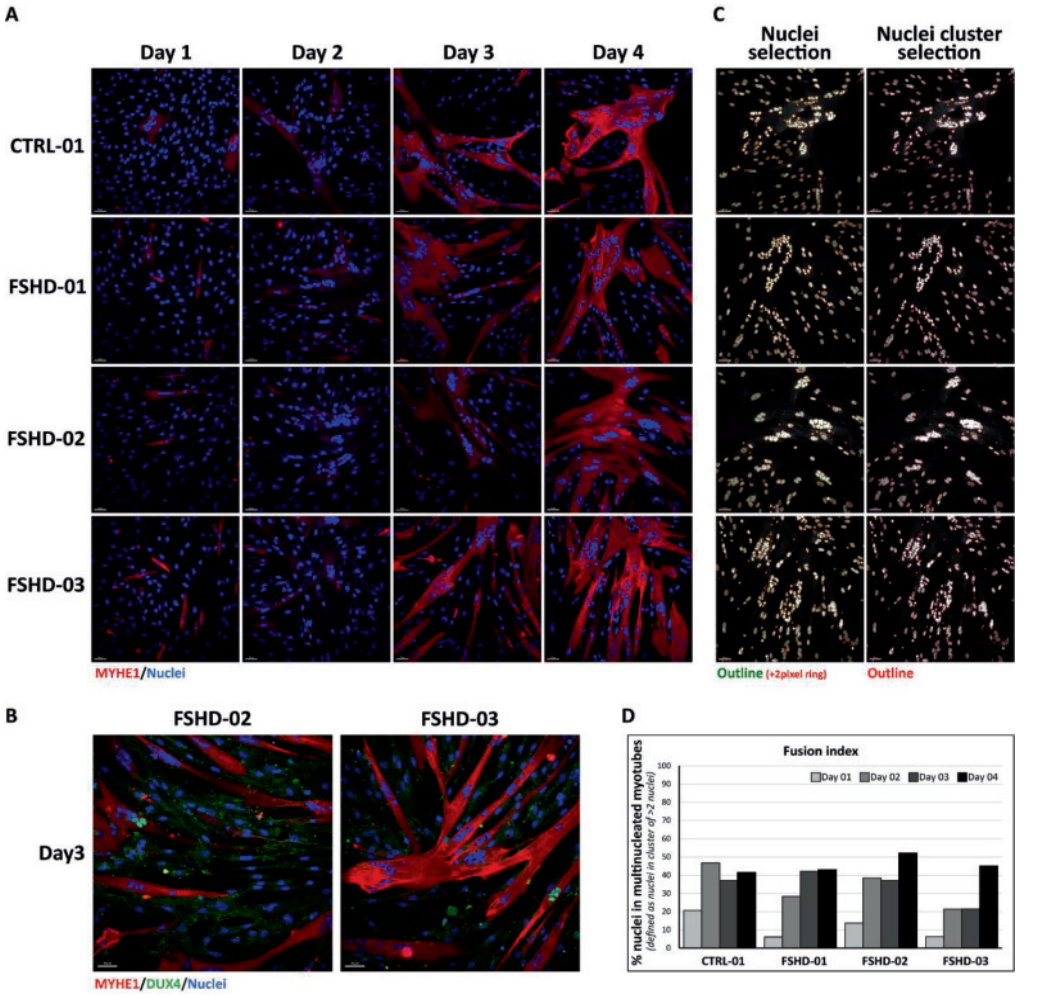

Figure S1. Morphology of differentiated myotubes of cell lines used for single-nucleus RNA-seq.

A) Immunofluorescence staining for MF20 (red) and nuclei (blue) in control and FSHD myotubes after 1, 2, 3, and 4 days of myogenic differentiation. Scale bar: 50 µm. **B)** Immunofluorescence staining for MF20 (red), DUX4 (green), and nuclei (blue) in FSHD myotubes after 3 days of myogenic differentiation. Representative images were selected to show the detection of DUX4 in cluster of nuclei within myotubes. Scale bar: 50 µm. **C)** Examples of nuclei and nuclei cluster selection with the CellProfiler software for fusion index calculation (see methods section). Day 4 differentiation images are shown. **D**) Fusion index for each cell line on differentiation day 1-4.

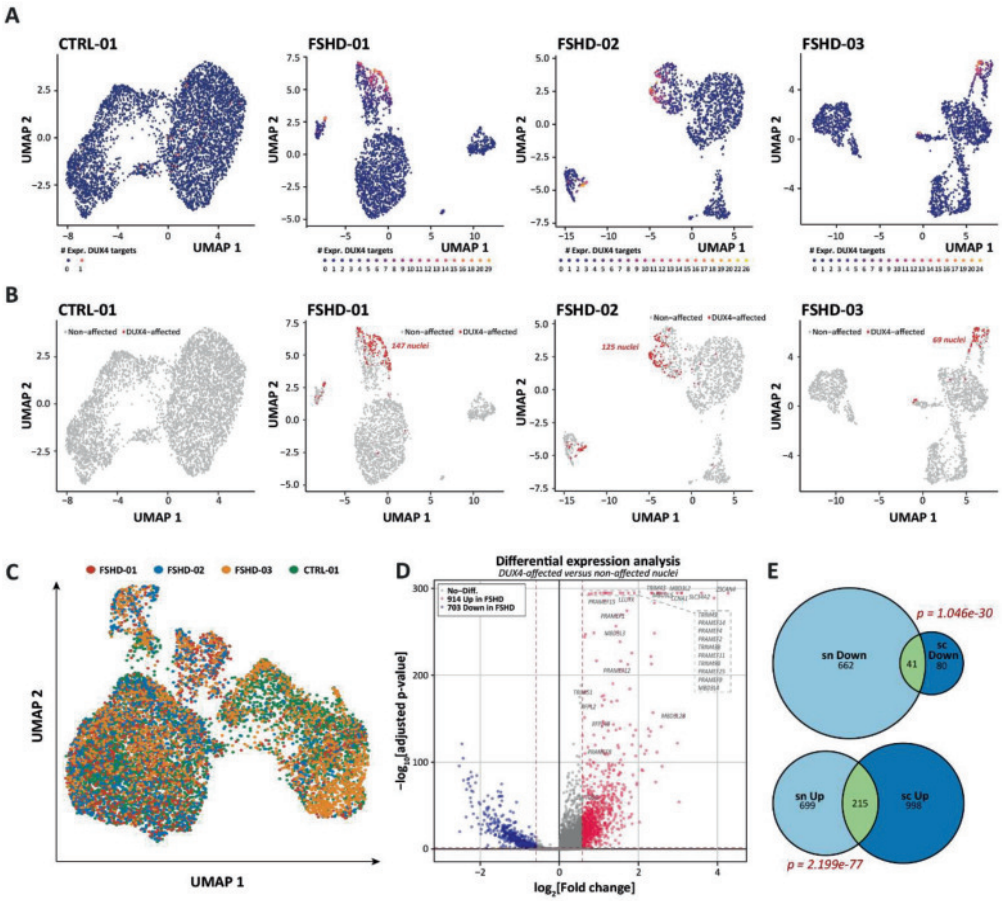

Figure S2. DUX4-affected nuclei classification and validation.

A) UMAP plots showing the number of expressed DUX4 targets in each nucleus. Each point depicts a single nucleus, colored according to the number of expressed DUX4 target genes **B)** UMAP plots as in A) depicting the final nuclei classification of DUX4-affected nuclei (in red, expressing DUX4 and/or ≥ 5 DUX4 target genes) versus non-affected nuclei (in grey). **C)** UMAP of the integrated dataset of all four samples, color-coded for sample ID, indicating contribution of each sample in each nuclei cluster, with a reduced contribution of the control sample in the DUX4-affected clusters I and II. **D)** Volcanoplot depicting the differential expression results of comparing all DUX4-affected nuclei versus all non-affected nuclei in the four samples. The red dashed lines indicate the thresholds for DE gene selection (fold change >1.5 or <-1.5 and adjusted *p*-value < 0.05). Significantly differentially expressed DUX4 target genes are labeled. **E)** Venn diagrams depicting the overlap in differentially expressed genes between DUX4-affected and non-affected nuclei in this study and that from our previously published scRNA-seq data in primary mononuclear myocytes (van den Heuvel et al. Hum. Mol. Genet. 2019). *P*-values depict the result of a hypergeometric test calculated with the online calculator from the Graeber Lab; https://systems.crump.ucla.edu/hypergeometric/.

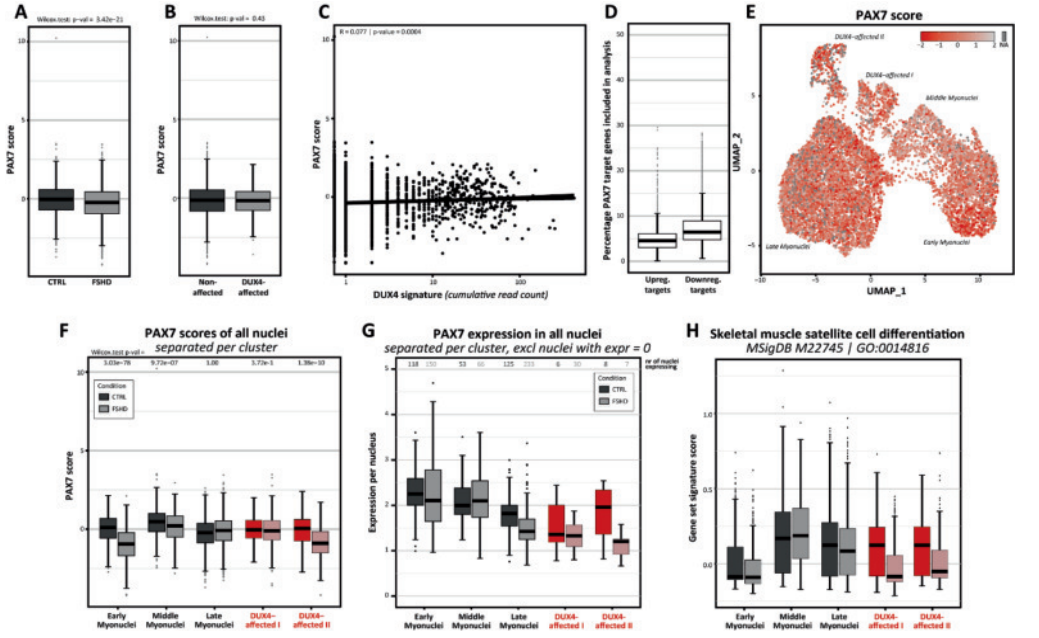

Figure S3. The PAX7 score is reduced in a subpopulation of early FSHD myonuclei, independent of DUX4 activation.

A) PAX7 score in FSHD and control nuclei. **B)** PAX7 score in DUX4-affected and non-affected nuclei. **C)** Linear correlation between PAX7 score and DUX4 signature score. *P*-values depict the result of a Pearson correlation. **D)** Percentage of PAX7 target genes with detected reads per nucleus. **E)** UMAP plot of the integrated dataset of all four samples, color-coded for the PAX7 score in each nucleus. **F)** PAX7 scores in FSHD and control nuclei grouped by cluster. **G)** PAX7 expression levels in FSHD and control nuclei grouped per cluster. **H)** Gene set signature score in FSHD and control nuclei, grouped per cluster.

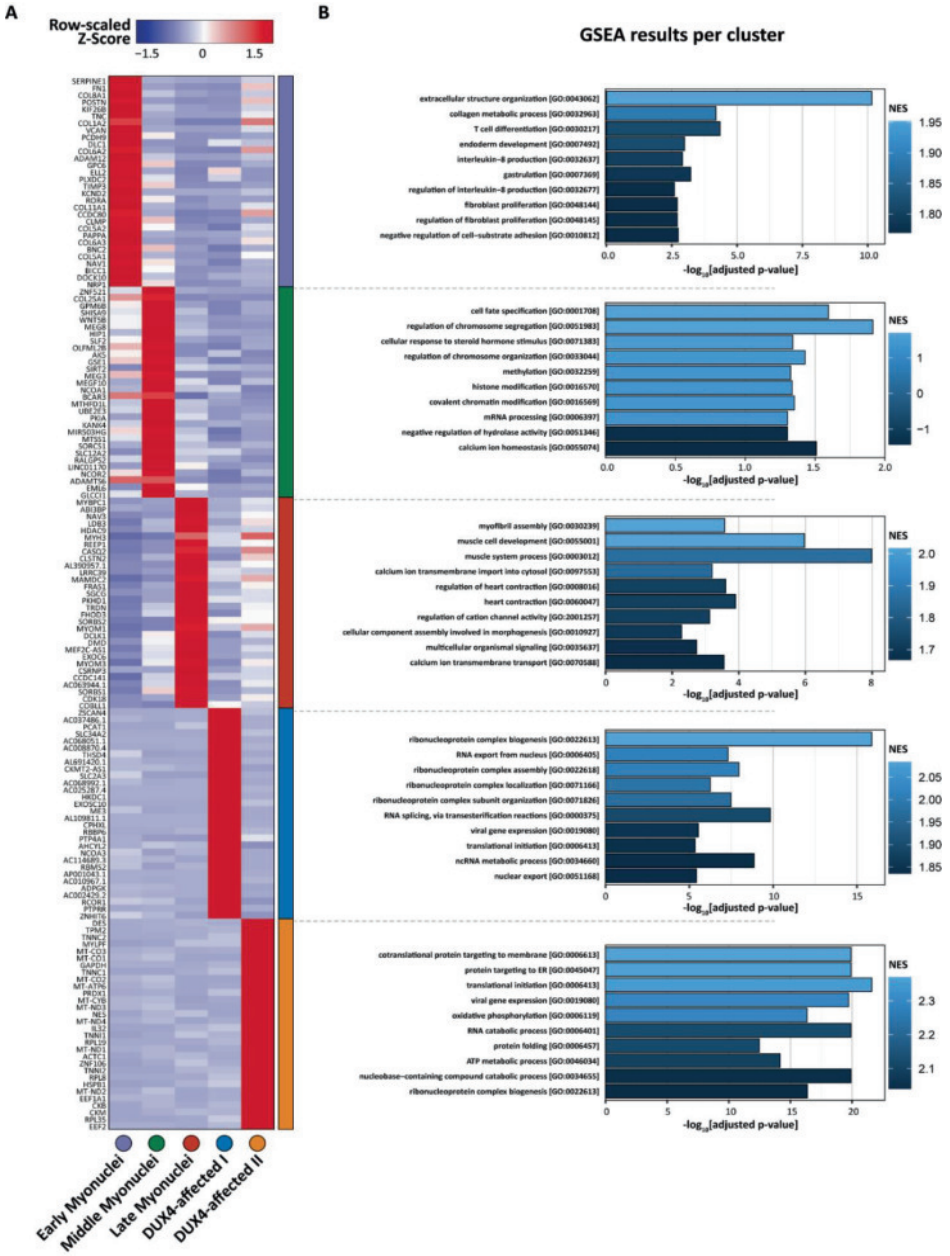

Figure S4. Cluster classification of the integrated snRNA-seq dataset from late multinuclear myotubes.

A) Expression heatmap showing the top 30 representative genes of each cluster. **B)** Bar graph displaying the top 10 statistically significantly enriched biological processes (GO terms) in each cluster (adjusted *p*-value <0.05).

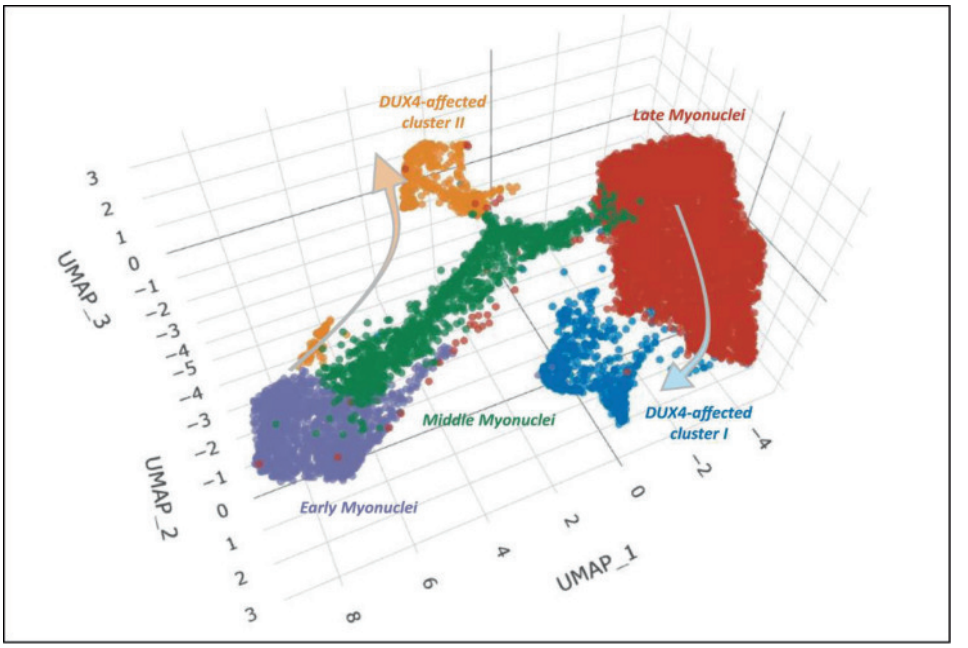

Figure S5. 3D UMAP visualization of the integrated snRNA-seq dataset from multinuclear myotubes.

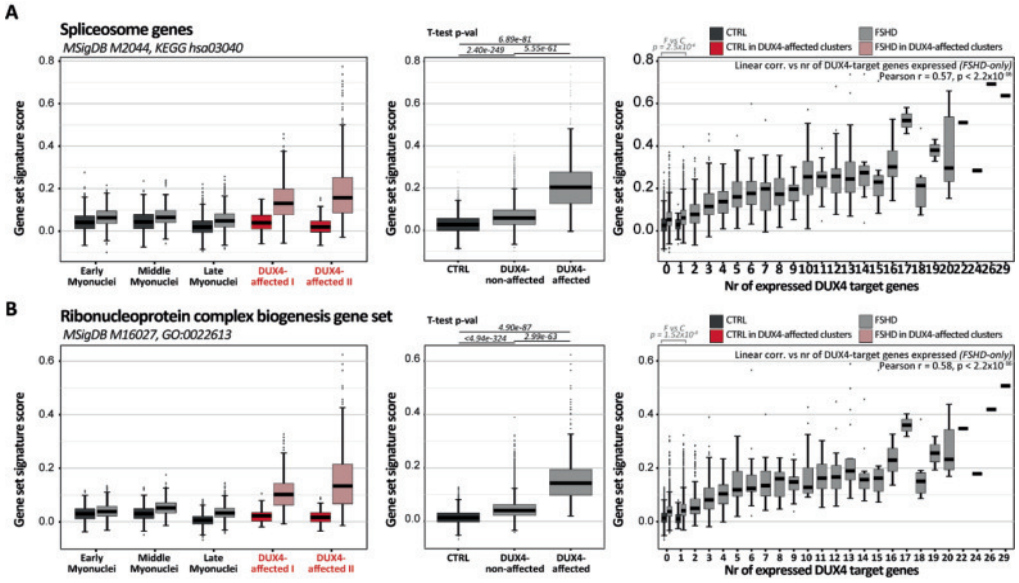

Figure S6. Several direct DUX4-induced responses correlate with DUX4 signature activation, independent of the cluster they are located in.

A-B) Gene set signature scores in FSHD and control nuclei, grouped per cluster (left), per DUX4-affected state (middle) or per nr of DUX4-target genes expressed (right), for a A) KEGG Spliceosome gene set (MSigDB M2044, KEGG;hsa03040), B) Ribonucleoprotein complex biogenesis gene set (MSigDB M16027, GO:0022613). MSigDB; Molecular Signature Database, KEGG; Kyoto Encyclopedia of Genes and Genomes, GO; Gene Ontology.

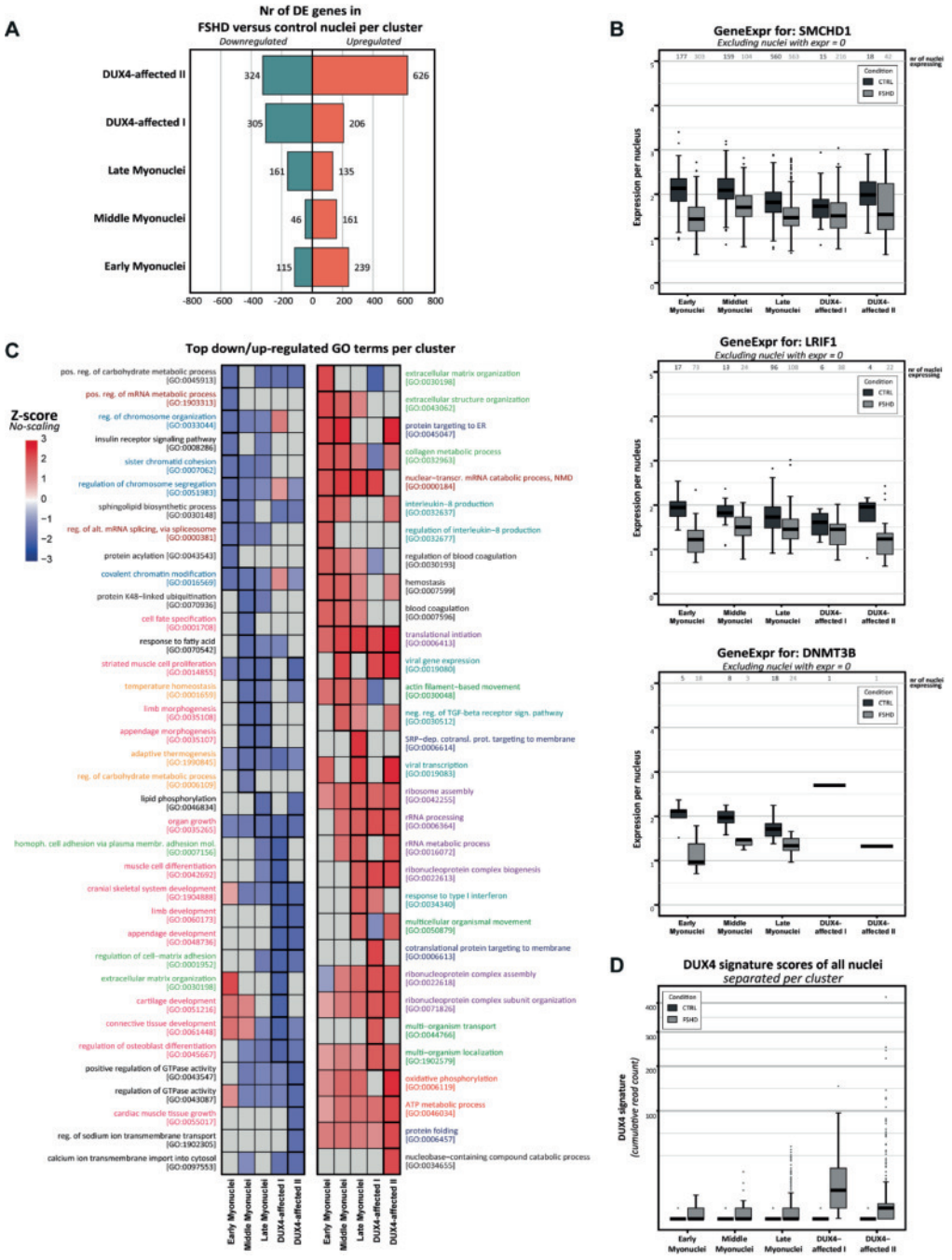
A
Nr of DE genes in
FSHD versus control nuclei per cluster
Downregulated
Upregulated
DUX4-affected II
DUX4-affected I
Late Myonuclei
Middle Myonuclei
Early Myonuclei
324
626
305
206
161
135
46
161
115
239
-800 -600 -400 -200 0 200 400 600 800
B
GeneExpr for: SMCHD1
Excluding nuclei with expr = 0
GeneExpr for: LRIF1
Excluding nuclei with expr = 0
GeneExpr for: DNMT3B
Excluding nuclei with expr = 0
Expression per nucleus
Condition
CTRL
FSHD
Early Myonuclei
Middle Myonuclei
Late Myonuclei
DUX4-affected I
DUX4-affected II
C
Top down/up-regulated GO terms per cluster
Z-score
No-scaling
pos. reg. of carbohydrate metabolic process [GO:0045913]
pos. reg. of mRNA metabolic process [GO:1903313]
reg. of chromosome organization [GO:0033044]
insulin receptor signaling pathway [GO:0008286]
sister chromatid cohesion [GO:0007062]
regulation of chromosome segregation [GO:0051983]
sphingolipid biosynthetic process [GO:0030148]
reg. of alt. mRNA splicing, via spliceosome [GO:0000381]
protein acylation [GO:0043543]
covalent chromatin modification [GO:0016569]
protein K48-linked ubiquitination [GO:0070936]
cell fate specification [GO:0001708]
response to fatty acid [GO:0070542]
striated muscle cell proliferation [GO:0014855]
temperature homeostasis [GO:0001659]
limb morphogenesis [GO:0035108]
appendage morphogenesis [GO:0035107]
adaptive thermogenesis [GO:1990845]
reg. of carbohydrate metabolic process [GO:0006109]
lipid phosphorylation [GO:0046834]
organ growth [GO:0035265]
homoph. cell adhesion via plasma membr. adhesion mol. [GO:0007156]
muscle cell differentiation [GO:0042692]
cranial skeletal system development [GO:1904888]
limb development [GO:0060173]
appendage development [GO:0048736]
regulation of cell-matrix adhesion [GO:0001952]
extracellular matrix organization [GO:0030198]
cartilage development [GO:0051216]
connective tissue development [GO:0061448]
regulation of osteoblast differentiation [GO:0045667]
positive regulation of GTPase activity [GO:0043547]
regulation of GTPase activity [GO:0043087]
cardiac muscle tissue growth [GO:0055017]
reg. of sodium ion transmembrane transport [GO:1902305]
calcium ion transmembrane import into cytosol [GO:0097553]
extracellular matrix organization [GO:0030198]
extracellular structure organization [GO:0043062]
protein targeting to ER [GO:0045047]
collagen metabolic process [GO:0032963]
nuclear-transcr. mRNA catabolic process, NMD [GO:0000184]
interleukin-8 production [GO:0032637]
regulation of interleukin-8 production [GO:0032677]
regulation of blood coagulation [GO:0030193]
hemostasis [GO:0007599]
blood coagulation [GO:0007596]
translational initiation [GO:0006413]
viral gene expression [GO:0019080]
actin filament-based movement [GO:0030048]
reg. reg. of TGF-beta receptor sign. pathway [GO:0030512]
SRP-dep. cotransl. prot. targeting to membrane [GO:0006614]
viral transcription [GO:0019083]
ribosome assembly [GO:0042255]
rRNA processing [GO:0006364]
rRNA metabolic process [GO:0016072]
ribonucleoprotein complex biogenesis [GO:0022613]
response to type I interferon [GO:0034340]
multicellular organismal movement [GO:0050879]
cotranslational protein targeting to membrane [GO:0006613]
ribonucleoprotein complex assembly [GO:0022618]
ribonucleoprotein complex subunit organization [GO:0071826]
multi-organism transport [GO:0044766]
multi-organism localization [GO:1902579]
oxidative phosphorylation [GO:0006119]
ATP metabolic process [GO:0046034]
protein folding [GO:0006457]
nucleobase-containing compound catabolic process [GO:0034655]
Early Myonuclei
Middle Myonuclei
Late Myonuclei
DUX4-affected I
DUX4-affected II
D
DUX4 signature scores of all nuclei
separated per cluster
DUX4 signature
(cumulative read count)
Condition
CTRL
FSHD
Early Myonuclei
Middle Myonuclei
Late Myonuclei
DUX4-affected I
DUX4-affected II

Figure S7. Global FSHD-associated responses in all 5 clusters.

A) Number of DE genes in FSHD versus control nuclei per cluster. **B)** Gene expression of three known FSHD regulators (SMCHD1, LRIF1 and DNMT3B) in FSHD and control nuclei per cluster. **C)** Heatmap of GSEA analysis results of DE genes in FSHD versus control nuclei per cluster. Highlighted are the top GO-terms for each cluster. GO-term text colors indicate groups of related GO-terms. **D)** DUX4 signature scores in FSHD and control nuclei per cluster.

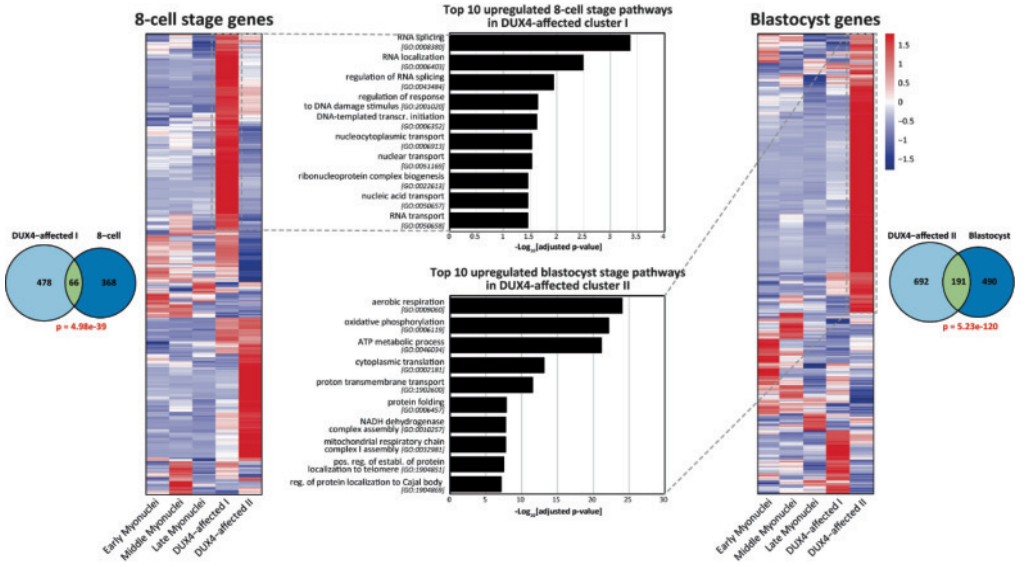

Figure S8. Pathways involved in the embryonic-like transcriptome program in DUX4-affected clusters are mostly involved in transcription initiation, RNA processing, protein production and energy metabolism.

This figure depicts an expression heatmap of the individual embryonic stage signature genes (i.e., left; 8-cell stage signature, right; blastocyst stage signature) in the five clusters of our myotube snRNA-seq dataset. The bar charts in the middle indicate the Top-10 enriched pathways in the most highly expressed genes in the respective signatures; the 8-cell stage signature in DUX4-affected cluster I and the blastocyst stage signature in DUX4-affected cluster II. The Venn diagrams depict the signature gene set overlap between the DUX4-affected cluster I and the 8-cell stage signature (left) and between the DUX4-affected cluster II and the blastocyst stage signature (right). *P*-values depict the result of a hypergeometric test calculated with the online calculator from the Graeber Lab;https://systems.crump.ucla.edu/hypergeometric/.

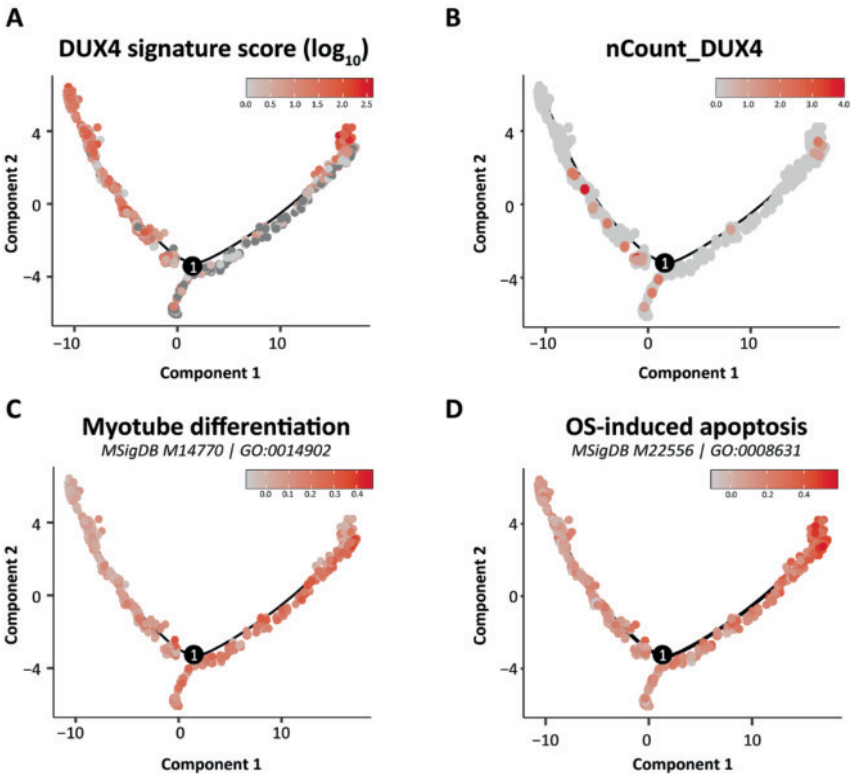

Figure S9. A pseudotime trajectory analysis representing a heterogeneous mix of DUX4-induced responses.

A-C) Pseudotime trajectory analysis of all nuclei in DUX4-affected cluster I and II (as in Figure 5), **A)** color-coded for the DUX4 signature expression, **B)** color-coded for DUX4 expression itself, **C)** color-coded for Myotube differentiation gene set signature score (MSigDB M14770/GO:0014902) and **D)** color-coded for oxidative stress-induced intrinsic apoptosis gene set signature score (MSigDB M22556/GO:0008631). MSigDB; Molecular Signature Database. GO; Gene Ontology.