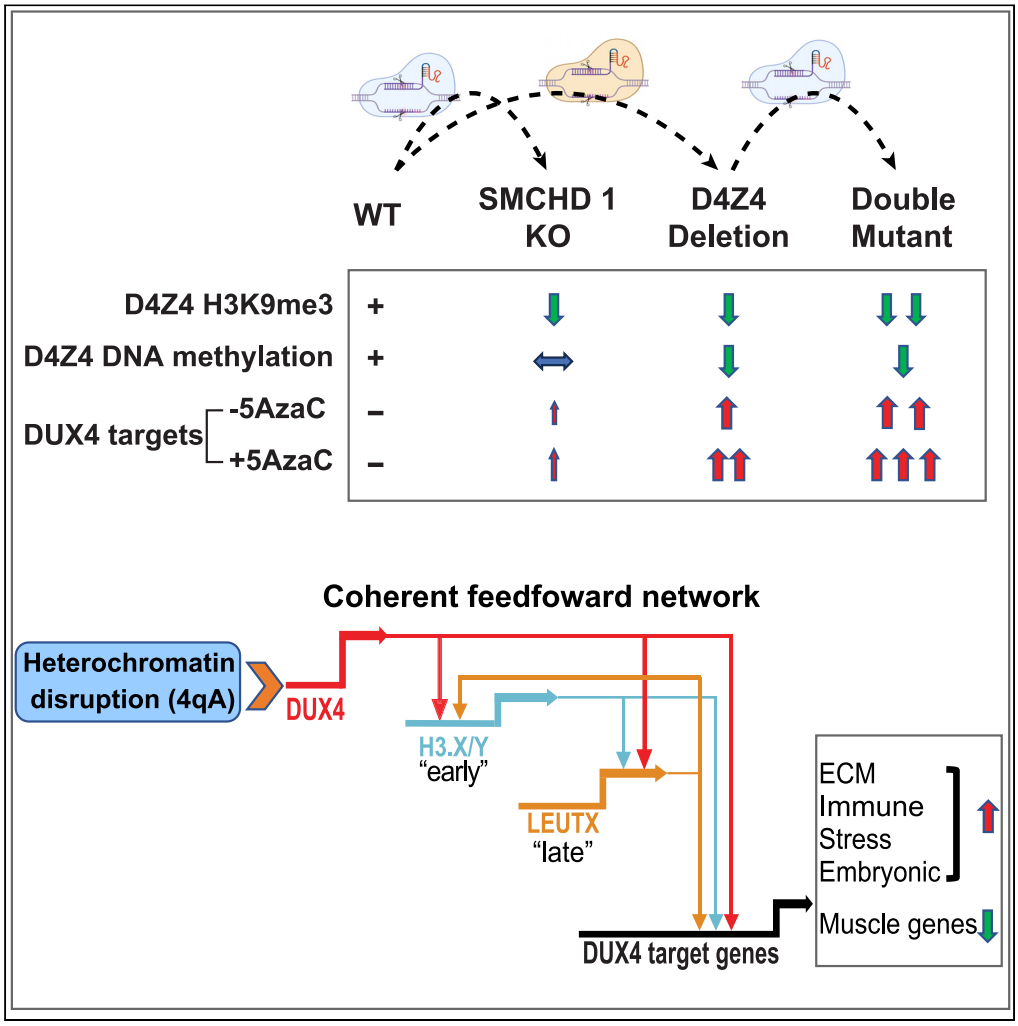
iScience

CellPress
OPEN ACCESS

Article

Engineered FSHD mutations results in D4Z4 heterochromatin disruption and feedforward DUX4 network activation



Xiangduo Kong, Nam Viet Nguyen, Yumeng Li, ..., Rabi Tawil, Ali Mortazavi, Kyoko Yokomori

ali.mortazavi@uci.edu (A.M.)
kyokomor@uci.edu (K.Y.)

Highlights

D4Z4 repeat removal causes variegated DUX4 gene network activation in human myotubes

SMCHD1 mutation or 5AzaC synergizes with D4Z4 deletion for DUX4 target activation

SMCHD1 mutation reduces H3K9me3 in adult myocytes

H3.X/Y and LEUTX enhance DUX4 target expression through feedforward mechanism

Kong et al., iScience *27*, 109357
April 19, 2024
https://doi.org/10.1016/j.isci.2024.109357

iScience

CellPress
OPEN ACCESS

Article

Engineered FSHD mutations results in D4Z4 heterochromatin disruption and feedforward DUX4 network activation

Xiangduo Kong,[1,7] Nam Viet Nguyen,[1,7] Yumeng Li,[1] Jasmine Shaaban Sakr,[2] Kate Williams,[2] Sheila Sharifi,[1] Jonathan Chau,[1] Altay Bayrakci,[1] Seiya Mizuno,[3] Satoru Takahashi,[3,4] Tohru Kiyono,[5] Rabi Tawil,[6] Ali Mortazavi,[2,*] and Kyoko Yokomori[1,8,*]

SUMMARY

Facioscapulohumeral dystrophy (FSHD) is linked to contraction of D4Z4 repeats on chromosome 4q with *SMCHD1* mutations acting as a disease modifier. D4Z4 heterochromatin disruption and abnormal upregulation of the transcription factor DUX4, encoded in the *D4Z4* repeat, are the hallmarks of FSHD. However, defining the precise effect of D4Z4 contraction has been difficult because D4Z4 repeats are primate-specific and DUX4 expression is very rare in highly heterogeneous patient myocytes. We generated isogenic mutant cell lines harboring D4Z4 and/or *SMCHD1* mutations in a healthy human skeletal myoblast line. We found that the mutations affect D4Z4 heterochromatin differently, and that *SMCHD1* mutation or disruption of DNA methylation stabilizes otherwise variegated DUX4 target activation in D4Z4 contraction mutant cells, demonstrating the critical role of modifiers. Our study revealed amplification of the DUX4 signal through downstream targets, H3.X/Y and LEUTX. Our results provide important insights into how rare DUX4 expression leads to FSHD pathogenesis.

INTRODUCTION

Facioscapulohumeral dystrophy (FSHD) is one of the most common muscular dystrophies with a prevalence of ~1 in 8,333. FSHD causes progressive wasting of facial, shoulder, and upper arm as well as lower leg musculature.[1,2] The majority of FSHD cases (>95%) are caused by monoallelic contraction of 3.3 kb D4Z4 macrosatellite repeat sequences located at the subtelomeric region of chromosome 4q (4qter D4Z4) (FSHD1 (MIM 158900)).[3,4] FSHD1 is associated with 1~10 copies of D4Z4 repeats in the contracted allele in contrast to 11–150 copies in the intact allele. Lower copy numbers (1–3 copies) are tied to earlier onset and more severe phenotypes compared to higher copy numbers (8–10 copies). However, clinical manifestations are variable, suggesting a potential contribution of additional modifier gene(s).[2,5] FSHD2, which is the rare form of FSHD (<5% of cases), is mainly linked to mutations of *SMCHD1* (MIM 158901).[2,6] Although FSHD2 was thought previously to involve no D4Z4 repeat contraction, accumulating evidence indicates its association with relatively short D4Z4 repeats (8–20 repeats).[2] Moreover, mutations of the *SMCHD1* gene have been found in severe cases of FSHD1, suggesting that it acts as a modifier gene to increase the disease severity.[7,8] Curiously, however, *SMCHD1* mutations (in some cases the exact same mutations) are also linked to isolated arhinia and Bosma arhinia microphthalmia syndrome, a developmental disorder with phenotypes distinct from FSHD.[9–11] Disease specificities are most likely dictated by the type of mutations as well as the presence of additional modifier genes in the relevant cell types. The direct consequences of FSHD mutations (i.e., D4Z4 contraction and/or SMCHD1 mutation) in myocytes and their contributions to the FSHD patient phenotype have not been fully elucidated.

D4Z4 contains an open reading frame for the double-homeobox transcription factor *DUX4* gene.[12–14] *DUX4* gene is normally expressed early in embryogenesis and participates in zygotic genome activation.[15,16] Expression of short and long isoforms have been detected but only the full-length *DUX4* transcript (*DUX4fl*) that includes a transactivation domain can activate target genes, and its abnormal upregulation is associated with FSHD.[14,17,18] The *DUX4* gene, embedded in the D4Z4 repeat, lacks a poly-adenylation (poly(A)) signal sequence within the repeat, and only those individuals with "permissive" 4qA haplotypes carrying the canonical poly(A) signal downstream of the last D4Z4 repeat

[1]Department of Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, CA, USA
[2]Department of Development and Cell Biology, School of Biological Sciences, University of California, Irvine, Irvine, CA, USA
[3]Laboratory Animal Resource Center in Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki, Japan
[4]Department of Anatomy and Embryology, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan
[5]Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Kashiwa, Chiba, Japan
[6]Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA
[7]These authors contributed equally
[8]Lead contact
*Correspondence: ali.mortazavi@uci.edu (A.M.), kyokomor@uci.edu (K.Y.)
https://doi.org/10.1016/j.isci.2024.109357

develop FSHD, strongly suggesting that the disease is tightly linked to functional *DUX4* mRNA production from the last D4Z4 copy.[17] However, the *DUX4fl* transcript is expressed at extremely low level (sometimes not detectable[5,14]), and DUX4 protein is detectable only in <0.1% of patient muscle cells.[14,17–19] Furthermore, *DUX4fl* expression can occasionally be observed even in unaffected individuals.[5,14] Nevertheless, activation of many DUX4 target genes was observed in patient muscle cells in multiple studies, supporting the significance of the DUX4 target gene network in FSHD, and target genes are used as markers for the disease.[5,13,20–23] How the limited DUX4 expression results in robust target gene activation and drives the disease, however, is not well understood.

Hypomethylation of D4Z4 DNA is a signature change in FSHD patient cells.[24–26] We also found that D4Z4 repeats contain heterochromatic regions marked by histone H3 lysine 9 trimethylation (H3K9me3), heterochromatin binding protein HP1γ and the higher-order chromatin organizer cohesin.[27] This H3K9me3 heterochromatin structure is compromised in both FSHD1 and FSHD2.[27] Indeed, reduction of H3K9me3 at D4Z4 by inhibition or depletion of SUV39H1, a responsible histone methyltransferase, causes *DUX4fl* expression.[27,28] Thus, perturbation of D4Z4 heterochromatin appears to be directly linked to FSHD pathogenesis.[18,29] Curiously, no significant loss of H3K9me3 was observed in cells from a clinically unaffected individual with prominent DNA hypomethylation at D4Z4, indicating that H3K9me3 is not a mere downstream consequence of DNA hypomethylation and raising the possibility that H3K9me3 reduction is prerequisite for the disease development.[27] SMCHD1 binds to D4Z4, and its depletion results in derepression of DUX4fl expression, indicating a direct role of SMCHD1 in DUX4fl regulation.[6] Previously, we demonstrated that SMCHD1 binding to D4Z4 is H3K9me3-dependent.[28] This raised the possibility that even in FSHD1 with no mutation in *SMCHD1*, SMCHD1 binding to D4Z4 may be compromised (due to the loss of H3K9me3).[28] In severe cases of FSHD1, this effect may be further exacerbated by the actual mutations in *SMCHD1* itself.[7,8] SMCHD1 has been implicated in regulation of DNA methylation at CpG islands and at the inactive X chromosome,[30–32] and thus, it is assumed to also regulate DNA methylation at D4Z4. However, SMCHD1 represses gene expression in both DNA methylation-dependent and independent ways.[33] Despite the close correlation between D4Z4 DNA hypomethylation and SMCHD1 mutations in FSHD2 patients, whether SMCHD1 modulates DUX4fl expression through D4Z4 DNA methylation has not been explicitly determined.

Overexpression of recombinant DUX4 in *in vitro* myoblasts and in *in vivo* model organisms is highly toxic,[34,35] and there is an ongoing effort to characterize the mechanism of DUX4-induced cell toxicity.[36–39] However, unlike the recombinant DUX4-inducible systems, the frequency of DUX4 expression in patient cells is often less than 1%. While there is an evidence for cell death in patient myocytes,[23] recent single cell/nucleus-sequencing and *in situ* RNA detection analyses indicated that significant acute cell death is not necessarily a major phenotype in FSHD patient myocytes that endogenously express DUX4.[40,41] Moreover, the subcellular localization of the recombinant *DUX4fl* mRNA differs from that of the endogenous *DUX4fl* mRNA, and not all the cellular phenotypes induced by the recombinant DUX4 overexpression are found in patient cells.[40,42] Collectively, these observations raise the question whether DUX4 overexpression-induced acute cytotoxicity reflects a physiologically relevant mechanism of disease pathogenesis. Thus, efforts are being made to express DUX4 at low level in an inducible fashion in mice.[43,44] D4Z4 repeats (and *DUX4* within), however, are primate-specific. Likewise, DUXA and LEUTX, two main transcription factors activated by DUX4, are absent in mice.[45] Consequently, the genes and molecular network activated by human *DUX4* introduced in non-primate model organisms may be different from those in patient muscle. Patient-derived myoblasts (often immortalized) are important resources for FSHD research. However, availability of high-quality patient myoblasts is limited, and variable genetic backgrounds of control and patient cells make it difficult to tease out disease-specific changes from individual heterogeneity. Myoblasts with or without D4Z4 contraction isolated from mosaic patient muscle are available.[46] However, additional modifier gene mutations that enabled the development of the disease may be present in this patient. Thus, to complement available model systems describe above and to experimentally address the consequences of D4Z4 and SMCHD1 mutations, we generated CRISPR-engineered isogenic mutant myoblast cell lines carrying either D4Z4 mutations, SMCHD1 mutation or both using a control cell line derived from a healthy individual. Analyses of these isogenic mutant cell lines revealed a feedback loop between SMCHD1 and H3K9me3 at D4Z4 and synergistic effect of double mutations on DUX4 target gene induction. Clear stabilization of variegated mutant phenotype by additional heterochromatin disruption highlighted the characteristics of FSHD as a heterochromatin abnormality disorder. Furthermore, our results uncovered differentiation-insensitive and -sensitive DUX4 target genes and their cross-regulation, revealing the complex mechanism of the DUX4 gene network activation. Taken together, our results provide important new insights into FSHD pathogenesis.

RESULTS

Elimination of SMCHD1 has a minor effect on *DUX4* expression

FSHD occurs only in individuals with "permissive" 4qA haplotypes, in which the presence of a poly(A) signal downstream of the last D4Z4 copy would allow the expression of functional DUX4fl mRNA.[17] We first depleted SMCHD1 using shRNA in a healthy control skeletal myoblast cell line of a permissive haplotype (subsequently used as a parental control line) (Figure S1). Partial depletion of SMCHD1 was sufficient to cause upregulation of DUX4 target genes at the myoblast (MB) stage in a statistically significant fashion compared to control shRNA-treated cells (Figures S1A and S1B). Notably, DUX4 target gene expression was further stimulated upon myotube (MT) differentiation (Figure S1B). However, DUX4 target gene expression in SMCHD1-depleted myotubes is much lower than that in FSHD2 myotubes carrying *SMCHD1* mutation (Figure S1C). Next, gRNAs specific for *SMCHD1* were designed for CRISPR knockout (KO) (Figure S2A). Despite screening 300 clones, we failed to obtain heterozygous mutant cells due to high efficiency of CRISPR mutation. Thus, unlike FSHD2 cells, in which *SMCHD1* mutation is heterozygous, our *SMCHD1* (SM) mutant cells are *SMCHD1* null (Figure 1A). As in HCT116 colorectal cancer cells,[47] *SMCHD1* KO in adult myoblasts is not lethal. SM mutation in permissive haplotype cells upregulated DUX4 target genes in a statistically significant fashion compared to the parental control cells (Figure 1D). However, even after a complete loss of SMCHD1, the amount of DUX4 target gene

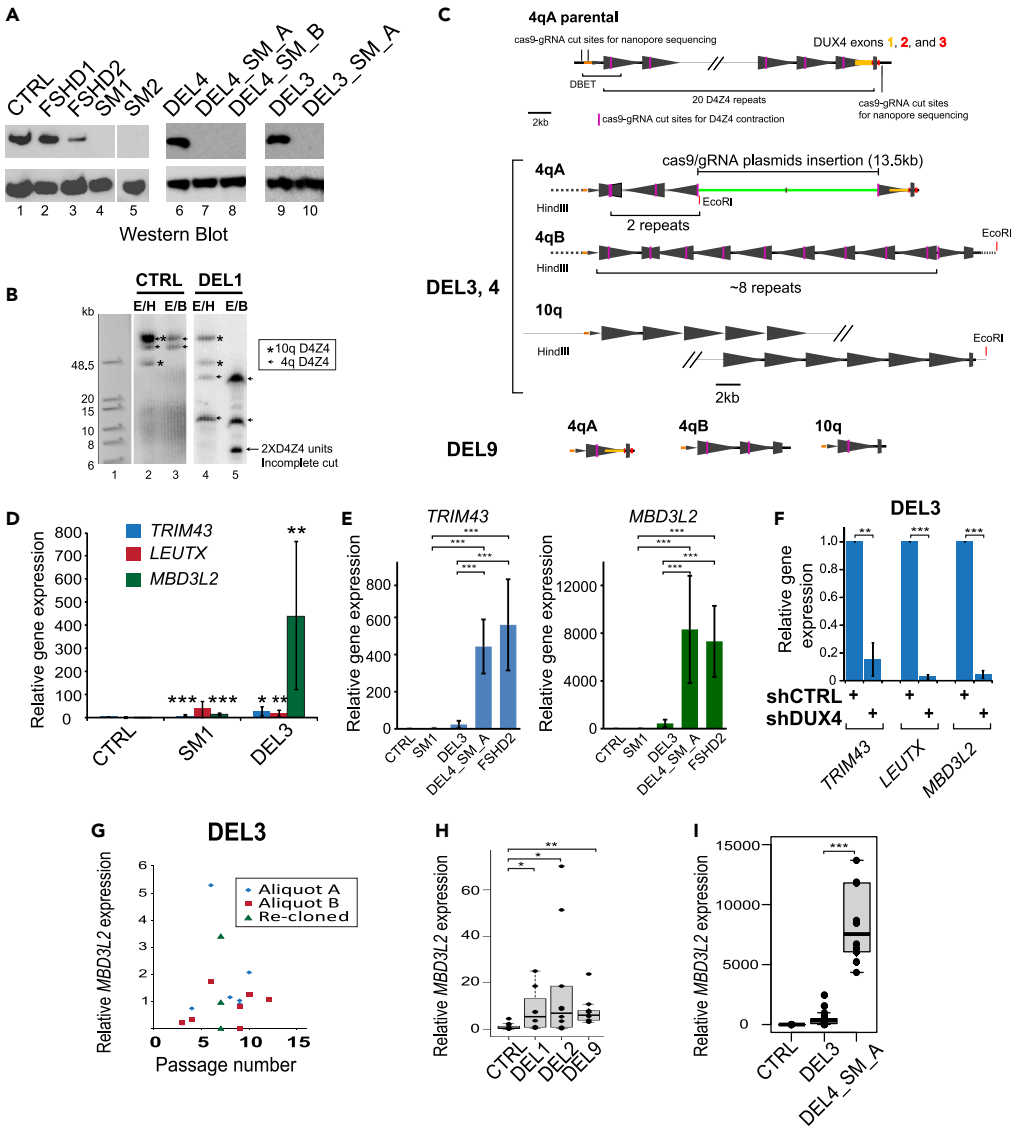


Figure 1. Generation of SMCHD1 and/or D4Z4 mutant cells from health permissive skeletal myoblast

(A) Western blot analysis the SMCHD1 protein expression in the cell lines used in the study. Lysates of immortalized control and FSHD1 and FSHD2 patient myoblasts, SMCHD1 mutants (SM), D4Z4 deletion mutants (DEL) and double mutants (DEL_SM) were subjected to western blot analysis using antibody specific for SMCHD1. β-Actin serves as a loading control.

(B) Determination of the 4q and 10q D4Z4 repeat number. Examples of control and DEL1 mutant clones are shown. Genomic DNA was digested with EcoRI/HindIII (E/H) or EcoRI/BlnI (E/B) and subjected to PFGE. They were then blot-hybridized with the 4q/10q specific "1-kb" D4Z4 probe. E/H digestion leaves intact two 4q and two 10q D4Z4 arrays, while BlnI in an E/B only cleaves 10qD4Z4 repeat units. Size markers (in kb) are shown on the left. Arrowheads and stars indicate 4q and 10q D4Z4, respectively. The arrow indicates a band around 6.6 kb, which should be 2 D4Z4 repeat units caused by incomplete digestion. The two 4q D4Z4 repeat arrays are contracted, while the 10q D4Z4 bands size show no change.

(C) D4Z4 gRNA targeting resulted in repeat contraction and recombination, leaving the last repeat with the *DUX4* gene intact at the 4qA allele in DEL mutant cells. Top: schematic diagram of D4Z4 array in 4qA allele of parental cell with gRNA target sites for D4Z4 deletion (purple bars) as well as crRNA target sites designed for nanopore sequencing were shown at the top panel. D4Z4 cluster in 4qA, 4qB and 10q alleles of DEL3 were shown below. 10q D4Z4 sequences were confirmed by SNP analysis. The large triangle represented a 3.3 kb D4Z4 unit and its orientation. The small and partial triangle represented partial D4Z4 units and their orientation. The endonucleases (EcoRI/HindIII) cut sites, which generated the fragments detected in PFGE, are indicated.

(D) Control, SMCHD1 mutant SM1 and D4Z4 contraction mutant DEL3 myoblast clones were differentiated and analyzed for DUX4 target (*TRIM43, LEUTX* and *MBD3L2*) RNA expression levels. Data are expressed as relative expression (mean ± standard deviation (SD)). Experiments are repeated (6 times for control, 5 times for SM1, and 8 times for DEL3). The gene expression over *GAPDH* was normalized to the corresponding gene expression value of the Control. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$ vs. control by unpaired Student's *t* test.

Figure 1. *Continued*

(E) Corresponding data from double mutant (DEL4_SM_A) and FSHD2 were added to (D) for comparison. Experiments were repeated 10 times for DEL4_SM_A and 3 times for FSHD2 cells. Error bars indicate the standard deviation from the mean values. ***p < 0.001. Multiple unpaired Student's t-tests were performed, with each test comparing two groups of data.

(F) DUX4 depletion by lentiviral shRNA abolished activation of DUX4 target genes (*TRIM43*, *LEUTX* and *MBD3L2*) compared to control shRNA (shCTRL). Cells were harvested at 4 days of myotube differentiation. DUX4 target expression levels were determined by RT-qPCR. Y axis is relative expression (mean ± SD) with the expression in shCTRL-transduced samples as one. **p < 0.01 and ***p < 0.001 by unpaired Student's *t* test, n = 3.

(G) Stochasticity of target gene expression in the DEL3 cell line. Two aliquots of DEL3 (aliquots A and B) at different passage numbers and 3 replicates of a re-cloned DEL3 (total 23 samples) were differentiated for 4 days and analyzed for *MBD3L2* expression by RT-qPCR.

(H) Similar variegation was also observed as in (G) in 3 other DEL mutant clones (DEL1 (n = 8), DEL2 (n = 10) and DEL9 (n = 11). Control (n = 11). *p < 0.05 and **p < 0.01 by unpaired Student's *t* test.

(I) Comparison of *MBD3L2* expression level in early myotube of Control (n = 11), single mutant DEL3 (n = 23), and double mutant DEL4_SM_A (n = 10). The dots on each boxplot represent the individual data, which was normalized to the mean value of Control, in each repeat. ***p < 0.001 by unpaired Student's *t* test.

transcripts was much lower than that in FSHD2 patient cells with a heterozygous *SMCHD1* mutation and robust DUX4 target gene expression (Figure 1E). Thus, neither partial shRNA depletion nor somatic KO mutation of *SMCHD1* in adult myocytes is sufficient to upregulate DUX4 target genes to the extent observed in FSHD2 patient cells used in our study.

D4Z4 contraction induces variegated DUX4 target gene expression, which is stabilized and enhanced by concurrent loss of SMCHD1

A set of gRNAs was designed at the single-nucleotide polymorphism (SNP) regions to enhance mutations of 4q over 10q D4Z4 repeats (DEL mutants) (Figures 1A–1C; Figure S2B). Parental control cells have 24 and 19 copies of D4Z4 on chromosome 4q and 23 and 14 copies on chromosome 10q based on PFGE, which was further confirmed by optical mapping and genomic nanopore sequencing (Figure 1B; Figure S3). We found that CRISPR-CAS9-mediated D4Z4 disruption was sufficient to upregulate DUX4 target genes (e.g., *TRIM43*, *LEUTX* and *MBD3L2*) in early myotubes, which was therefore used as a screening phenotype (Figure 1D; Figure S2B). We generated DEL mutant clones using DNA plasmid- and protein/RNA-based CRISPR-Cas9 systems by screening ~300 and 180 clones, respectively (see STAR methods and Figure S2B). We observed a tendency of inverse correlation between high DUX4 target gene expression and efficient myotube differentiation (Figure S2C). It should be noted that for our experiments, we selected clones with a relatively high DUX4 target gene expression as well as efficient proliferation and differentiation capabilities (Figure S2C). We found that D4Z4 mutant (DEL) clones showed higher DUX4 target gene expression than *SMCHD1* mutant (SM) clones, but still much lower than FSHD patient cells (DEL3 as an example in Figures 1D and 1E). Importantly, target gene upregulation is DUX4-dependent as DUX4 shRNA depletion abolished the expression of target genes (Figure 1F). It should be noted, however, that target gene expression is highly variable even in the same clone with different passage numbers and/or in different experiments (Figure 1G).

Although PFGE analyses indicated repeat contraction (Figure 1B), we analyzed the regions and resulting transcripts using genomic and RNA nanopore long-read sequencing, respectively. For DEL mutants generated by the plasmid-based CRISPR-CAS9 treatment, we found that gRNA-mediated cutting of D4Z4 repeats in the 4qA allele resulted not only in deletion of repeats upstream of the last copy, but also in insertion of inverted 2.5 copies of D4Z4 repeats as well as 2 copies of a gRNA plasmid sequence separated by a small fragment of a CAS9 plasmid sequence (Figure 1C; Figures S2D and S3B). These inverted repeats could give rise to the ~2 repeat signal in PFGE when in fact only the last copy is present downstream of this insertion, thus creating the 4qA allele with one D4Z4 repeat (Figure 1B). For 4qB, we found the repeat shortening and inversion, the length consistent with the PFGE band, which would not yield any significant DUX4 gene expression due to the absence of canonical poly(A) signal sequence[17] (Figure 1C). No rearrangements were detected in 10q D4Z4 alleles, consistent with the expectation with 4q-tailored gRNAs (Figure 1C). The insertion of plasmid sequences resulted in expression of EGFP from the PKG promoter as well as small spliced RNA fragments containing *DUX4* exons 2 and 3 (corresponding to the 3′-UTR of *DUX4* mRNA) (termed a "chimeric DUX4 3′-UTR") (Figure S2D). We confirmed that overexpression of the chimeric DUX4 3′-UTR fragment has no effect on DUX4 target gene expression (Figure S2E) or other genes which were found to be altered in mutant cells (Figure S2F). As a comparison, we generated a second set of mutants using protein/RNA-based CRISPR-Cas9 mutagenesis (to minimize the insertion problem). We screened for 120 clones and picked 7 clones which showed SMCHD1 KO. We validated 2 clones of the double mutant DEL9-SM for further experiments. The other 5 clones didn't differentiate well and therefore were not chosen for the experiments. This approach resulted in repeat contraction down to one copy of D4Z4 at 4qA and 2 copies at 4qB, and also reduced the copy number of 10qD4Z4 alleles to one copy as confirmed by nanopore sequencing (Figure 1C, DEL9). DEL9 cells exhibit comparable upregulation of DUX4 target genes as the plasmid-based DEL mutants (Figure 1H, *MBD3L2* as an example). Consistent with the one intact *DUX4* gene in the 4qA allele, RNA nanopore sequencing of both types of DEL mutant clones confirmed an intact wild type *DUX4fl* transcript, indicating that the last D4Z4 copy at 4qA retained the ability to express *DUX4fl* (Figure S2D). Thus, we used both types of DEL mutant clones for further analyses.

Although DEL mutants exhibit upregulation of *DUX4* and target genes, their expression is highly variegated and tends to be much lower than in FSHD patient cells (Figure 1E). In FSHD1 patients, however, the lower D4Z4 repeat numbers associate with more severe clinical phenotypes.[2,48] Our results suggest, therefore, that D4Z4 contraction (down to even one repeat copy) is not sufficient, and additional mechanism(s) are required for efficient *DUX4* and target gene activation. As *SMCHD1* mutations are associated with severe cases of FSHD1

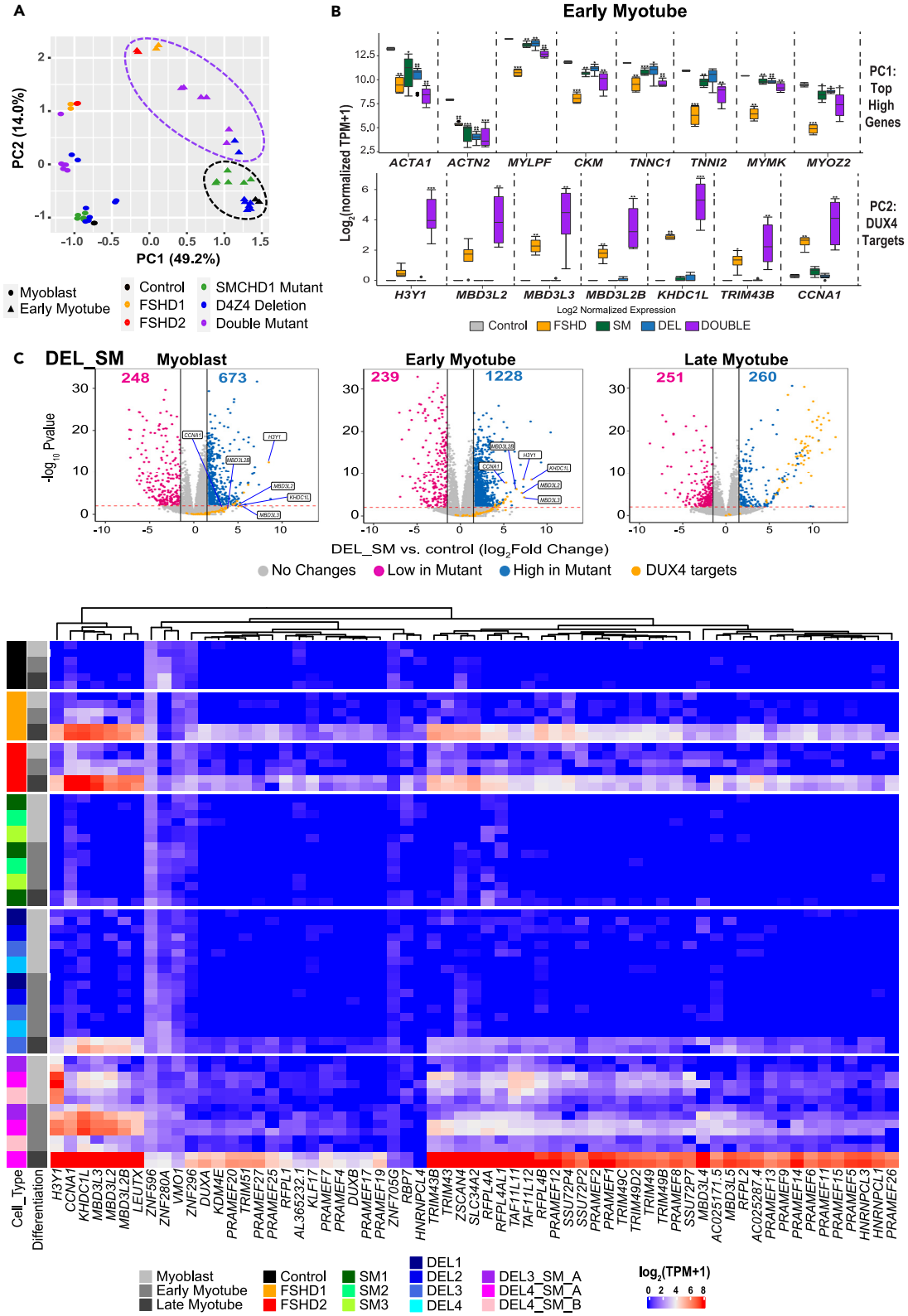

A
PC2 (14.0%)
PC1 (49.2%)
Myoblast
Early Myotube
Control
FSHD1
FSHD2
SMCHD1 Mutant
D4Z4 Deletion
Double Mutant
B
Early Myotube
Log2(normalized TPM+1)
ACTA1
ACTN2
MYLPF
CKM
TNNC1
TNNI2
MYMK
MYOZ2
PC1: Top High Genes
H3Y1
MBD3L2
MBD3L3
MBD3L2B
KHDC1L
TRIM43B
CCNA1
PC2: DUX4 Targets
Log2 Normalized Expression
Control
FSHD
SM
DEL
DOUBLE
C
DEL_SM
Myoblast
Early Myotube
Late Myotube
248
673
239
1228
251
260
-log10 Pvalue
DEL_SM vs. control (log2Fold Change)
No Changes
Low in Mutant
High in Mutant
DUX4 targets
D
Control
FSHD1
FSHD2
SM
DEL
DEL_SM
Cell_Type
Differentiation
H3Y1
CCNA1
KHDC1L
MBD3L3
MBD3L2
MBD3L2B
LEUTX
ZNF596
ZNF280A
VMO1
ZNF296
DUXA
KDM4E
PRAMEF20
TRIM51
PRAMEF27
PRAMEF25
RFPL1
AL365232.1
KLF17
PRAMEF7
PRAMEF4
DUXB
PRAMEF17
PRAMEF19
ZNF705G
RBP7
HNRNPCL4
TRIM43B
TRIM43
ZSCAN4
SLC34A2
RFPL4A
RFPL4AL1
TAF11L11
TAF11L12
RFPL4B
PRAMEF12
SSUT2P4
SSUT2P2
PRAMEF2
PRAMEF1
TRIM49C
TRIM49D2
TRIM49
TRIM49B
PRAMEF8
SSUT2P7
MBD3L4
AC025171.5
MBD3L5
RFPL2
AC025287.4
PRAMEF13
PRAMEF9
PRAMEF14
PRAMEF6
PRAMEF11
PRAMEF15
PRAMEF5
HNRNPCL3
HNRNPCL1
PRAMEF26
Myoblast
Early Myotube
Late Myotube
Control
FSHD1
FSHD2
SM1
SM2
SM3
DEL1
DEL2
DEL3
DEL4
DEL3_SM_A
DEL4_SM_A
DEL4_SM_B
log2(TPM+1)

Figure 2. Double mutants closely recapitulate patient cells

(A) PCA analysis of myoblasts and early myotubes across all the cell types and clones. Top genes for each component are included in the Table S2. Differentiation days are indicated by shapes and cell types are indicated by colors according to the label legend.

(B) Expression comparison of selected genes from PC1 and PC2 from (A). Top: boxplots of the selected top high genes expression of PC1 in control, FSHD patients and 3 types of mutant early myotubes. Bottom: boxplots of 7 DUX4 targets expression from the top 500 high genes of PC2. Expression values are in $\log_2$ (normalized TPM +1). Significant values were calculated by Wilcoxon t-test (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05).

(C) Volcano plots of significance ($\log_{10}$ p-value) and $\log_2$ fold change of double mutants (DEL_SM) compared to control at myoblast, early and late myotube stages. Significantly upregulated (blue) and downregulated (pink) genes and DUX4 targets (orange) are shown. DUX4 target genes in PC2 (B) are indicated in myoblasts and early myotubes.

(D) Hierarchical heatmap of DUX4 target gene expression. A total of 63 target genes were selected based on previous studies.[41,45] Expression values are in normalized TPM and log transformed. Gray shades indicate differentiation and colors indicate cell types.

(combined with D4Z4 contraction),[7,8] we generated double mutant cells (DEL_SM), by introducing *SMCHD1* mutation in DEL clones (Figure S2B). In comparison to single DEL mutants, DEL_SM mutants consistently upregulated DUX4 target genes at a higher level (Figures 1E and 1I). These results indicate that the loss of SMCHD1 acts synergistically with D4Z4 contraction to enhance and stabilize the DUX4 target gene expression in our mutant cells. The results were somewhat unexpected as somatic *SMCHD1* mutation (even the complete KO) alone only exhibited very weak phenotype (see above, Figure 1E).

Synergistic effect of double mutations recapitulates patient cell phenotype

To determine the genome-wide effects of these engineered mutations, we performed RNA-seq of 3–5 clones of each mutant type (SM, DEL, and DEL_SM) in comparison to the isogenic control in myoblasts and two time points during myotube differentiation. We also included one representative each of FSHD1 and FSHD2 patient cells as protypes with clear DUX4 target gene induction and efficient proliferation and differentiation for comparison. Although we chose clones with relatively efficient differentiation, we consistently observed prominent delays in differentiation of DEL and DEL_SM clones even in those with comparable doubling time (Figure S4A). A mild delay was also observed in FSHD patient cells whereas the delay was minimal in *SMCHD1* mutant (SM) cells compared to the control cells. Time course principal component analyses (PCA) of gene expression during differentiation also confirmed delays in DEL and DEL_SM mutant cells compared to the isogenic control and SM cells (Figure S4B). To minimize the data skewing due to differentiation delay, we compared DEL and DEL_SM samples at days 4∼5 and 13–14 of with control and SM samples on days 3 and 12 of differentiation as "early" and "late" myotubes, respectively, based on our previous time course analyses of DUX4 target gene expression.[40] Muscle sarcomere protein genes, MYL3 and MYH1, are expressed significantly higher in late myotubes (Figure S4C), consistent with more mature morphology with increased striation.[40,49]

Using edgeR, we identified up- and downregulated differential expressed genes (DEGs) (p-values < 0.01 and Log_2 fold change >1.5) in each mutant type (Table S1). At both myoblast and early myotube stages, the number of genes that are upregulated in the DEL_SM mutant cells is significantly more (673 and 1228, respectively) than those in the DEL- or SM-only mutants combined (203 and 228, respectively) (Figure 2C; Figure S5A). PCA of myoblast and early myotube stages revealed that DEL_SM clones and FSHD patient myocytes cluster together (Figure 2A). While PC1 (49.2% variance) largely separates myoblast and differentiated myotube stages, there is a significant clustering of patient and DEL_SM myotubes comparing to the control and single mutants, and two clusters are further separated in PC2 (14.0% variance) (Figure 2A). The top DEGs in PC1 include myogenesis genes that are downregulated (Table S2). Boxplots show that albeit at varying degrees, these genes are commonly downregulated in all mutants at early myotube stage (Figure 2B, top; Table S1). Thus, all three types of mutations (DEL, SM and DEL_SM) impede expression of muscle genes, which may explain the delayed myotube differentiation. The top misexpressed genes in PC2 include a subset of DUX4 target genes that are significantly upregulated in patient and DEL_SM, but not in single mutant, early myotubes (Figure 2B, bottom; Figure S5B).

Expression of 63 representative DUX4 target genes[41,45] increased in DEL_SM during differentiation and became predominant in late myotubes (Figure 2C, orange). These genes include 54 genes that we previous defined as "FSHD-induced genes" differentially upregulated in our 6-day time course RNA-seq analyses of primary patient, but not control, myotube differentiation.[41] All of them except 1 were previously identified as possible DUX4 targets from myoblasts with inducible DUX4,[50] endogenous DUX4[23] and FSHD biopsies.[45] We also added additional genes from the list of robust DUX4 target genes in the study by Yao et al.[45] for our analyses. Hierarchically clustered heatmaps in three types of mutant clones compared to isogenic control and FSHD patient myocytes indicated that DUX4 target gene expression is strongest at late myotube stage in DEL, DEL_SM and patient cells (Figure 2D). In SM mutants, however, DUX4 target expression was over all very low throughout differentiation (Figure 2D). Consistent with the RT-qPCR analyses (Figures 1E and 1I), target gene expression is highly prominent in DEL_SM mutants, which is clear even in myoblast and early myotube stages (Figure 2D; Figures S5B and S6A). This is in a stark contrast to DEL mutants, in which target gene expression is relatively weak in early stages (Figure 2D; Figures S5B and S6A). These results further highlight the synergy between D4Z4 contraction and *SMCHD1* KO.

Characteristics of non-DUX4 target DEGs are recapitulated in mutant cells

We performed Gene Ontology (GO) enrichment analysis on the DEGs from DEL_SM for three stages of differentiation (myoblasts, early and late myotubes) (Figures 3A and 3B; Table S3). While similar GO terms in muscle contraction and development were enriched for downregulated genes at all three stages, we found largely distinct GO term enrichment for upregulated genes at each stage, suggesting that upregulation of DEGs are closely linked to differentiation stages of myocytes (Figures 3A and 3B). Namely, we observed prominent upregulation of

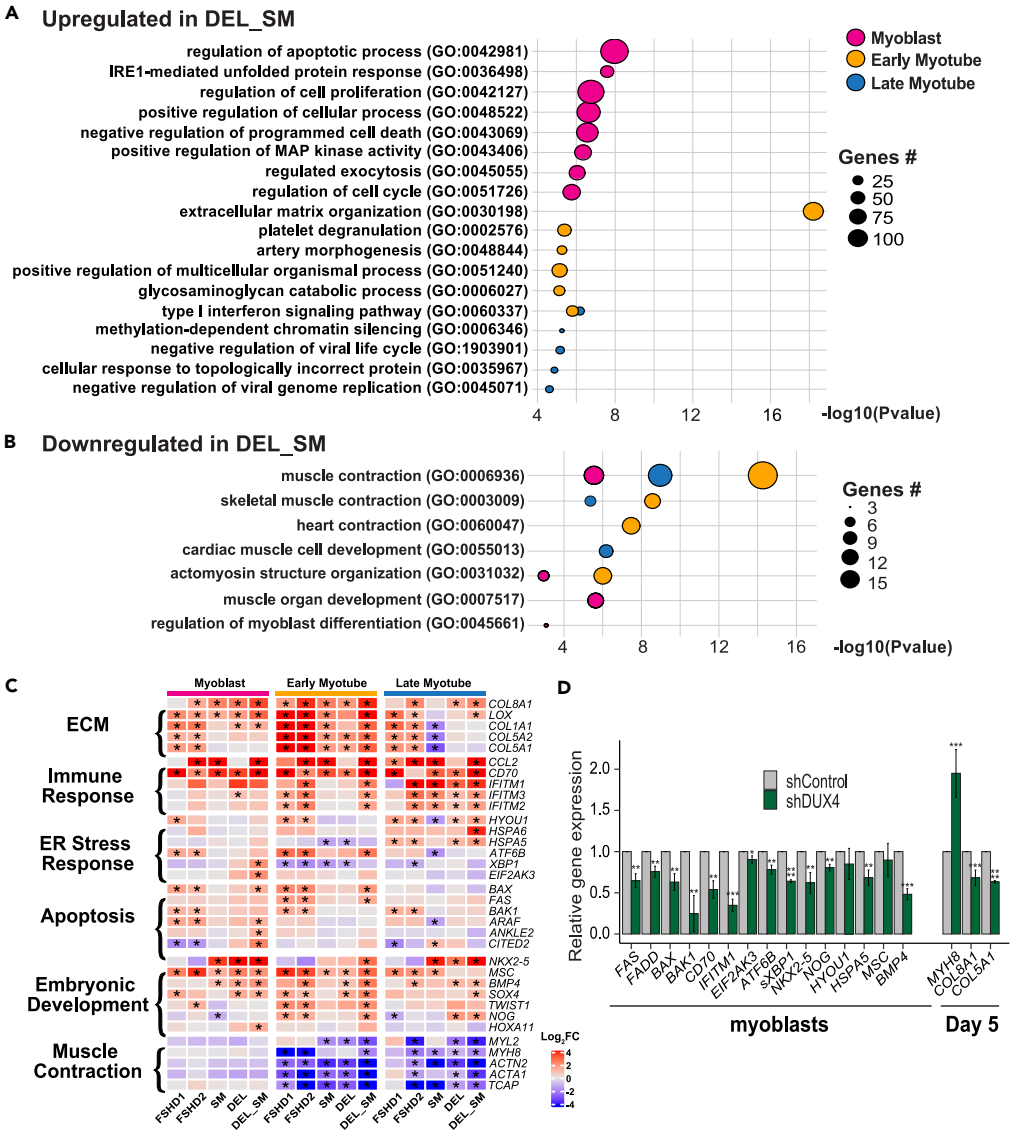


Figure 3. Ontology analyses of common and distinct gene expression in patient and double mutant cells

(A) The bubble plot shows gene ontology enrichment analysis of upregulated genes in the double mutants in myoblasts (pink), early (yellow) and late (blue) myotubes. The plot shows the selected top terms for each differentiation stage. X axis displays $\log_{10}$ p-value and bubble size indicates number of genes in each term as indicated.

(B) As in (A), bubble plot for downregulated genes.

(C) Heatmap of $\log_2$ fold change expression for the selective genes in FSHD patients and mutants and their related pathway. Asterisks indicate p value <0.01 and $\log_2$FC > +1.5 or < −1.5.

(D) The expression level of selected genes from (C) in double mutant myoblast (DEL4_SM_A) transduced with lentivirus carrying shControl or shDUX4. Real-time RT-qPCRs were performed for three biological replicates for each sample. Data are presented as mean ± SD; **p < 0.01, ***p < 0.001, by one-tailed Student's *t* test. Results presented as fold difference compared to shControl sample.

genes related to extracellular matrix (ECM), immune response, ER stress, apoptosis, and embryonic genes as well as down regulation of muscle-related genes across patient and mutant cells (Figure 3C). Similar GO terms were previously reported in patient myocytes and recombinant DUX4 overexpression studies.[13,19,38,50–53] DUX4 depletion reversed these changes of representative genes, suggesting that most of these changes are triggered by the mutation-induced DUX4 expression (Figure 3D). However, this may be indirect as the DUX4 depletion effect was not as drastic as that on DUX4 target genes (Figure 1F).

We observed that both positive and negative regulators of apoptosis are upregulated especially in patient, DEL, and DEL_SM myoblasts and early myotubes (e.g., *BAX, BAK1, FAS* and *FADD* for pro-apoptosis, and *CITED2, ANKLE2,* and *ARAF* for cell survival) (Table S3).

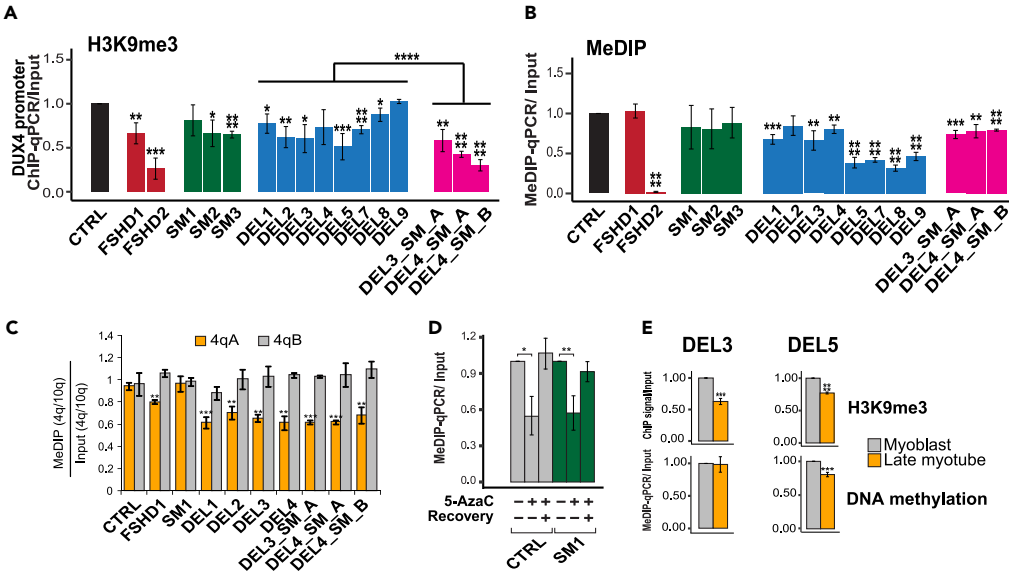


Figure 4. Heterochromatin changes in mutant cells

(A) H3K9me3 ChIP-qPCR analysis of the *DUX4* promoter region in FSHD1, FSHD2 and mutant myoblasts. Reduction of H3K9me3 is enhanced in double mutant cells. For both (A) and (B), signals were normalized to input of the corresponding samples. Error bars indicate the standard deviation from the mean values. Significant comparisons to the control are shown with the asterisks calculated by Student's *t* test. The mean value of DEL mutant group is also compared to that of DEL_SM group using Student's *t* test. (****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05).

(B) DNA methylation from MeDIP analysis. No reduction of DNA methylation was observed in SMCHD1 only and no additional effect in double mutant myoblasts. Error bars indicate the standard deviation from the mean values. Significant values were calculated by Student's *t*-test (****p < 0.0001, ***p < 0.001, **p < 0.01).

(C) Comparison of DNA methylation levels at 4qA, 4qB, and 10q D4Z4 regions among the control, FSHD, and mutant cells. The MeDIP and input samples from (B) were amplified by using 4q/10q-D4Z4 specific PCR primers. The PCR products were sequenced and the 4qA, 4qB, and 10q D4Z4 specific sequence reads were analyzed. The 4qA(orange)/10q and 4qB(gray)/10q ratios of MeDIP were normalized with that of input. The data indicated relatively lower methylation at 4qA (but not 4qB) D4Z4 regions of FSHD1, D4Z4 deletion mutants and double mutants. p-values for significant differences versus the control sample are shown. Significant values were calculated by Student's *t*-test (***p < 0.001, **p < 0.01). All values are mean ± SD.

(D) The effect of 5AzaC treatment on SMCHD1 mutants on DNA methylation. The control or SM1 cells were treated with 5AzaC for 24 h and allowed for 48 h of recovery before harvesting for MeDIP-qPCR. Error bars indicate the standard deviation from the mean values. Significant values were calculated by Student's *t*-test (**p < 0.01, *p < 0.05).

(E) H3K9me3 ChIP-qPCR and MeDIP analysis at the DUX4 promoter region were performed to compare between Day 0 and Day 14 of DEL3 or Day 12 of DEL5 mutants. Error bars indicate the standard deviation from the mean values. Significant values were calculated by Student's *t*-test (****p < 0.0001, ***p < 0.001).

Unexpectedly, these gene expression changes appear to taper off in late myotubes despite the higher expression of DUX4 target genes (Figures 2C and 2D, 3C). Although DUX4 expression was linked to immediate cell death,[23,50] our results suggest more complex fate of DUX4-activated myotubes.

DEL mutations diminish D4Z4 heterochromatin, which is exacerbated by H3K9me3 reduction induced by SMCHD1 KO

It is well known that disruption of H3K9me3 and DNA methylation is linked to DUX4 de-repression in both FSHD1 and FSHD2 myoblasts.[18,27,28,54] Importantly, we showed that the loss of H3K9me3 induces DUX4 expression.[28] Thus, we examined the two heterochromatin marks at the *DUX4* promoter using ChIP-qPCR and MeDIP, respectively (Figures 4A and 4B). Unexpectedly, the reduction in H3K9me3 was observed in both DEL and SM, and was more substantial and consistent in DEL_SM (Figure 4A). Thus, the loss of *SMCHD1* alone affects H3K9me3, and further enhances H3K9me3 reduction in D4Z4 mutants, correlating with their synergy on DUX4 target gene expression (Figures 2 and 4A). We previously showed that SMCHD1 binding to D4Z4 is H3K9me3-dependent.[28] Taken together, our results uncover a positive feedback relationship between SMCHD1 and H3K9me3 at D4Z4.

MeDIP qPCR results indicate that DEL mutants exhibit reduction of DNA methylation at the DUX4 promoter region (Figure 4B). DNA methylation analyses of nanopore sequencing data of the D4Z4 genomic region confirmed that the proximal DBET region is hypomethylated whereas the last D4Z4 repeat is hypermethylated in the parental control line (Figure S3B). Both types of DEL mutants (i.e., DEL3/DEL4 with the exogenous vector insertion in addition to D4Z4 deletion, and DEL5/DEL9 with D4Z4 deletion) showed that the DUX4 promoter region in the last copy of D4Z4 is less methylated compared to the corresponding region of the parental control (Figure S3B). Interestingly, a significant loss of DNA methylation was observed both at the DUX4 promoter and gene body in DEL5, which expresses higher DUX4 target genes than DEL9 (Figures S3B and S3C). In contrast to FSHD2 patient cells carrying *SMCHD1* mutations, loss of *SMCHD1* in our cell lines (SM mutants) did not affect DNA methylation (Figure 4B). Thus, a somatic mutation of *SMCHD1* does not recapitulate the DNA methylation changes observed in

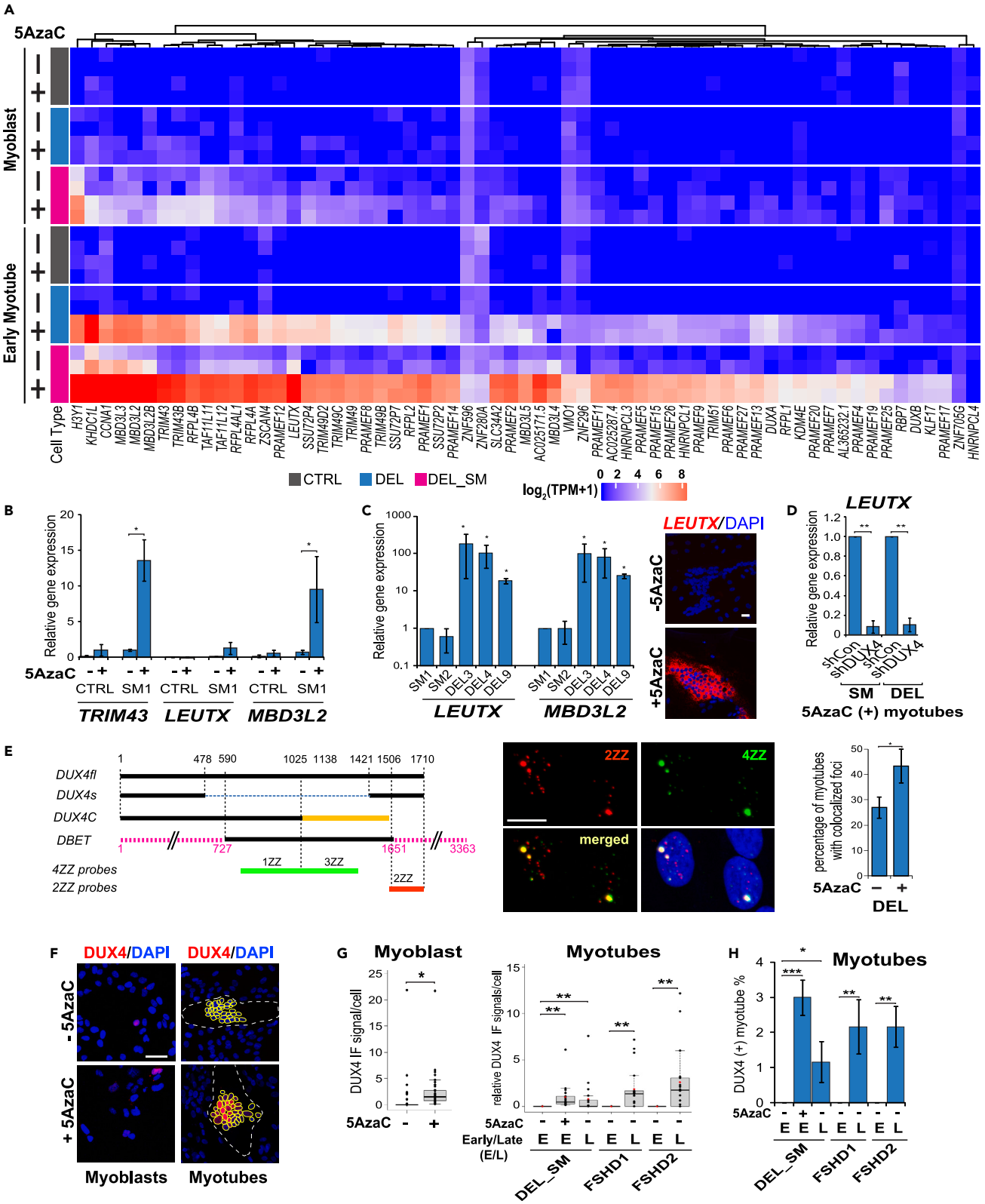

A
5AzaC
Myoblast
Early Myotube
Cell Type
CTRL
DEL
DEL_SM
log2(TPM+1)
B
Relative gene expression
5AzaC
CTRL SM1 CTRL SM1 CTRL SM1
TRIM43 LEUTX MBD3L2
C
Relative gene expression
SM1 SM2 DEL3 DEL4 DEL9
LEUTX MBD3L2
LEUTX/DAPI
-5AzaC
+5AzaC
D
LEUTX
Relative gene expression
shCon shDUX4 shCon shDUX4
SM DEL
5AzaC (+) myotubes
E
DUX4fl
DUX4s
DUX4C
DBET
4ZZ probes
2ZZ probes
2ZZ
4ZZ
merged
percentage of myotubes with colocalized foci
5AzaC
DEL
F
DUX4/DAPI
DUX4/DAPI
- 5AzaC
+ 5AzaC
Myoblasts
Myotubes
G
Myoblast
DUX4 IF signal/cell
5AzaC
Myotubes
relative DUX4 IF signals/cell
5AzaC
Early/Late (E/L)
DEL_SM
FSHD1
FSHD2
H
Myotubes
DUX4 (+) myotube %
5AzaC
DEL_SM
FSHD1
FSHD2

Figure 5. Inhibition of DNA methylation increases DUX4fl RNA and protein and robustly upregulates target genes in mutant cells

(A) Hierarchical heatmap of DUX4 target gene expression for control, DEL3, and DEL3_SM_A mutants with or without 5AzaC at myoblast and early myotube. A total of 63 target genes were selected based on previous studies.[41,45] Expression values are in normalized TPM and log transformed.

(B) DUX4 target genes were greatly affected by 5AzaC treatment in mutant cells. Control and SM1 myoblasts were treated with or without 5AzaC for 48 h. Then 5AzaC was removed from the media, and differentiation was induced. 4 days later, RT-qPCR of DUX4 target genes were performed. The gene expression data were normalized to *GAPDH* level in each sample, and then normalized to the *LEUTX* value of 5AzaC treated SM1. Data are presented as mean ± SD; *p < 0.05, by one-tailed Student's t-test.

(C) Comparison of DUX4 target genes level between early myotubes of 5AzaC treated SM and DEL mutants. SM and DEL myoblasts were treated with 5AzaC for 48 h right before differentiation. At day 5 of differentiation, the mRNA expression level of DUX4 target genes was assessed by real-time RT-PCR, relative to SM1. Data are presented as mean ± SD; *p < 0.05, by one-tailed Student's t-test. Representative images of *in situ* detection of *LEUTX* RNA (red) with or without 5AzaC treatment are shown on the right (blue: DAPI). Scale bar 10 μm.

(D) DUX4 depletion inhibited DUX4 target gene upregulation induced by 5AzaC treatment in mutant cells. SM1 and DEL3 cells were treated with 5AzaC and induced differentiation same as (B). During 5AzaC treatment, cells were infected with lentivirus containing shCTRL or shDUX4. For each cell line, LEUTX expression level after DUX4 depletion was shown as fold difference compared to the control. Data are presented as mean ± SD; **p < 0.01, by one-tailed Student's t-test.

(E) 5AzaC facilitated DUX4fl expression in DEL3 early myotubes. Left: the schematic diagrams of mRNA transcripts for *DUX4fl*, the *DUX4s* isoform and *DUX4* homologs (*DUX4c* and *DBET*), the black regions, which represent >99% homology to *DUX4fl*, could be detected by corresponding *DUX4* 4ZZ probes or 2ZZ probes, but not by both. Therefore, the overlapping signals from 4ZZ and 2ZZ probes represent the *DUX4fl* transcripts. Middle panel, example images of the RNAScope results of the 4ZZ probes (green), 2ZZ probes (red) and the overlapping foci (yellow). DAPI is in blue. Scale bar = 10 μm. Right, 5AzaC treatment increased the percentage of myotubes with overlapping foci of 4ZZ and 2ZZ probes. Data from 3 independent experiments are presented as mean ± SD. *p < 0.05, by one-tailed Student's t-test.

(F) Examples of DUX4 protein expression in double mutant myoblasts and myotubes. Immunofluorescence for DUX4 on DEL4_SM_A cells after 5 days of differentiation. Nuclei were counterstained with DAPI (blue). Scale bar 50 μm.

(G) Quantification of DUX4 protein expression with and without 5AzaC in double mutant myoblasts (left) and early myotubes (right). Mutant late myotubes and FSHD1 and FSHD2 patient early and late myotubes are shown for comparison. The DUX4 integrated density values in myoblasts/myotubes were measured using ImageJ software.[40] Top 3% values in each group were used for graph and data analysis. All the data were normalized to the corresponding mean value of the 5AzaC-treated DEL_SM samples. *p < 0.05, **p < 0.01, ***p < 0.001, by unpaired Student's *t* test. Totally 600 myotubes or 1200 myoblasts were observed in each group.

(H) Replotting the data in (G, right panel) for the frequency of DUX4 IF staining positive myotubes. All values are mean ± SD. *p < 0.05, **p < 0.01, ***p < 0.001, by unpaired Student's *t* test.

FSHD2 myoblasts. We also separated reads specific to 4qA, 4qB and 10q D4Z4 based on their SNPs[27] and analyzed the read count ratio of 4q/10q in MeDIP compared to input genomic DNA at the *DUX4* promoter region. The results further confirm a significant decrease of DNA methylation at 4qA D4Z4 over 10q in FSHD1, DEL and DEL_SM mutants (Figure 4C). Importantly, we observed no significant difference between control and SM1 mutant cells as well as between DEL and DEL_SM cells (Figure 4C). These results suggest a negligible contribution of *SMCHD1* KO to the maintenance of D4Z4 DNA methylation in adult myocytes either with or without D4Z4 contraction.

An earlier study suggested that SMCHD1 is important for *de novo* methylation of D4Z4 repeats.[47] Therefore, we compared recovery of DNA methylation in control- or SM1 mutant myoblasts after treatment with the DNA methylation inhibitor, 5AzaC. Following treatment with 5AzaC for 48 h, cells were allowed to recover for 48 h. MeDIP revealed no differences in recovery between the control and *SMCHD1* mutant cells (Figure 4D). Taken together, our results indicate that SMCHD1 is not required for either the maintenance or the re-establishment of DNA methylation at the DUX4 promoter in adult myoblasts. Importantly, however, this somatic SMCHD1 mutation effectively stimulates DUX4 target gene activation in our double mutants (DEL_SM) (Figure 2). Thus, the results strongly suggest that the observed stimulatory effect of SMCHD1 mutation in double mutant cells on DUX4 target gene induction is not through change in D4Z4 DNA methylation. Taken together with the effect of SMCHD1 mutation on H3K9me3 reduction described above, it is possible that the effect may be at least in part through reduction of H3K9me3.

At late myotube stages in DEL mutants, we observed a further reduction in DNA methylation and/or H3K9me3 at the *DUX4* promoter, accompanied by more efficient expression of *DUX4* and target genes (Figures 2D and 4E). Our results highlight the intimate relationship between reduction of D4Z4 heterochromatin marks and DUX4 gene network activation.

Inhibition of DNA methylation boosts expression of the DUX4fl network in mutant cells

One major difference between our somatic mutant cells and FSHD2 cells that were tested was the level of DNA methylation (Figure 4B). To examine the role of DNA methylation, we treated control and mutant cells with 5AzaC. We observed robust induction of DUX4 target genes in DEL and DEL_SM mutant, but not in control cells (Figure 5A; Figure S6B). Importantly, this treatment has no effect in terms of DUX4 target gene activation in the parental control cells (Figure 5A). Upregulation is more prominent in early myotubes than at myoblast stages (Figure 5A). Although DUX4 target genes were significantly upregulated in SM myotubes by 5AzaC, they were induced over 100-fold higher in DEL myotubes (Figures 5B and 5C). Importantly, in both cases, upregulation was DUX4-dependent (Figure 5D; Figure S7A). Because 5AzaC also inhibits RNA methylation, we tested 5AzadC, which only gets incorporated into DNA, and obtained the same results (Figure S7B). Moreover, MeDIP indicates that both 5AzaC and 5AzadC caused a comparable reduction in DNA methylation at D4Z4 (Figure S7C). We conclude that inhibition of DNA methylation leads to an induction of the DUX4 gene network in DEL and DEL_SM mutant cells.

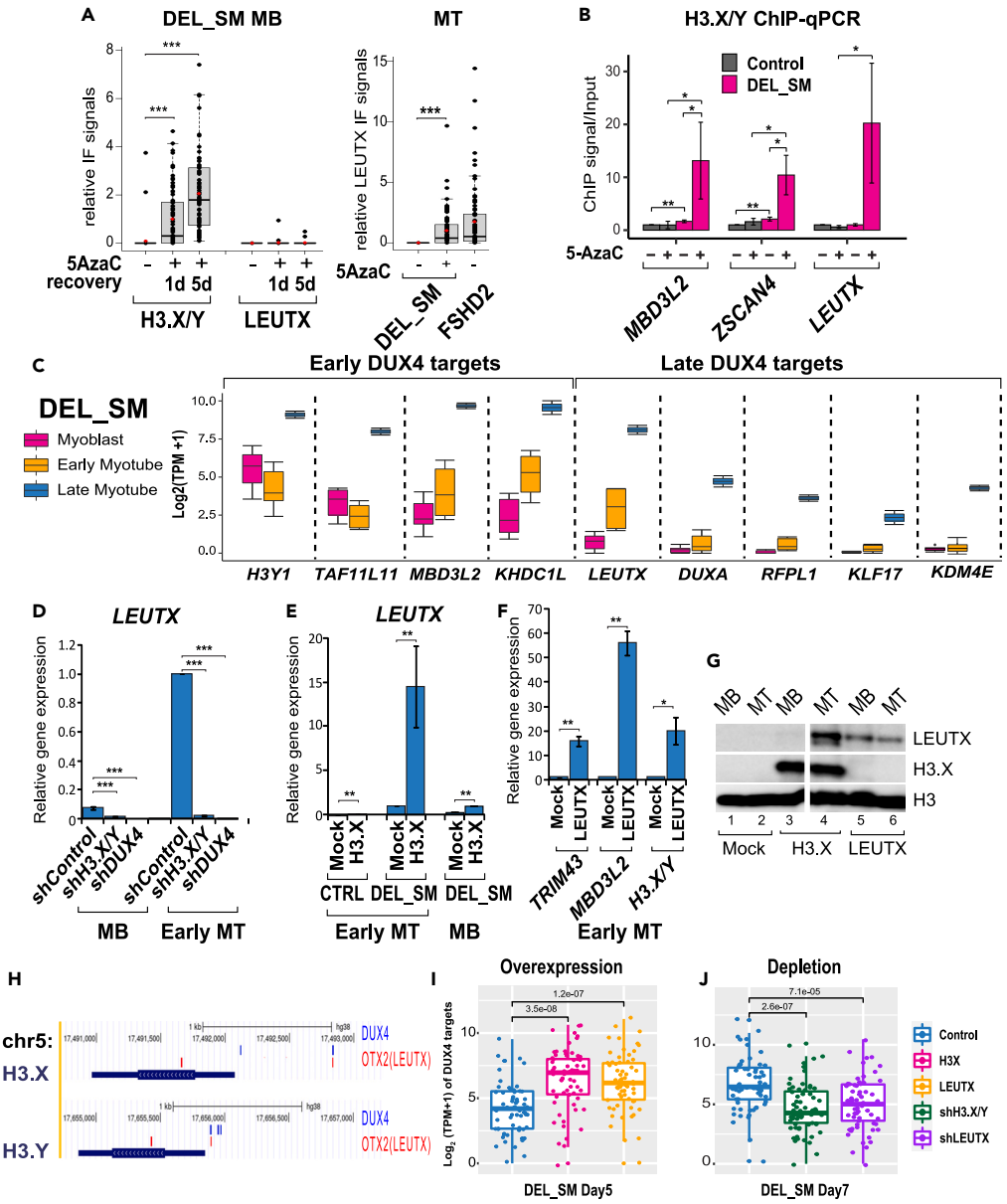


Figure 6. Identification of early and late DUX4 target genes that are differentially sensitive to myotube differentiation

(A) Double mutant cells were treated with or without 5AzaC as indicated. IF signals of H3.X/Y and LEUTX in myoblasts (left panel) and LEUTX in early myotubes (right panel) were quantified as integrated intensity in each myoblast/myotube using ImageJ software. FSHD2 patient early myotubes are shown for comparison. The DUX4 integrated density values in each myoblast/myotube were measured using ImageJ software. Based on the highest positive myoblasts/myotubes number of all, same number of values in each group were used for graph and data analysis. All the data were normalized to the corresponding mean value of the 5AzaC treated samples (release day1 for the myoblasts). Red dots represent mean values. (n = 300 myotubes or 1000 for myoblasts). ***$p < 0.001$, by one-tailed Student's *t*-test.

(B) Incorporation of H3.X/Y into DUX4 targets in control and double mutant at Day 4 with or without 5AzaC. Cells were treated with 5AzaC for 48 h before differentiation. Significant incorporation of H3.X/Y is shown by the asterisks with the indicated comparisons. Significant values were calculated by Student's *t*-test (**$p < 0.01$, *$p < 0.05$). All values are mean ± SD.

(C) Identification of early and late DUX4 target genes that are insensitive and sensitive to myotube differentiation, respectively. Boxplots of representative early and late DUX4 target gene expression in double mutant cells were shown to compare myoblast, early and late myotube stages as indicated. Expression values are in $\log_2$ normalized TPM.

(D) The effects of H3.X/Y or DUX4 shRNA depletion on LEUTX expression. Double mutant DEL4_SM_A myoblasts were infected with lentivirus expressing control, H3.X/Y or DUX4 shRNA and differentiated into myotubes. RNA was harvested at early myotubes stage (differentiation days 3–5) as described in the STAR methods section. Three biological replicates for each sample were performed. Data are presented as mean ± SD; **$p < 0.01$, ***$p < 0.001$, by one-tailed Student's *t* test. Results presented as fold difference compared to shControl differentiated sample.

Figure 6. *Continued*

(E) Control and DEL4_SM_A myoblasts were transduced with a lentiviral empty vector or a lentiviral vector expressing H3. X. Differentiation was induced at 48 h after transduction. For myoblasts or early myotubes as indicated, the mRNA expression level of the downstream target genes was assessed by real-time RT-qPCR. Data are presented as mean ± SD; *p < 0.05, **p < 0.01, ***p < 0.001, by one-tailed Student's *t* test. Results presented as fold difference compared to empty vector infected double mutant cells.

(F) Similar experiments as in (E), but DEL4_SM_A myoblasts were transduced with a lentiviral vector expressing LEUTX. Data are presented as mean ± SD; *p < 0.05, **p < 0.01, by one-tailed Student's *t* test. Results presented as fold difference compared to empty vector infected double mutant cells.

(G) Overexpression of H3.X and LEUTX in MB or MT was assessed by western blot. Pan histone H3 antibody was used as control as indicated. Lanes 1 and 2: mock transfection. Lanes 3 and 4: H3.X OE. Lanes 5 and 6: LEUTX OE. The endogenous LEUTX is upregulated in H3.X OE myotubes (lane 4).

(H) TF binding motifs at the promoter of H3.X/Y. Binding motifs for DUX4 and the putative LEUTX motif (OTX2) within 1 kb upstream and 0.5 kb downstream of the transcription start site for H3.X/Y were identified by using the MoLoTool provided in HOCOMOCO v11, with p-values less than or equal to 0.001. Visualization was done on the UCSC genome browser using GENCODE v36 for the H3.X/Y genes model.

(I) The effects of control, H3.X/Y or LEUTX overexpression (as in E–G) on 63 DUX4 target genes in DEL_SM myotubes Day 5 are assessed by RNA-seq and displayed in boxplots. p-values are calculated using Wilcoxon t-test indicated at the top.

(J) Similar analysis was performed with H3.X/Y or LEUTX shRNA depletion compared to the same control as in (I) on DUX4 target gene expression in DEL_SM myotubes Day 7.

To specifically detect the *DUX4fl* transcript that can be translated into the protein, we split our RNAScope probes to 4ZZ specific to the middle region and 2ZZ specific to the 3′end of the DUX4fl transcript (Figure 5E). Therefore, colocalization of the two probes should reflect the presence of DUX4fl. Using this strategy, we were able to confirm that the prominent nuclear foci of *DUX4* RNA previously detected[40] represents *DUX4fl* transcripts (Figure 5E). 5AzaC treatment indeed increased the colocalized signals of *DUX4fl* in DEL mutant myotubes (Figure 5E, right). Consistently, DUX4 protein expression at the early myotube stage is significantly induced by 5AzaC treatment to a level that is comparable to those at late myotube stages of DEL_SM and patient cells (Figures 5F–5H; Figure S6C). Collectively, these findings indicate that heterochromatin disruption stimulates target gene induction through upregulation of DUX4fl expression in mutant cells. Our results reveal that contraction of D4Z4 specifically makes this locus susceptible to additional heterochromatin destabilization to induce DUX4 expression.

Coherent feedforward loop of early and late DUX4 target genes

LEUTX is a DUX4 target gene that encodes for a transcription factor critical for zygotic genome activation in early development.[55] We observed significant increase of *LEUTX* transcript and protein signals in early mutant myotubes after 5AzaC treatment (Figures 5C and 6A, respectively). Interestingly, however, LEUTX induction in myoblasts is very weak (Figure 6A). This is in contrast to the significant induction of another DUX4 target H3.X/Y in myoblasts by 5AzaC (Figure 6A). We found that genes, such as *H3Y1*, *MBD3L2*, *KHDC1L*, are among the top DUX4 target genes activated, and are efficiently expressed and can be stimulated by 5AzaC in the myoblast stage in DEL_SM cells (Figures 2D and 6C; Figure S8). In contrast, *LEUTX*, *DUXA*, *RFPL1* and *KLF17*, are expressed only at a low level and fail to get stimulated by 5AzaC in the myoblast stage and are induced much more efficiently after cells were differentiated into myotubes (Figure 6C; Figure S8). These results reveal that there are two classes of DUX4 targets: early response genes insensitive to the differentiation status, and differentiation-dependent late target genes.

Histone H3Y1 (H3.Y) is one of the highest induced DUX4 targets in mutant cells (Figure 2D). H3.X/Y was found to be incorporated in the target gene regions and increased their expression.[56] Indeed, H3.X/Y binding to *MBD3L2*, *ZSCAN4* and *LEUTX* gene regions was all induced in 5AzaC-treated DEL_SM, but not control myotubes, providing an additional mechanism for increased target gene expression by 5AzaC (Figure 6B). Consistently, shRNA depletion of H3.X/Y effectively blocks *LEUTX* activation in DEL_SM myotubes (Figure 6D). The effect is nearly comparable to DUX4 depletion, emphasizing the strong reliance of DUX4 target gene expression on H3.X/Y (Figure 6D). Conversely, overexpression of H3.X significantly stimulated *LEUTX* expression in DEL_SM cells (Figure 6E). However, consistent with the notion that *LEUTX* is a differentiation-dependent late target gene, induction of LEUTX mRNA and protein is much more robust in myotubes than in myoblasts, despite comparable H3.X/Y overexpression in both cells (Figures 6E and 6G).

Overexpression of LEUTX also stimulated *H3.X/Y* (as well as *MDB3L2* and *TRIM43*), indicating a positive feedback loop (Figures 6F and 6G). Indeed, H3.X and H3.Y gene promoters both contain LEUTX (OTX2) binding motifs, raising the possibility that LEUTX directly upregulates *H3.X/Y* genes (Figure 6H). Intriguingly, genome-wide analyses of overexpression and depletion of H3.X/Y and LEUTX by RNA-seq reveal that both globally affect DUX4 target gene expression (Figures 6I and 6J). Thus, the results indicate a positive feedback loop between H3.X/Y and LEUTX (and possibly other TFs) enforcing target gene expression (Figure 7B).

DISCUSSION

In the present study, we investigated the effects of two major FSHD-related genetic mutations, D4Z4 contraction and/or loss of *SMCHD1*, in healthy human skeletal myoblasts during differentiation. SMCHD1 mutation is not only linked to FSHD2, but also increases the disease severity in FSHD1.[7,8] We found that, although loss of *SMCHD1* alone in adult myoblasts has only a minor effect on DUX4 target genes, this mutation strongly synergizes with D4Z4 contraction, resulting in significant activation of the DUX4 target gene network. Interestingly, this effect appears to be not through DNA methylation, but possibly through reduction of H3K9me3 (Figure 7A). Our results also reveal that D4Z4 contraction alone results in unstable/variegated phenotype, but makes this locus more susceptible to additional heterochromatin

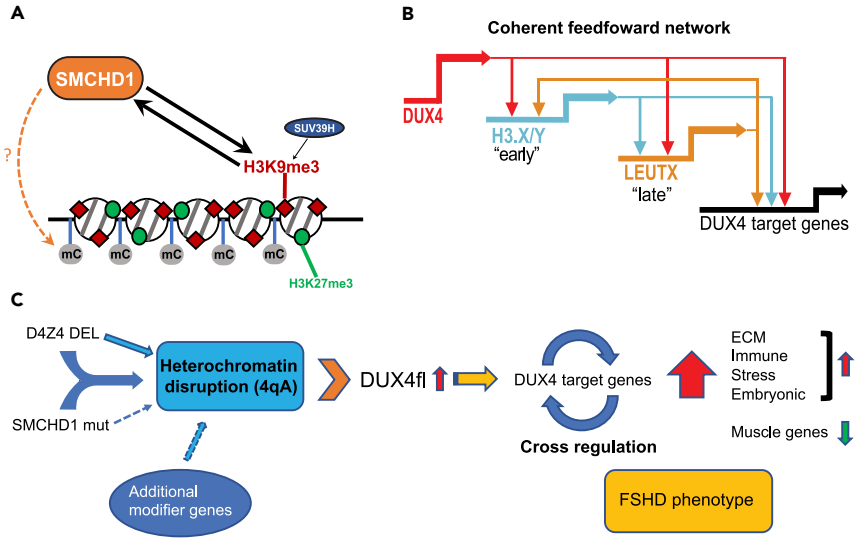


Figure 7. Schematic models of the establishment of the FSHD gene expression phenotype

Figure360▷ For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.isci.2024.109357.

(A) Positive feedback loop between H3K9me3 and SMCHD1. SMCHD1 interacts with D4Z4 chromatin in an H3K9me3-dependent manner[28] and also maintain H3K9me3.

(B) Coherent feedforward mechanism of DUX4 and target gene expression. While DUX4 is critical for the initial activation of its target genes, the early target H3.X/Y expression is essential for efficient expression of other downstream target genes, including the late target TF, LEUTX. LEUTX in turn promotes further expression of H3.X/Y. H3.X/Y as well as LEUTX (and possibly other DUX4 target TFs) contribute significantly to the expression of other DUX4 target genes.

(C) Two key processes in FSHD pathogenesis. D4Z4 heterochromatin disruption induced synergistically by D4Z4 and SMCHD1 mutations (and/or other epigenetic modifiers) enables stabilization and enhancement of DUX4fl expression. Once activated by DUX4, DUX4 target genes undergo cross-regulation contributing to the establishment of the FSHD gene expression phenotype.

disruption leading to efficient and more consistent DUX4 activation. Our results further provide the evidence for the self-amplification of the DUX4 gene network. We experimentally demonstrated two key processes in the establishment of the robust FSHD gene expression phenotype: (1) disruption of heterochromatin at D4Z4 repeats and (2) coherent feedforward loop of the DUX4 target gene expression, involving histone H3.X/Y and LEUTX (Figures 7B and 7C). Importantly, the analysis of late myotubes demonstrate that transcriptional activation of the endogenous DUX4 does not necessarily result in rapid self-elimination of myotubes by immediate cell death, but triggers gene expression amplification within the target gene network, which may be relevant to FSHD disease progression.

CRISPR-induced D4Z4 and SMCHD1 mutations to dissect FSHD mechanisms

While patient myoblasts remain important experimental tools to study FSHD, individual variability remains a problem, which motivated us to create isogenic mutant cells to specifically evaluate the effects of D4Z4 and SMCHD1 mutations and to gain insights into the mechanism of FSHD. Activating the endogenous DUX4 in a comparable frequency and level as in-patient cells in our mutant cells complement DUX4-overexpression cell system and mouse models. SMCHD1 CRISPR mutation resulted in homozygous knockout, rather than haploinsufficiency observed in ~30% of FSHD2 cases.[57,58] Nevertheless, SMCHD1 KO is not lethal in adult myoblasts, exhibited comparable phenotype as partial depletion by shRNA in terms of DUX4 target expression, and importantly exhibited robust phenotype when combined with D4Z4 deletion, thus providing a useful tool to investigate the relationship between SMCHD1 and D4Z4.

There have been limited applications of CRISPR approach to manipulate long repeat sequences: one study removed telomere repeats in human neuroblastoma cells.[59] In our first set of mutants, CRISPR mutation using gRNA specific to 4q D4Z4 repeat upstream of the *DUX4* transcription start site has led to repeat contraction as initially assessed by PFGE. Using nanopore long-read genomic sequencing, however, we found that repeats are susceptible to inversion and insertion, which cannot be discerned by PFGE. Nevertheless, we were able to confirm the expression of *DUX4fl* by nanopore long-read RNA sequencing. Although we detected a chimeric small transcript expressed from the inserted plasmid sequence, we confirmed that it does not affect target gene expression. It should be mentioned, however, even including the inverted repeats, the total repeat number is 8, which is significantly shortened and within the FSHD1 disease range (less than 10 copies). Perhaps it is not surprising that repetitive sequences can be recombinogenic though CRISPR targeting appears to be 4q-specific and 10q remained intact. Using RNA-based CRISPR delivery in the second set of deletion mutants, we were able to avoid the unwanted insertion of DNA plasmid into the cut site as well as repeat inversion, and generated D4Z4 deletion mutants with one D4Z4 repeat at 4qA locus. In this approach however, CRISPR appears to become less selective to 4qA and repeat deletion occurred in both 4q and 10q D4Z4 alleles. We did not see any significant phenotype difference in terms of DUX4 target gene expression between the first and second sets of mutants.

SMCHD1 knockout synergizes with D4Z4 contraction to stimulate DUX4 target gene expression without affecting DNA methylation at D4Z4

It is now appreciated that D4Z4 repeats are semi-contracted in FSHD2, which may synergize with *SMCHD1* mutation.[7] Despite relatively short 4qA repeats (~20) in our parental cells, however, our somatic *SMCHD1* mutation alone resulted in a weak phenotype with no effect on DNA methylation. Mouse Smchd1 was originally identified to play a role in the maintenance of DNA methylation at CpG islands and inactive X chromosome.[30–32] Since FSHD2 cells (with *SMCHD1* mutations) tend to exhibit a strong DNA hypomethylation phenotype[24] (and this study), it was speculated that *SMCHD1* mutation results in loss of DNA methylation leading to DUX4 upregulation. Indeed, it was shown that D4Z4 remethylation during reprogramming cannot take place in FSHD2 iPSCs.[47] A recent study showed that mouse Smchd1 binds and antagonizes Tet enzymes.[60] Consequently, loss of *Smchd1* in mouse ES cells leads to reduced DNA methylation and upregulation of *Dux,* a functional homolog of human *DUX4*.[60] Thus, it is possible that SMCHD1 mutations interfere with the initial establishment of DNA methylation at D4Z4 during the early development. However, *SMCHD1* gene correction in FSHD2 iPSCs failed to increase DNA methylation.[61] Thus, the role of SMCHD1 in D4Z4 DNA methylation remains obscure. Interestingly, 5AzaC-induced DNA hypomethylation in our *SMCHD1* mutant cells was not sufficient to recapitulate the robust DUX4 target gene upregulation seen in FSHD2 patient cells (with *SMCHD1* mutation) used in our study. These observations hint at the existence of yet another modifier(s) that synergizes with *SMCHD1* mutations to in FSHD2 (Figure 7C). Nevertheless, when combined with further deletion of D4Z4 repeats, SMCHD1 mutation led to robust DUX4 target gene expression, strongly suggesting that this synergy is linked to the increased disease severity seen in patients with double mutations. This is consistent with a recent report that increased DUX4 target expression correlates with increased pathology.[53]

SMCHD1-H3K9me3 feedback at D4Z4

The fact that *SMCHD1* mutation did not affect D4Z4 DNA methylation, yet cooperated strongly with D4Z4 contraction in activating DUX4 target gene expression in our mutant cells, strongly suggests that SMCHD1 suppresses DUX4 in a DNA methylation-independent manner. It has been reported that SMCHD1 can suppress gene expression in a DNA methylation-independent manner.[33] We previously demonstrated that D4Z4 heterochromatin is marked by H3K9me3, which is required for the recruitment of HP1γ, cohesion, and SMCHD1.[27,28] Reduction of H3K9me3 leads to decrease of SMCHD1 binding, accompanied by increased expression of DUX4.[28] In the currently study, albeit more variable than in DEL mutants, we found that *SMCHD1* mutation reduced H3K9me3 and increased DUX4 target gene expression in a modest, but statistically significant manner, suggesting the positive feedback loop between SMCHD1 and H3K9me3 contributing to *DUX4* suppression. H3K9me3 was further reduced in double mutants compared to D4Z4 deletion alone, also supporting this notion. Interestingly, reduction of H3K9me3 was also observed in D4Z4 transgenic mice crossed with smchd1 mutant mice.[62] Furthermore, restoration of SMCHD1 expression in FSHD2 iPSCs increased the level of H3K9me3 and HP1γ at D4Z4 chromatin.[61] SMCHD1 is recruited to and compacts human inactive X chromosome in part through interaction with HP1-bound H3K9me3 chromatin via HP1-binding protein HBiX1 (LRIF1).[63,64] Suggestively, a homozygous mutation of *LRIF1* has also been linked to FSHD2.[65] Thus, disruption of the observed positive feedback loop between SMCHD1 and H3K9me3 at D4Z4 may be critical for FSHD pathogenesis.

Robust DUX4 target gene expression in late myotubes

In addition to DUX4 target genes, D4Z4 disruption combined with loss of SMCHD1 in our mutant cells cluster closer to patient cells in PCA plot, and yielded a gene-expression pattern that recapitulated key features of patient myocytes. These include upregulation of genes related to ECM, immune and stress responses as well as embryonic genes, and downregulation of muscle genes.[13,19,38,51,66] Crucially, DUX4 depletion reversed these changes in gene expression, strongly suggesting that they are DUX4-dependent. In contrast to the almost complete suppression of the DUX4 target gene network, this reversal is partial, suggesting that expression changes of these pathways are indirect downstream effects. Notably, changes in the expression of genes involved in the regulation of apoptosis (both anti- and pro) are more prominent in myoblasts, and appear to taper off later in myotube differentiation when DUX4 target genes are most highly induced. We previously failed to detect any significant cell death in patient cells on differentiation day 7 when DUX4 target gene expression was detectable in 40% of myotubes.[40] Immortalization may make patient and mutant cells more protective from cell death. However, lack of overt cytotoxic phenotype or apoptotic transcriptomic signature was also observed in our previous single nucleus RNA-seq analyses of primary patient myoblasts/myotubes.[41] Taken together, our results indicate that activation of the DUX4 target gene network can be robust without triggering apoptotic genes. Therefore, even though DUX4 expression was shown to be immediately toxic in early myotubes,[23] there may be more complex possibly dose-dependent downstream gene expression phenotype in late myotubes. Further analyses of dynamics and consequences of DUX4 and target gene expression during muscle differentiation will be important.

Differentiation-insensitive "early" H3.X/Y and sensitive "late" LEUTX are part of coherent positive feedforward amplification of DUX4 target genes

Our results revealed that DUX4 target genes can be sub-divided into "early" and "late" genes based, not on timing but, on their differential dependence on myotube differentiation. A group of "early" genes, including histone variants H3.X and/or H3.Y, is efficiently induced in myoblasts. H3.X/Y were previously shown to be incorporated into the DUX4 target gene regions and promote their expression.[56] Consistent with this, we found that H3.X/Y depletion almost completely inhibits the expression of LEUTX, a differentiation-dependent target gene. Interestingly, another differentiation-independent early DUX4 target, MBD3L, can disrupt the repressive functions of MBD2 or MBD3, and upregulate

DUX4 expression as a positive feedback regulator.[67,68] Thus, these early DUX4 targets may act as epigenetic initiators that set the stage to promote activation of the DUX4 gene network following differentiation. In contrast, "late" genes such as LEUTX, are not efficiently induced until cells differentiate into myotubes. It should be noted that LEUTX is expressed during ZGA in embryonic stem cells and in the recombinant DUX4-overexpressed myoblasts.[50,69,70] Thus, the results suggest presence of an additional differentiation-coupled mechanism of DUX4 target gene activation in the context of myocytes with the endogenous DUX4, which may be more relevant to the disease. Nevertheless, overexpression or depletion of H3.X/Y as well as LEUTX globally affects DUX4 target expression. It is possible that LEUTX may directly bind and control these genes and/or indirectly promote their expression through feedback activation of H3.X/Y (and possibly other TFs) in differentiated myotubes. Alternatively, H3.X/Y and LETUX may contribute to further upregulation of DUX4 though there is no strong evidence that they directly upregulate the DUX4 gene.[56,71] Thus, in this case, the effect may be indirect, for example through upregulation of MDB3L.[67,68] Taken together, our results strongly suggest that DUX4 triggers the sequential induction of epigenetic regulators and downstream transcription factors to ensure a "coherent feedforward effect"[72] on the DUX4 gene network (Figure 7C).

In conclusion, our results demonstrate that contraction of D4Z4 alone results in unstable DUX4 gene network activation, which is effectively stabilized by the loss of a disease modifier SMCHD1 and/or DNA methylation inhibition, strongly suggesting that stabilization of variegated phenotype is the critical driver of the disease. Our results suggest that at least in adult myocytes, SMCHD1 modulates H3K9me3, but not DNA methylation, at D4Z4. Furthermore, our results revealed a hierarchy within the DUX4 target genes and significant contributions of differentiation-independent and -dependent transcription regulators in the DUX4 signal amplification. However, since mutations are introduced in adult myocytes, developmental effects of these mutations cannot be assessed. Thus, examining the mutation effects earlier in development would be important. Also, it will be highly informative to analyze the phenotypes of these mutations in different muscle cell types as well as in different genetic backgrounds. Although CRISPR mutations do not completely mirror mutations observed in patients, the results lay an important ground to dissect the consequences of D4Z4 contraction and its interaction with disease modifiers, and provided important insights into the molecular mechanism of FSHD as a heterochromatin abnormality disorder. It will be interesting, for example, to test the effect of other known modifiers, such as DNMT3B, LRIF1, FAT1 and estrogens, in these mutant cells.[73–75] The mutant cells described in the present study will be valuable resources for further investigation of the disease pathogenesis, and may serve as a possible platform for biomarker discovery and therapy development.

Limitations of the study

Although this is the first study to artificially contract D4Z4 in healthy myoblasts to observe activation of the endogenous DUX4 gene network, mutations induced by CRISPR-Cas9 are not exactly the same as the D4Z4 contractions seen in patients (the first set of mutants inadvertently acquired CRISPR plasmid insertion and repeat inversion though 10qD4Z4 repeats were spared; the second set of mutants showed contractions (without any insertion or inversion) at both 4q and 10q alleles). Immortalization may also affect the phenotype. Since mutations were introduced in adult myoblasts, their effects during early development were not assessed. Although SMCHD1 knockout had no effect on DNA methylation in adult myoblasts, it is possible that it may affect DNA methylation during the early stage of development. Although DUX4 dependency of the target gene induction was clearly demonstrated by shRNA depletion experiments, it is not possible to quantitatively detect very rare expression of the endogenous *DUX4* gene products unless cells were treated with 5AzaC or later in differentiation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Generation of SMCHD1 knockout mutants in immortalized permissive control myoblast with CRISPR-Cas9
  - Generation of D4Z4 contraction mutants with CRISPR-Cas9
  - Cell culture and differentiation
  - Pulsed-field gel electrophoresis (PFGE) and Southern blotting for 4q/10q D4Z4 repeat array length analysis
  - Optical genome mapping
  - Genomic and RNA nanopore long-read sequencing
  - Methylation analysis of nanopore long-read sequencing
  - Immunofluorescent staining
  - Western blotting
  - RNA isolation and quantitative real-time RT-PCR (RT-qPCR)
  - RNA-seq and data processing
  - RNAScope *in situ* hybridization

- ChIP-qPCR analysis
- MeDIP
- Detection the ratio of 4qA- and 10q-specific nucleotide polymorphisms (SNPs) using amplicon sequencing
- 5-Azacytidine (5AzaC) and 5-Aza-2′-deoxycytidine (5AzadC) treatment
- ShRNA depletion or overexpression of proteins using lentiviral systems

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109357.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of the Chao Family Comprehensive Cancer Center Optical Biology Core (LAMMP/OBC) Shared Resource and UCI Genomics Research and Technology Hub. This work was funded in part by National Institutes of Health grants P01NS069539 (R.T.) and R01AR071287 (K.Y. and A.M.), and Japan Agency for Medical Research and Development grant 20jk0210009 (TK).

AUTHOR CONTRIBUTIONS

Conceptualization, K.Y., X.K., and A.M.; Methodology, X.K., N.N., J.S., T.K., S.T., T.K., K.W., A.M., and K.Y.; Investigation, X.K., N.N., Y.L., J.S., J.C., S.M., K.W., S.S., and A.B.; Visualization, N.N., X.K., J.C., and A.B.; Supervision, K.Y. and A.M.; Writing—original draft: N.N., X.K., J.S., and K.Y.; Writing—review & editing, N.N., X.K., K.T., A.M., and K.Y.

DECLARATION OF INTERESTS

Authors declare that they have no competing interests.

Received: June 9, 2023
Revised: September 20, 2023
Accepted: February 23, 2024
Published: February 29, 2024

REFERENCES

1. Deenen, J.C.W., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J.G.M., Bakker, E., Weinreich, S.S., Verbeek, A.L.M., and van Engelen, B.G.M. (2014). Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurol. *83*, 1056–1059.
2. Wang, L.H., and Tawil, R. (2021). Current therapeutic approaches in FSHD. J. Neuromuscul. Dis. *8*, 441–451.
3. van der Maarel, S.M., and Frants, R.R. (2005). The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. *76*, 375–386.
4. van der Maarel, S.M., Tawil, R., and Tapscott, S.J. (2011). Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol. Med. *17*, 252–258.
5. Jones, T.I., Chen, J.C.J., Rahimov, F., Homma, S., Arashiro, P., Beermann, M.L., King, O.D., Miller, J.B., Kunkel, L.M., Emerson, C.P., Jr., et al. (2012). Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Hum. Mol. Genet. *21*, 4419–4430.
6. Lemmers, R.J.L.F., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W.E., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat. Genet. *44*, 1370–1374.
7. Sacconi, S., Lemmers, R.J.L.F., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. Am. J. Hum. Genet. *93*, 744–751.
8. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Müller, C.R. (2015). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. Eur. J. Hum. Genet. *23*, 808–816.
9. Gordon, C.T., Xue, S., Yigit, G., Filali, H., Chen, K., Rosin, N., Yoshiura, K.I., Oufadem, M., Beck, T.J., McGowan, R., et al. (2017). De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. Nat. Genet. *49*, 249–255.
10. Shaw, N.D., Brand, H., Kupchinsky, Z.A., Bengani, H., Plummer, L., Jones, T.I., Erdin, S., Williamson, K.A., Rainger, J., Stortchevoi, A., et al. (2017). SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. Nat. Genet. *49*, 238–248.
11. Mul, K., Lemmers, R.J.L.F., Kriek, M., van der Vliet, P.J., van den Boogaard, M.L., Badrising, U.A., Graham, J.M., Jr., Lin, A.E., Brand, H., Moore, S.A., et al. (2018). FSHD type 2 and Bosma arhinia microphthalmia syndrome: Two faces of the same mutation. Neurology *91*, e562–e570. https://doi.org/10.1212/wnl.0000000000005958.
12. Gabriëls, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., and Belayew, A. (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene *236*, 25–32.
13. 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., Tawil, R., and Tapscott, S.J. (2012). DUX4 activates germline genes, retroelements, and immune mediators: Implications for facioscapulohumeral dystrophy. Dev. Cell *22*, 38–51.
14. Snider, L., Geng, L.N., Lemmers, R.J.L.F., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., and Miller, D.G. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. *6*, e1001181.
15. 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., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. *49*, 925–934.
16. Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the

DUX4-family gene network. Nat. Genet. *49*, 935–940.
17. Lemmers, R.J.L.F., van der Vliet, P.J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science *329*, 1650–1653.
18. Himeda, C.L., Jones, T.I., and Jones, P.L. (2015). Facioscapulohumeral muscular dystrophy as a model for epigenetic regulation and disease. Antioxidants Redox Signal. *22*, 1463–1482.
19. Tsumagari, K., Chang, S.C., Lacey, M., Baribault, C., Chittur, S.V., Sowden, J., Tawil, R., Crawford, G.E., and Ehrlich, M. (2011). Gene expression during normal and FSHD myogenesis. BMC Med. Genom. *4*, 67.
20. Broucqsault, N., Morere, J., Gaillard, M.C., Dumonceaux, J., Torrents, J., Salort-Campana, E., Maues De Paula, A., Bartoli, M., Fernandez, C., Chesnais, A.L., et al. (2013). Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy. Hum. Mol. Genet. *22*, 4206–4214.
21. Ferreboeuf, M., Mariot, V., Bessières, B., Vasiljevic, A., Attié-Bitach, T., Collardeau, S., Morere, J., Roche, S., Magdinier, F., Robin-Ducellier, J., et al. (2014). DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Hum. Mol. Genet. *23*, 171–181.
22. Rahimov, F., King, O.D., Leung, D.G., Bibat, G.M., Emerson, C.P., Jr., Kunkel, L.M., and Wagner, K.R. (2012). Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. Proc. Natl. Acad. Sci. USA *109*, 16234–16239.
23. 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. Hum. Mol. Genet. *24*, 5901–5914.
24. van Overveld, P.G.M., Lemmers, R.J.F.L., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.-J.B., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat. Genet. *35*, 315–317.
25. de Greef, J.C., Wohlgemuth, M., Chan, O.A., Hansson, K.B., Smeets, D., Frants, R.R., Weemaes, C.M., Padberg, G.W., and van der Maarel, S.M. (2007). Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. Neurol. *69*, 1018–1026.
26. de Greef, J.C., Lemmers, R.J.L.F., van Engelen, B.G.M., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R., and van der Maarel, S.M. (2009). Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum. Mutat. *30*, 1449–1459.
27. Zeng, W., de Greef, J.C., Chen, Y.-Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1γ/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet. *5*, e1000559.
28. Zeng, W., Chen, Y.Y., Newkirk, D.A., Wu, B., Balog, J., Kong, X., Ball, A.R., Jr., Zanotti, S., Tawil, R., Hashimoto, N., et al. (2014). Genetic and Epigenetic Characteristics of FSHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. Hum. Mutat. *35*, 998–1010.
29. Zeng, W., Ball, A.R., and Yokomori, K. (2012). The epigenetics of facioscapulohumeral muscular dystrophy. In Epigenomics: From Chromatin Biology to Therapeutics, K. Appasani, ed. (Cambridge University Press), pp. 347–361.
30. Ashe, A., Morgan, D.K., Whitelaw, N.C., Bruxner, T.J., Vickaryous, N.K., Cox, L.L., Butterfield, N.C., Wicking, C., Blewitt, M.E., Wilkins, S.J., et al. (2008). A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome Biol. *9*, R182.
31. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat. Genet. *40*, 663–669.
32. Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell *23*, 265–279.
33. Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M., Heard, E., and Brockdorff, N. (2013). Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Mol. Cell Biol. *33*, 3150–3165.
34. Bosnakovski, D., Xu, Z., Gang, E.J., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppée, F., et al. (2008). An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. EMBO J. *27*, 2766–2779.
35. Vanderplanck, C., Ansseau, E., Charron, S., Stricwant, N., Tassin, A., Laoudj-Chenivesse, D., Wilton, S.D., Coppée, F., and Belayew, A. (2011). The FSHD atrophic myotube phenotype is caused by DUX4 expression. PLoS One *6*, e26820.
36. Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S.M., Tapscott, S.J., and Bradley, R.K. (2015). A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. Elife *4*, e04996. https://doi.org/10.7554/eLife.04996.
37. 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 Genet. *13*, e1006658.
38. Jagannathan, S., Ogata, Y., Gafken, P.R., Tapscott, S.J., and Bradley, R.K. (2019). Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. Elife *8*, e41740.
39. Lek, A., Zhang, Y., Woodman, K.G., Huang, S., DeSimone, A.M., Cohen, J., Ho, V., Conner, J., Mead, L., Kodani, A., et al. (2020). Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. Sci. Transl. Med. *12*, eaay0271.
40. Chau, J., Kong, X., Viet Nguyen, N., Williams, K., Ball, M., Tawil, R., Kiyono, T., Mortazavi, A., and Yokomori, K. (2021). Relationship of DUX4 and target gene expression in FSHD myocytes. Hum. Mutat. *42*, 421–433.
41. 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 Genet. *16*, e1008754.
42. Homma, S., Beermann, M.L., Boyce, F.M., and Miller, J.B. (2015). Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. Ann. Clin. Transl. Neurol. *2*, 151–166.
43. Bosnakovski, D., Chan, S.S.K., Recht, O.O., Hartweck, L.M., Gustafson, C.J., Athman, L.L., Lowe, D.A., and Kyba, M. (2017). Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. Nat. Commun. *8*, 550.
44. Jones, T., and Jones, P.L. (2018). A cre-inducible DUX4 transgenic mouse model for investigating facioscapulohumeral muscular dystrophy. PLoS One *13*, e0192657.
45. Yao, Z., Snider, L., Balog, J., Lemmers, R.J.L.F., 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. Hum. Mol. Genet. *23*, 5342–5352.
46. Krom, Y.D., Dumonceaux, J., Mamchaoui, K., den Hamer, B., Mariot, V., Negroni, E., Geng, L.N., Martin, N., Tawil, R., Tapscott, S.J., et al. (2012). Generation of Isogenic D4Z4 Contracted and Noncontracted Immortal Muscle Cell Clones from a Mosaic Patient: A Cellular Model for FSHD. Am. J. Pathol. *181*, 1387–1401.
47. Dion, C., Roche, S., Laberthonnière, C., Broucqsault, N., Mariot, V., Xue, S., Gurzau, A.D., Nowak, A., Gordon, C.T., Gaillard, M.C., et al. (2019). SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. Nucleic Acids Res. *47*, 2822–2839.
48. Goselink, R.J.M., Schreuder, T.H.A., van Alfen, N., de Groot, I.J.M., Jansen, M., Lemmers, R.J.L.F., van der Vliet, P.J., van der Stoep, N., Theelen, T., Voermans, N.C., et al. (2018). Facioscapulohumeral Dystrophy in Childhood: A Nationwide Natural History Study. Ann. Neurol. *84*, 627–637. https://doi.org/10.1002/ana.25326.
49. Morris, T.A., Naik, J., Fibben, K.S., Kong, X., Kiyono, T., Yokomori, K., and Grosberg, A. (2020). Striated myocyte structural integrity: Automated analysis of sarcomeric z-discs. PLoS Comput. Biol. *16*, e1007676.
50. 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. Hum. Mol. Genet. *25*, 4419–4431.
51. Winokur, S.T., Chen, Y.W., Masny, P.S., Martin, J.H., Ehmsen, J.T., Tapscott, S.J., van der Maarel, S.M., Hayashi, Y., and Flanigan, K.M. (2003). Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Hum. Mol. Genet. *12*, 2895–2907.
52. Jones, T.I., Chew, G.L., Barraza-Flores, P., Schreier, S., Ramirez, M., Wuebbles, R.D., Burkin, D.J., Bradley, R.K., and Jones, P.L. (2020). Transgenic mice expressing tunable levels of DUX4 develop characteristic facioscapulohumeral muscular dystrophy-like

pathophysiology ranging in severity. Skeletal Muscle *10*, 8. https://doi.org/10.1186/s13395-020-00227-4.

53. 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., Campbell, A.E., et al. (2019). MRI-informed muscle biopsies correlate MRI with pathology and DUX4 target gene expression in FSHD. Hum. Mol. Genet. *28*, 476–486.
54. Daxinger, L., Tapscott, S.J., and van der Maarel, S.M. (2015). Genetic and epigenetic contributors to FSHD. Curr. Opin. Genet. Dev. *33*, 56–61. https://doi.org/10.1016/j.gde.2015.08.007.
55. Jouhilahti, E.M., Madissoon, E., Vesterlund, L., Töhönen, V., Krjutškov, K., Plaza Reyes, A., Petropoulos, S., Månsson, R., Linnarsson, S., Bürglin, T., et al. (2016). The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. Development *143*, 3459–3469.
56. 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 Rep. *29*, 1812–1820.e5.
57. Balog, J., Goossens, R., Lemmers, R.J.L.F., Straasheijm, K.R., van der Vliet, P.J., Heuvel, A.V.D., Cambieri, C., Capet, N., Feasson, L., Manel, V., et al. (2018). Monosomy 18p is a risk factor for facioscapulohumeral dystrophy. J. Med. Genet. *55*, 469–478. https://doi.org/10.1136/jmedgenet-2017-105153.
58. Lemmers, R.J.L.F., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2015). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Hum. Mol. Genet. *24*, 659–669. https://doi.org/10.1093/hmg/ddu486.
59. Kim, H., Ham, S., Jo, M., Lee, G.H., Lee, Y.S., Shin, J.H., and Lee, Y. (2017). CRISPR-Cas9 mediated telomere removal leads to mitochondrial stress and protein aggregation. Int. J. Mol. Sci. *18*, 2093. https://doi.org/10.3390/ijms18102093.
60. Huang, Z., Yu, J., Cui, W., Johnson, B.K., Kim, K., and Pfeifer, G.P. (2021). The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. Sci. Adv. *7*, eabb9149. https://doi.org/10.1126/sciadv.abb9149.
61. 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. Hum. Mol. Genet. *27*, 4024–4035.
62. de Greef, J.C., Krom, Y.D., den Hamer, B., Snider, L., Hiramuki, Y., van den Akker, R.F.P., Breslin, K., Pakusch, M., Salvatori, D.C.F., Slütter, B., et al. (2018). Smchd1 haploinsufficiency exacerbates the phenotype of a transgenic FSHD1 mouse model. Hum. Mol. Genet. *27*, 716–731. https://doi.org/10.1093/hmg/ddx437.
63. Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struct. Mol. Biol. *20*, 566–573.
64. Brideau, N.J., Coker, H., Gendrel, A.V., Siebert, C.A., Bezstarosti, K., Demmers, J., Poot, R.A., Nesterova, T.B., and Brockdorff, N. (2015). Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. Mol. Cell Biol. *35*, 4053–4068.
65. Hamanaka, K., Šikrová, D., Mitsuhashi, S., Masuda, H., Sekiguchi, Y., Sugiyama, A., Shibuya, K., Lemmers, R.J.L.F., Goossens, R., Ogawa, M., et al. (2020). Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. Neurol. *94*, e2441–e2447.
66. 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. Hum. Mol. Genet. *27*. R153-r162. https://doi.org/10.1093/hmg/ddy162.
67. Campbell, A.E., Shadle, S.C., Jagannathan, S., Lim, J.W., Resnick, R., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2018). NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. Elife *7*, e31023. https://doi.org/10.7554/eLife.31023.
68. Jin, S.G., Jiang, C.L., Rauch, T., Li, H., and Pfeifer, G.P. (2005). MBD3L2 interacts with MBD3 and components of the NuRD complex and can oppose MBD2-MeCP1-mediated methylation silencing. J. Biol. Chem. *280*, 12700–12709. https://doi.org/10.1074/jbc.M413492200.
69. Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., and Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program in vitro. Cell Stem Cell *29*, 449–459.e6. https://doi.org/10.1016/j.stem.2022.01.014.
70. Yu, X., Liang, S., Chen, M., Yu, H., Li, R., Qu, Y., Kong, X., Guo, R., Zheng, R., Izsvák, Z., et al. (2022). Recapitulating early human development with 8C-like cells. Cell Rep. *39*, 110994. https://doi.org/10.1016/j.celrep.2022.110994.
71. Gawriyski, L., Jouhilahti, E.M., Yoshihara, M., Fei, L., Weltner, J., Airenne, T.T., Trokovic, R., Bhagat, S., Tervaniemi, M.H., Murakawa, Y., et al. (2023). Comprehensive characterization of the embryonic factor LEUTX. iScience *26*, 106172. https://doi.org/10.1016/j.isci.2023.106172.
72. Peter, I., and Davidson, E. (2015). Genomic Control Process: Development and Evolution (Elsevier).
73. Park, H.J., Lee, W., Kim, S.H., Lee, J.H., Shin, H.Y., Kim, S.M., Park, K.D., Lee, J.H., and Choi, Y.C. (2018). FAT1 Gene Alteration in Facioscapulohumeral Muscular Dystrophy Type 1. Yonsei Med. J. *59*, 337–340. https://doi.org/10.3349/ymj.2018.59.2.337.
74. Teveroni, E., Pellegrino, M., Sacconi, S., Calandra, P., Cascino, I., Farioli-Vecchioli, S., Puma, A., Garibaldi, M., Morosetti, R., Tasca, G., et al. (2017). Estrogens enhance myoblast differentiation in facioscapulohumeral muscular dystrophy by antagonizing DUX4 activity. J. Clin. Invest. *127*, 1531–1545. https://doi.org/10.1172/jci89401.
75. Tihaya, M.S., Mul, K., Balog, J., de Greef, J.C., Tapscott, S.J., Tawil, R., Statland, J.M., and van der Maarel, S.M. (2023). Facioscapulohumeral muscular dystrophy: the road to targeted therapies. Nat. Rev. Neurol. *19*, 91–108. https://doi.org/10.1038/s41582-022-00762-2.
76. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
77. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323.
78. McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297. https://doi.org/10.1093/nar/gks042.
79. Wick, R.R., Judd, L.M., and Holt, K.E. (2018). Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLoS Comput. Biol. *14*, e1006583. https://doi.org/10.1371/journal.pcbi.1006583.
80. Ehrlich, M., Jackson, K., Tsumagari, K., Camaño, P., and Lemmers, R.J.F.L. (2007). Hybridization analysis of D4Z4 repeat arrays linked to FSHD. Chromosoma *116*, 107–116.
81. Shiomi, K., Kiyono, T., Okamura, K., Uezumi, M., Goto, Y., Yasumoto, S., Shimizu, S., and Hashimoto, N. (2011). CDK4 and cyclin D1 allow human myogenic cells to recapture growth property without compromising differentiation potential. Gene Ther. *18*, 857–866.
82. Zeng, W., Jiang, S., Kong, X., El-Ali, N., Ball, A.R., Jr., Ma, C.I.H., Hashimoto, N., Yokomori, K., Mortazavi, A., Yokomori, K., and Mortazavi, A. (2016). Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity. Nucleic Acids Res. *44*, e158.
83. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44–53. https://doi.org/10.1126/science.abj6987.
84. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-SMCHD1 (1:2000)	Abcam	Cat# ab31865; RRID:AB_777986
Mouse monoclonal anti-DUX4 (1:500)	Novus Bio.	Cat# NBP2-12886
Rabbit polyclonal anti-LEUTX (1:2000)	Thermofisher	Cat# PA5-59595; RRID:AB_2643351
Mouse monoclonal anti-Actin (1:2000)	Sigma	Cat# A4700; RRID:AB 476730
Rabbit polyclonal anti-H3 (1:2000)	Abcam	Cat# ab18521; RRID:AB_732917
Rat monoclonal anti-H3. X/Y (1:1000 for WB, 1:500 for IF)	Sigma	Cat# MABE243-I
Mouse monoclonal anti-H3.X/Y (ChIP)	Active Motif	Cat# 61161; RRID:AB_2793533
Rabbit polyclonal anti-H3K9me3 (ChIP)	Abcam	Cat# ab8898; RRID:AB_306848
Alexa 555 Donkey anti-Mouse IgG (1:1000)	Thermofisher	Cat# A31570; RRID: AB_2536180
Alexa 555 Donkey anti-Rabbit IgG (1:1000)	Thermofisher	Cat# A31572; RRID:AB_162543
Alexa 488 Donkey anti-Rat IgG (1:1000)	Thermofisher	Cat# A21208 RRID:AB_2535794
Anti-Rabbit IgG, HRP linked (1:10000)	Promega	Cat#4018; RRID:AB_430833
Anti-Mouse IgG, HRP linked (1:10000)	Promega	Cat#4028; RRID:AB_430834
Anti-Rat IgG, HRP-linked (1:10000)	Abcam	Cat#ab97057; RRID:AB_10680316
Bacterial and virus strains		
Stbl3™ Chemically Competent E. coli	Thermofisher	Cat# C737303
Chemicals, peptides, and recombinant proteins		
Puromycin	Sigma	Cat# P8833
FBS	Omega Scientific	Cat# FB-02
DMEM	Gibco	Cat# 11965
Pen-Strep	Gibco	Cat# 15140122
Ultrasor G	Crescent Chemical Co.	Cat# 67042
ITS	Thermofisher	Cat# 51300044
UltraPure Low Melting Point Agarose	Bio-Rad	Cat# 1613111
Pronase	Sigma	Cat# P5147
Alt-R S.p. HiFi Cas9 Nuclease V3	IDT	Cat# 1081060
Alt-R CRISPR-Cas9 tracrRNA	IDT	Cat# 1073190
Nuclease-free duplex buffer	IDT	Cat# 11-01- 03-01
Prolong Diamond Antifade Mountant	Thermofisher	Cat# P36961
2x Laemmli Sample Buffer	Bio-Rad	Cat#1610737
4,6-Diamidine-2-phenylindole dihydrochloride (DAPI)	Sigma	Cat# D9542
Pierce Protein-Free T20 Blocking Buffer	Thermofisher	Cat# 37571
SuperSignal™ West Pico PLUS Chemiluminescent Substrate	Thermofisher	Cat# 34580
5-Azacytidine	Sigma	Cat# A2385
5-Aza-2′-deoxycytidine	Sigma	Cat# A3656
Critical commercial assays		
QuickExtract DNA extraction kit	Epicentre	Cat# QE09050
RNeasy Plus Mini kit	Qiagen	Cat# 74134
QIAquick PCR Purification Kit	Qiagen	Cat# 28104
SuperScript IV VILO Master	Thermofisher	Cat# 11756050

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AzuraView GreenFast qPCR Blue Mix LR	Azura Genomics	Cat# AZ-2320
TaqMan Fast Advanced Master Mix	Thermofisher	Cat# 4444557
EpiMark® Methylated DNA Enrichment Kit	New England Biolabs	Cat# E2600S
Lipofectamine™ CRISPRMAX™ Cas9 Transfection Reagent	Thermofisher	Cat# CMAX00003
Lipofectamine 3000	Thermofisher	Cat# L3000001
Bionano Genomics DNA Isolation Kit	Bionano Genomics	Cat# 90057
DLS Labeling Kit	Bionano Genomics	Cat# 80005
Cas9 Sequencing Kit	Oxford Nanopore Technologies	Cat# SQK-CS9109,
Flow Cell (R9.4.1)	Oxford Nanopore Technologies	Cat# FLO-MIN106D
MinION Mk1B	Oxford Nanopore Technologies	Cat# MIN-101B
Ligation Sequencing Kit	Oxford Nanopore Technologies	Cat# SQK-LSK110
NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing	New England Biolabs	Cat# E7180S
RNAScope Multiplex Fluorescent Reagent Kit v2	Advanced Cell Diagnostics	Cat# 323100
NEBuilder® HiFi DNA Assembly Cloning Kit	New England Biolabs	Cat# E5520S
Nextera DNA Flex Library Prep Kit	Illumina	Cat# 20018704
Deposited data		
RNA-seq and nanopore seq data	This paper	dbGaP: phs002554.v2
Experimental models: Cell lines		
293T cell line	ATCC	Cat# CRL-3216
Control	This paper	N/A
FSHD1	This paper	N/A
FSHD2	This paper	N/A
SM1	This paper	N/A
SM2	This paper	N/A
SM3	This paper	N/A
DEL1	This paper	N/A
DEL2	This paper	N/A
DEL3	This paper	N/A
DEL4	This paper	N/A
DEL5	This paper	N/A
DEL7	This paper	N/A
DEL8	This paper	N/A
DEL9	This paper	N/A
DEL3_SM_A	This paper	N/A
DEL4_SM_A	This paper	N/A
DEL4_SM_B	This paper	N/A
DEL9_SM_A	This paper	N/A
DEL9_SM_B	This paper	N/A
Oligonucleotides		
Primers used in this study	IDT	See Table S4
TaqMan Gene Expression Assay probes	Thermofisher	See Table S5
LEUTX RNAscope probe (Hs-LEUTX-C2)	Advanced Cell Diagnostics	Cat# 547251-C2
6ZZ DUX4fl RNAscope probe (HS-DUX4-O6-C1)	Advanced Cell Diagnostics	Cat# 546151
4ZZ DUX4fl RNAscope probe (Hs-DUX4-O7-C2)	Advanced Cell Diagnostics	Cat# 1089191-C2
2ZZ DUX4fl RNAscope probe (Hs-DUX4-O8-C3).	Advanced Cell Diagnostics	Cat# 1089201-C3

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
X330-U6-Chimeric_BB-CBh-hSpCas9	Addgene	pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid # 42230; http://n2t.net/addgene:42230; RRID:Addgene_42230)
hCas9	Addgene	hCas9 was a gift from George Church (Addgene plasmid # 41815; http://n2t.net/addgene:41815; RRID:Addgene_41815)
TPC1-mCherry	Addgene	TPC1-mCherry was a gift from Antony Galione (Addgene plasmid # 135182; http://n2t.net/addgene:135182; RRID:Addgene_135182)
shSMCHD1	Sigma	TRCN0000253777
shDUX4	Sigma	TRCN0000421072
shCTRL	Sigma	SHC002
pLV[shRNA]-PuroU6>Scramble_shRNA#1	VectorBuilder	Cat# VB010000-0005mme
pLV[shRNA]-Puro-U6>shDUX4	VectorBuilder	Cat# VB210621-1328hvw
pLV[shRNA]-Puro-U6>shH3.X/Y	VectorBuilder	Cat# VB210621-1327rfe
pLVX_H3.X	This paper	N/A
pLVX_LEUTX	This paper	N/A
pH082 pU6-gRNA2.0-GFP	This paper	pH082 pU6-gRNA2.0-GFP is a kind gift from Dr. Gerd A. Blobel, Children's Hospital of Philadelphia.
Software and algorithms		
R programming language (version 4.2.0)	R Core Team	https://www.r-project.org/
STAR (version 2.5.1b)	Dobin at el.[76]	https://github.com/alexdobin/STAR
RSEM (version 1.3)	Li and Dewey at el.[77]	https://github.com/deweylab/RSEM
EdgeR (version 3.38.2)	McCarthy at el.[78]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Adobe Illustrator (version 26.3.1)	Adobe	https://www.adobe.com/products/illustrator.html
ggplot2 (Version 3.3.6)	Wickham	https://cran.r-project.org/web/packages/ggplot2/index.html
EnFocus FSHD analysis	Bionano Genomics	https://bionanogenomics.com/wp-content/uploads/2017/03/30321-Rev.-C-Bionano-Solve-Theory-of-Operation-Bionano-EnFocus-FSHD-Analysis.pdf
Bionano Access (v. 1.7.2)/Bionano Solve (v.3.) software	Bionano Genomics	https://bionano.com/software-products-support/
MinKNOW	Oxford Nanopore Technologies	https://help.nanoporetech.com/en/collections/3738245-minknow
Guppy version 6.0.1 + 652ffd1	Oxford Nanopore Technologies	https://community.nanoporetech.com/docs/prepare/library_prep_protocols/Guppy-protocol
Porechop	Wick et al.[79]	https://github.com/rrwick/Porechop
Megalodon	Oxford Nanopore Technologies	https://github.com/nanoporetech/megalodon
Integrative Genomics Viewer	MIT	https://igv.org/
Other		
CHEF-DR III Pulsed Field Electrophoresis system	Bio-Rad	Cat#170-3700
Biodyne B Membrane	KPL	Cat#60-00-50

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
protein G Dynabeads	Thermofisher	Cat# 10003D
Bioruptor	Diagnode	Cat# B01060001
LSM510 confocal laser microscope	Zeiss	LSM510
Fujifilm LAS-4000 Image Analyzer	GE Healthcare	LAS 4000
Typhoon scanner	GE Healthcare	Amersham Typhoon
Bionano Saphyr system	Bionano Genomics	Cat# 60325

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kyoko Yokomori (kyokomor@uci.edu).

Materials availability

FSHD mutant cell lines generated in this study are available upon request from the lead contact.

Data and code availability

Raw RNA-seq and Nanopore-seq data are available in the dbGAP with accession number phs002554.v2.

This paper does not report original code.

Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Generation of SMCHD1 knockout mutants in immortalized permissive control myoblast with CRISPR-Cas9

The SMCHD1-specific gRNA-1 and gRNA-2 were designed by CRISPRdirect (https://crispr.dbcls.jp/) (Figure S2A). The CRISPR-SMCHD1-1 and 2 constructs were generated by cloning the SMCHD1 gRNA sequences into Cas9 and gRNA expression vector X330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene), respectively. ~3 × 10^5 Immortalized human control myoblasts were seeded in a 35 mm cell culture dish. One day later, the cells were transfected with both 1 μg of CRISPR-SMCHD1-1 and 1 μg of CRISPR-SMCHD1-2 or CRISPR-SMCHD1-1 alone together with 0.5 μg of a puromycin-resistance plasmid using Lipofectamine 3000 (Thermo Fisher Scientific). The media was changed after 4 h. The next day the media was replaced with fresh media containing 2 μg/mL puromycin (Sigma) for 3 days. Single cell clones were isolated by FACS sorting into 96 well plates. Ten to 14 days later, genomic DNA of the single cell clones with good proliferation was extracted using a QuickExtract DNA extraction kit (Epicentre). The exon 23–24 region was amplified by a pair of PCR primers (SMCHD1_PCR) to check for genomic deletions on a 1.5% agarose gel. The deletion mutants were confirmed by Sanger sequencing (Azenta) and western blot. SMCHD1 knock out mutants were identified by western blot and pooled amplicon sequencing. Using custom barcoded primers (SMCHD1_seqPCR), pooled amplicons from multiple individuals were sequenced at Azenta to determine the genomic sequence at the gRNA target site of each cell line. The off-target loci with the highest prediction scores were amplified by a PCR primer pair (SMCHD1_off_target_PCR) and sequenced by Sanger sequencing (Azenta). For the SMCHD1 mutant cell lines used in this paper, the corresponding off-target sequence (5′-TTTTCAATTTCAGTCAACGA-3′, chr9:+72135246) was not changed.

Generation of D4Z4 contraction mutants with CRISPR-Cas9

The CRISPR-Cas9 system was used to delete D4Z4 repeat units on 4qA in permissive control myoblasts. Guide RNAs (gRNAs) target a D4Z4 "1-kb" subregion sequence, which excludes the regions that are repeated elsewhere in the genome[80] as well as the DUX4 gene/promoter region to avoid to induce DUX4 mutation (Figure S2B). Based on this sequence, gRNAs were designed for 4q D4Z4 using the CRISPR Design Tool (https://zlab.bio/guide-design-resources) with low predicted off-target effects. Two rounds of CRISPR-Cas9 induced D4Z4 repeat array contraction were performed to obtain D4Z4 contraction mutants with satisfactory repeat number. hCas9 (Addgene), gRNA-D4Z4-1/2 (in gRNA expression vector pH082 pU6-gRNA2.0-GFP) and a puromycin-resistance plasmid were cotransfected into parental myoblast as indicated in Figure S2B. One mutant with 10 units of 4qA D4Z4 repeat was used as parental cells for the 2nd round of D4Z4 deletion mutant generation. Alternatively, the Alt-R CRISPR-Cas9 genome editing system (IDT) was used for the recombinant Cas9 protein and gRNA delivery for single and double mutations of D4Z4 and SMCHD1 (DEL5, DEL7, DEL8, DEL9 and DEL9_SM, respectively). Alt-R S.p. HiFi Cas9 Nuclease V3 (IDT)/tracrRNA (IDT)/crRNA (IDT) RNP were delivered to the myoblasts using Lipofectamine CRISPRMAX Cas9 Transfection Reagent (Thermo Fisher Scientific). Transfections were performed as described in the IDT protocol for "Alt-R CRISPR/Cas9 System". The differentiation efficiency and DUX4 target gene MBD3L2 expression of the single colony cell lines were tested. Based on the results, several cell lines were subjected to PFGE and blot hybridization to confirm the size of D4Z4 regions at 4q and 10q as well as nanopore genomic sequencing (see below).

Cell culture and differentiation

Immortalized control, FSHD1, FSHD2 and control-derived mutant skeletal myoblast cells were grown in high glucose DMEM (Gibco) supplemented with 20% FBS (Omega Scientific, Inc.), 1% Pen-Strep (Gibco), and 2% Ultrasor G (Crescent Chemical Co.). The FSHD1 cells used carry one 4qA allele with 6 D4Z4 repeats. The FSHD2 cells used carry a heterozygous deletion of 5 bp in exon 10 of the SMCHD1 gene producing a loss of frame with premature stop codon at the deletion (c.1302-1306delTGATA), and 2 4qA alleles with 15 and 19 D4Z4 repeats, respectively. These two patient lines were chosen as FSHD prototypes from >10 patient cells based on their significant DUX4 target gene expression as well as proliferation and differentiation efficiencies. Immortalization and single cell clone isolation of primary FSHD1 myoblasts were performed as same as for Control and FSHD2 myoblasts.[40,81] Briefly, control, FSHD1 and FSHD2 cells were immortalized using hTERT with p16INK4a-resistant R24C mutant CDK4 (mtCDK4) and cyclin D1 which effectively retain proliferation and differentiation capabilities as described.[81] Single cell RNA-seq data of immortalized myoblasts was compared to that of primary human skeletal myocyte myoblast (HSMM) data in our previous study demonstrating a reasonable similarity.[82] Upon reaching 80% confluence, myoblast differentiation was induced by using high glucose DMEM medium supplemented with 2% FBS and 1% ITS (Thermo Fisher Scientific) supplement. Fresh differentiation media was changed every day. Differentiating cells were harvested on days 3–5 or days 12–14 after the onset of differentiation, as "early" or "late" myotubes, respectively, with differentiation efficiency of 60–80%. Differentiation efficiency was determined by ([the initial cell number] – [the number of mononucleated cells left on each day])/[the initial cell number] (Figure S4A) as well as expression of differentiation markers (e.g., MYH1, MYH3, and MYOG) by RT-qPCR or RNA-seq. The muscle myosin genes (MYL3 and MYH1) are expressed significantly higher in late myotubes than in early myotubes (Figure S4C).

Pulsed-field gel electrophoresis (PFGE) and Southern blotting for 4q/10q D4Z4 repeat array length analysis

Suspended myoblasts were mixed with melted 1% UltraPure Low Melting Point Agarose (Bio-Rad) at 37°C to form plugs, each containing ~1.5 x10^6 cells. The plugs were digested with pronase (Sigma), rinsed, and treated with enzymes (EcoRI plus HindIII or EcoRI plus BlnI, Roche). The digested DNA in plugs was subjected to PFGE. Electrophoresis was done in CHEF-DR III system (Bio-Rad) at 6 V/cm and 15°C for 13 h, with the switch time increasing linearly from 1 to 6 s. The gel was washed in 0.25 M HCl for 30 min to depurinate DNA fragments, rinsed with H_2O, washed in denaturation buffer (0.6 M NaOH, 0.4 M NaCl) for 30 min, and transferred in that solution to a Biodyne B Membrane (KPL). The membrane was neutralized, cross-linked by UV irradiation, and was probed with the 1-kb 4q/10q specific probe.[80] Southern blots were visualized using Typhoon scanner (GE Healthcare).

Optical genome mapping

High molecular weight (HMW) genomic DNA (gDNA) was extracted from fresh Control myoblast cells by using Bionano Genomics DNA Isolation Kit (Bionano Genomics) following manufacturer's guidelines. The rest testing was performed in UCI Genomics Research and Technology Hub. The HMW gDNA was diluted over period 2 weeks to ~45 ng/μl using TE buffer, gentle mixing, and overnight homogenization. The HMW gDNA was then tagged with a fluorescent probe at each DLE-1 (Direct Labeling Enzyme 1) enzyme binding site using the Direct Label and Stain (DLS) DNA Kit (Bionano Genomics). The DLS HMW gDNA was then linearized and optically mapped on the Bionano Saphyr system (Bionano Genomics). Image scans by the Saphyr produced a molecules file, in which each molecule is one HMW gDNA strand with labels at each DLE-1 binding site. These data were processed with the Bionano Solve software to align labeled molecules against the HG38 reference sequence predicted label pattern. Molecules aligned to the reference chromosome 4q35 or 10q26 regions were further collected to infer the length of the D4Z4 arrays using the custom EnFocus FSHD analysis (Bionano Genomics).

Genomic and RNA nanopore long-read sequencing

Genomic Nanopore libraries constructed from genomic DNA using Cas9 Sequencing Kit (Oxford Nanopore Technologies), Nuclease-free duplex buffer (IDT), and the following Alt-R CRISPR reagents from IDT: tracrRNA resuspended at 100 μM in TE pH 7.5, Cas9 nuclease V3, 3 different S. pyogenes Cas9 Alt-RTM crRNAs (two upstream target sites 5′-CCTATTAAACGTCACGGACA-3′ and 5′-GATACCGACAGCAA TAGTCC-3′ and one downstream target site '5-AAATCTTCTATAGGATCCAC-3') resuspended at 100 μM in TE pH 7.5. The Long Fragment Buffer (LFB) from the Cas9 Sequencing Kit was used during the wash steps. Libraries were loaded on R9.4.1 Flow Cells (Oxford Nanopore Technologies) and sequenced on MinION Mk1B instrument (Oxford Nanopore Technologies) using the MinKNOW software. Oxford Nanopore's base calling software, Guppy version 6.0.1 + 652ffd1, was run in super accurate (sup) mode with the dna_r9.4.1_450bps_sup.cfg configuration file.

RNA-seq Nanopore libraries were constructed from 200 fmol Illumina libraries using Ligation Sequencing Kit (Oxford Nanopore Technologies) and NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (New England Biolabs). The Short Fragment Buffer (SFB) from the Ligation Sequencing Kit was used during the wash steps. 50 fmol sample libraries were loaded on R9.4.1 Flow Cells and sequenced on MinION Mk1B instrument using the MinKNOW software. Oxford Nanopore's base calling software, Guppy version 6.0.1 + 652ffd1, was run in super accurate (sup) mode with the dna_r9.4.1_450bps_sup.cfg configuration file. Adapters were trimmed from reads with the Porechop software package by adding a custom adapter sequence that includes both the illumina primer and nanopore adapter (top: 5′-AATGTACTTCGTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTAC-3′ and bottom: 5′-GTACTCTGCGTTGATACCACTGCT TAGCAATACGT-3′). Reversed reads were flipped using a custom script.

Methylation analysis of nanopore long-read sequencing

Megalodon version 2.5.0 was used to call CpG methylation and map to reference with the following configuration files and parameters: –guppy-config dna_r9.4.1_450bps_modbases_5mc_cg_sup.cfg –remora-modified-bases dna_r9.4.1_e8 sup 0.0.0 5mc CG 0 –mod-map-emulate-bisulfite –mod-map-base-conv C T –mod-map-base-conv m C. Reads were mapped to a custom reference which consists of CHM13[83] and an additional pseudo chromosome. The pseudo chromosome is a sequence consisting of DBET, one D4Z4 repeat and DUX4. Reads were visualized on the Integrative Genomics Viewer (IGV) version 2.14.1 in bisulfite mode (CG).

Immunofluorescent staining

Cells grown on coverslips were fixed in 4% paraformaldehyde for 10 min at room temperature, permeabilized with 0.5% Triton X-100 in PBS, and blocked in blocking buffer (0.02% saponin, 0.05% NaN3, 1% BSA, 4% horse serum and 0.1% gelatin in PBS) for 15 min at 37°C. The coverslips were incubated overnight with primary antibodies at 4°C followed by three PBS washes, then incubated with fluorescent secondary antibody for 30 min at 37°C, washed with PBS 3 times, counter-stained with DAPI(Sigma), and mounted with Prolong Diamond Antifade Mountant (Thermo Fisher Scientific). Images were acquired with a Zeiss LSM510 confocal laser microscope.

Western blotting

Cells were lysed in 2X Laemmli Buffer (Bio-Rad) with 5% beta-mercaptoethanol, sonicated, boiled, and separated by 4%–20% TEO-Tricine gel (Abcam). Then the samples were transferred to nitrocellulose membranes, blocked with Pierce Protein-Free T20 Blocking Buffer (Thermo Fisher Scientific), and blotted with the desired antibodies. Horseradish peroxidase-conjugated anti-mouse-IgG (Promega), anti-rabbit-IgG (Promega), or anti-rat-IgG (Abcam) were used as secondary antibodies. Immunoblots were developed with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific). Images were acquired with Fujifilm LAS-4000 Image Analyzer (GE Healthcare).

RNA isolation and quantitative real-time RT-PCR (RT-qPCR)

RNA was extracted using RNeasy Plus Mini kit (Qiagen), and complementary DNA (cDNA) was made using 500 ng of total RNA with SuperScript IV VILO Master (Thermo Fisher Scientific) following the manufacturer's instructions. qPCR was performed by using AzuraView GreenFast qPCR Blue Mix LR (Azura Genomics Inc.). The genes and their corresponding PCR primers are listed in Table S4. For some genes, the expression was detected by probe qPCR, which was performed by using TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific). The commercially available TaqMan Gene Expression Assay probes (Thermo Fisher Scientific) were listed in Table S5.

RNA-seq and data processing

Total RNA was extracted by using the RNeasy kit (QIAGEN). Between 12 and 50 ng of RNA was converted to cDNA using the Smart-Seq 2 protocol.[41] DNA libraries were constructed using Nextera DNA Flex Library Prep Kit (Illumina). DNA samples were sequenced on the Illumina NextSeq500 platform using paired-end 43 bp mode with around 15 million reads per sample. RNA-seq raw reads were aligned with STAR (version 2.5.1b) using human genome reference hg38. Alignment default parameters were applied except with a maximum of 10 mismatches per pair, a ratio of mismatches to read length of 0.07, and a maximum of 10 multiple alignments. Read count was performed using RSEM (version 1.3) by defaults with gene annotations from GENCODE v28, and raw read counts were extracted for downstream analysis. Genes were filtered based on raw counts with at least 2 counts in at least 2 samples. Raw count was normalized by TMM in EdgeR and then converted to TPM (transcript per millions). Differential genes were calculated using the cut off p-values $\leq$ 0.01 and Log_2 Fold Change $\geq$1.5.

RNAScope *in situ* hybridization

RNAScope was performed using RNAScope Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics, Inc.) according to the manufacturer's protocol.[40] The following RNAScope probes (Advanced Cell Diagnostics, Inc.) were used: LEUTX probe set (Hs-LEUTX-C2), 6ZZ DUX4fl probe set (HS-DUX4-O6-C1), 4ZZ DUX4fl probe set (Hs-DUX4-O7-C2), and 2ZZ DUX4fl probe set (Hs-DUX4-O8-C3). The 4ZZ and 2ZZ DUX4fl probes target 701–1388 and1481-1697 of DUX4fl mRNA (NM_001306068.2) respectively (Figure 5E).

ChIP-qPCR analysis

Around 3 x 10^6 cells were crosslinked with 1% formaldehyde for 10 min and quenched with 0.125 M glycine for 5 min. Cells were washed with ice-cold PBS and harvested in cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40, protease inhibitors). Nuclear extract was collected in RIPA buffer (1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, protease inhibitors). Nuclei were sonicated for 20 cycles (30 s on and off) using the Bioruptor (Diagnode) to obtain fragment length around 150–500 bp. Chromatin extract was quickly spun down to remove cell debris. Around 20 μg or 10 μg of chromatin were incubated overnight at 4°C with 10 μg H3.X/Y (Active Motif) or 2.5 μg H3K9me3 (Abcam) antibody respectively. Chromatin-antibody extracted and incubated with protein G Dynabeads (Thermo Fisher Scientific) for an hour. The mixture was washed with Lithium Chloride 5 times and TE buffer once on ice. Chromatin was eluted with 0.1 M NaHCO3 and 1% SDS. Proteinase K was added to chromatin and the mixture was proceeded to reverse-crosslinking for 2 h at 55°C. DNA was purified with QIAquick PCR Purification Kit (Qiagen). ChIP DNA was quantified by qPCR with specific primers.

MeDIP

Cells were washed with PBS and harvested in SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris HCl, pH 8.1). Nuclear extract was sonicated with 3 cycles (30 s on and off). Chromatin extract was quickly spun down and supernatant was collected. Chromatin was added with Proteinase K and incubated for 2 h at 55°C. DNA was purified with QIAquick PCR Purification Kit (Qiagen). Between 0.5 and 1 μg of chromatin was proceeded to MeDIP (Methylated DNA Immunoprecipitation) according to the manufacture protocol (EpiMark Methylated DNA Enrichment Kit, New England Biolabs). MeDIP DNA was quantified by qPCR with specific primers and amplicon sequencing.

Detection the ratio of 4qA- and 10q-specific nucleotide polymorphisms (SNPs) using amplicon sequencing

Input and MeDIP DNA were amplified by PCR using Phusion DNA Polymerase in two steps: The first PCR primers (D4Z4_seqPCR) are derived from the 4q/10q specific Q-PCR primers[27] with adapters for the 2nd PCR primers attachment. The second PCR primer pair (Illumina_seqPCR) was used to attach Illumina adaptors and to barcode samples. Amplification was carried out with 18 cycles for the first PCR and 24 cycles for the second PCR. Resulting amplicons from the second PCR were gel extracted, quantified, mixed, and sequenced using NovaSeq6000 (Illumina).[84] The ratio of 4qA/B - and 10q-derived D4Z4 sequences was calculated based on the reads number of their specific SNPs.

5-Azacytidine (5AzaC) and 5-Aza-2′-deoxycytidine (5AzadC) treatment

Six μM of 5AzaC (Sigma) was added to myoblasts at ~70% confluency for 48 h. With or without differentiation, the cells were harvested or fixed at indicated day after the drug treatment for subsequent RNA-seq, ChIP-qPCR, RT-qPCR, MeDIP and IFA analyses. Alternatively, myoblasts were incubated with 6 μM of 5AzadC (Sigma) for 24 h, followed by 2 days release. To assess the SMCHD1 depletion effect, cells were treated with 5 μM of 5AzaC for 48 h. For RNA-seq experiment, myoblasts were allowed to grow for additional 48 h before collection or before changing to differentiating media. For SMCHD1 mutant maintenance experiment, cells were allowed to recover for 48 h before collection.

ShRNA depletion or overexpression of proteins using lentiviral systems

The shRNA plasmids for SMCHD1 (shSMCHD1, 5′-TTATTCGAGTGCAACTAATTT-3′), DUX4 (shDUX4, 5′-AGATTTGGTTTCAGAATGAGA-3′) and a control shRNA (shCTRL, 5′-CAACAAGATGAAGAGCACCAA-3′), were obtained from the Sigma Mission library. The shRNA pLV[shRNA]-Puro-U6>shH3.X/Y (pLVshH3.X/Y, 5′- GCGGGAAATCAGAAAGTAC-3′, the siH3.X/Y targeting sequence in previous paper[56]), pLV[shRNA]-Puro-U6>shDUX4 (pLVshDUX4, 5′-GGCAAACCTGGATTAGAGTT-3′[21]), and corresponding pLV[shRNA]-PuroU6>Scramble_shRNA#1 (pLVshControl, 5′-CCTAAGGTTAAGTCGCCCTCG-3′) were synthesized and cloned in pLV[shRNA]-Puro-U6 vector by VectorBuilder. To construct the H3.X and LEUTX overexpression lentiviral plasmids pLVX_H3.X and pLVX_LEUTX, human H3.X and LEUTX ORF sequences were amplified from FSHD1 myotube cDNA using primer pairs H3.X_PCR and LEUTX_PCR respectively. Lentiviral plasmid TPC1-mCherry (Addgene) was digested with XbaI and XhoI to remove the insertion and used as pLVX Lentiviral vector. The PCR products were combined into the vector by using NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs) following the manufacturer's instructions except that Stbl3 Chemically Competent E. coli (Thermo Fisher Scientific) was used as competent cell. The overexpression lentiviral plasmids were confirmed by Sanger sequencing (Azenta). An empty pLVX vector was used as a negative control. Myoblasts in multiple wells were infected twice with lentivirus (produced in 293T packaging cell line) at a 24-h interval. One well of the infected cells was used to confirm the expected infection efficiency (>99%) by 4 μg/mL puromycin selection for 72 h. Cells in the rest of the wells were harvested without differentiation at 2–3 days after the 2nd infection (myoblasts) or induced to differentiate at 25 h after the 2nd infection (myotubes). Differentiated myotubes (60–80% fusion efficiency) were harvested at 3–7 days of differentiation as indicated in each figure.

QUANTIFICATION AND STATISTICAL ANALYSIS

Experiments were repeated at least three times. Excel (Microsoft Office 365, Microsoft) and R were used to perform statistical analyses. Quantitative data are presented as mean ± standard deviation (SD). Statistical comparisons were made using the unpaired Student's *t* test and Wilcoxon's t-test. Statistical significance between two samples was determined by a p value of less than 0.05. Details of statistical test used for each dataset are provided in the corresponding figure legend.