


# SCIENTIFIC REPORTS



OPEN

## Analysis of the 4q35 chromatin organization reveals distinct long-range interactions in patients affected with Facio-Scapulo-Humeral Dystrophy

Marie-Cécile Gaillard<sup>1</sup>, Natacha Broucqsault<sup>1</sup>, Julia Morere<sup>1</sup>, Camille Laberthonnière<sup>1</sup>, Camille Dion<sup>1</sup>, Cherif Badja<sup>1</sup>, Stéphane Roche<sup>1</sup>, Karine Nguyen<sup>1,2</sup>, Frédérique Magdinier<sup>1</sup> & Jérôme D. Robin<sup>1</sup> 

Facio-Scapulo Humeral dystrophy (FSHD) is the third most common myopathy, affecting 1 amongst 10,000 individuals (FSHD1, OMIM #158900). This autosomal dominant pathology is associated in 95% of cases with genetic and epigenetic alterations in the subtelomeric region at the extremity of the long arm of chromosome 4 (q arm). A large proportion of the remaining 5% of cases carry a mutation in the *SMCHD1* gene (FSHD2, OMIM #158901). Here, we explored the 3D organization of the 4q35 locus by three-dimensions DNA *in situ* fluorescent hybridization (3D-FISH) in primary fibroblasts isolated from patients and healthy donors. We found that D4Z4 contractions and/or *SMCHD1* mutations impact the spatial organization of the 4q35 region and trigger changes in the expression of different genes. Changes in gene expression were corroborated in muscle biopsies suggesting that the modified chromatin landscape impelled a modulation in the level of expression of a number of genes across the 4q35 locus in FSHD. Using induced pluripotent stem cells (iPSC), we further examined whether chromatin organization is inherited after reprogramming or acquired during differentiation and showed that folding of the 4q35 region is modified upon differentiation. These results together with previous findings highlight the role of the D4Z4 macrosatellite repeat in the topological organization of chromatin and further indicate that the D4Z4-dependent 3D structure induces transcriptional changes of 4q35 genes expression.

Over the recent years, sequence distribution within the nuclear space, nuclear topology and long-distance interactions have emerged as key elements in the regulation of gene expression and cell fate determination. The genome is separated from the cytoplasm by the nuclear envelope. The inner nuclear membrane is composed of specialized proteins, which contribute to higher-order organization of the genome within the nuclear space and the regulation of gene expression, with repressed loci mainly located at the nuclear periphery<sup>1,2</sup>. Interphase chromosomes reside in minimally overlapping chromosome territories with active genes mainly oriented toward the nuclear interior<sup>3</sup>. Individual genes are largely confined to their respective chromosome territories (CT). However, in certain developmental contexts, certain loci such as the *Hox* genes or genes on the inactive X chromosomes can loop out of their CTs<sup>4,5</sup>. Furthermore, long distance dynamic interactions between sequences or subnuclear domains have been identified by various genome-wide approaches<sup>6-8</sup>. Thus, different topological levels shape the human genome, from large scale folding involving *cis*- or *trans*- long-distance interactions to smaller interactions between regulatory elements located in close proximity<sup>9</sup>. Interestingly, topological conformation is highly conserved between cell types and species suggesting that it is sequence-dependent rather than controlled by cell

<sup>1</sup>Aix Marseille Univ, INSERM, MMG, U 1251, Marseille, France. <sup>2</sup>APHM, Laboratoire de Génétique Médicale, Hôpital de la Timone, Marseille, France. Marie-Cécile Gaillard, Natacha Broucqsault, Frédérique Magdinier and Jérôme D. Robin contributed equally. Correspondence and requests for materials should be addressed to F.M. (email: [frederique.magdinier@univ-amu.fr](mailto:frederique.magdinier@univ-amu.fr)) or J.D.R. (email: [Jerome.robin@univ-amu.fr](mailto:Jerome.robin@univ-amu.fr))

differentiation processes<sup>10,11</sup>. However, so far only a few sequences with topological activity have been identified and characterized<sup>12–14</sup>. Amongst them, repetitive elements involved in pathologies such as expanded triplets<sup>15</sup> or the D4Z4 macrosatellite<sup>12,13,16</sup> have been shown to cause topological modifications.

Regarding this disease-associated topological disorganization, one intriguing locus is the 4q35 subtelomere<sup>17,18</sup>. This gene-poor locus characterized by the presence of gene deserts and large blocks of repetitive DNA sequence, including blocks of (CA)<sub>n</sub> microsatellites, variable number of tandem 3.3 Kb D4Z4 macrosatellite repeats and  $\beta$  satellite elements is localized at the nuclear periphery, a nuclear compartment enriched in heterochromatin<sup>12,17–20</sup>. The 4q35 genes are organized in clusters separated by domains associated with the nuclear lamina (Lamin Associated Domains, LAD) and limited by CTCF-dependent boundaries<sup>21,22</sup>.

We have previously described the D4Z4 macrosatellite as the only known element able to tether any telomere at the nuclear periphery and control the replication timing of its abutting telomeric region. As usually observed for heterochromatin-rich regions, the 4q35 telomere replicates late, at the end of the S phase<sup>23</sup> but displays features of repressed euchromatin rather than constitutive heterochromatin<sup>24</sup>.

Shortening of the D4Z4 array at the 4q35 locus is linked to Facio-Scapulo-Humeral Dystrophy (FSHD), the third most common hereditary myopathy. In 95% of cases, FSHD is linked to the deletion of an integral number of D4Z4, on one of the two 4q35 alleles<sup>25,26</sup>. Affected individuals exhibit between 1–10 repeats whereas the general population have a number of repeats above 10 (up to 100)<sup>27</sup>. Shortening of this repetitive the array is associated with DNA hypomethylation and relaxation of the macrosatellite chromatin suggesting the involvement of epigenetic changes in the disease<sup>16,28–30</sup>. In agreement, the remaining 5% of patients usually display a profound hypomethylation and 80% of them carry a mutation in gene encoding the SMCHD1 (Structural Maintenance of Chromosomal Hinge Domain Containing 1) protein<sup>31</sup>. In mice, *Smchd1* loss of function results in early lethality for female embryos, attributed to derepression of genes on the inactive X chromosome<sup>32</sup>. *Smchd1* is also involved in silencing of repetitive DNA sequences and formation of long distance loops at the *Hox* genes loci and inactive X chromosome<sup>33</sup>.

As a consequence of the D4Z4 array hypomethylation, the current patho-physiological model for FSHD implicates the ectopic expression of *DUX4*, a retrogene present in each D4Z4 unit but whose protein is encoded by the last D4Z4 and downstream sequence. In this scheme, functional *DUX4* transcripts are produced by the last unit and stabilized in individuals carrying a permissive 4qA haplotype, that includes a poly adenylation signal (pLAM) in the direct proximity of the last D4Z4 unit. Three subtelomeric 4q35 haplotypes have been identified (A, B, C), with A and B being predominant and equally distributed in the general population<sup>34</sup>. Of note, D4Z4 units are also found at other loci such as the short arm of acrocentric chromosomes. In particular, the subtelomeric 10q locus shares 98% homology with the 4q35 locus but is not associated with FSHD.

By using a three-dimensional Fluorescent *In Situ* Hybridization approach (3D FISH), we have previously shown the existence of functional interactions between D4Z4, the nuclear lamina and the telomere<sup>12,13,16,23</sup> with a key role for D4Z4 in the organization and regulation of long-distance interactions at this locus. These observations were further corroborated by chromatin conformation capture approaches that revealed the influence of the contracted units on the 3D organization with chromatin loops encompassing domains beyond the D4Z4 units<sup>35–37</sup>. The higher-order regulation of this region is also modulated by telomere length<sup>16</sup>, suggesting possible cooperation between different specialized genomic elements in the topological organization of this region. Hence, partitioning of the 4q35 region into different topological domains raises the question of the regulation of long-distance interactions across the locus and the role of the D4Z4 array in this higher-order organization.

In order to get a precise insight into the 3D DNA interactions within the 4q35 locus in normal and pathological conditions, *i.e.*, influence of D4Z4 array size or *SMCHD1* mutation, we investigated the three-dimensional chromatin organization of the locus using hybridization techniques in primary cells isolated from FSHD individuals and controls. We identified different loops organizing the 4q35 locus into different chromatin domains. In agreement with previous findings and a role for D4Z4 in chromatin topology, these chromatin domains are modified when the number of D4Z4 units is decreased. Moreover, we analyzed expression of a number of genes localized within or outside of these loops in our cells and confronted our results to those obtained in muscle biopsies from affected and unaffected individuals.

## Results

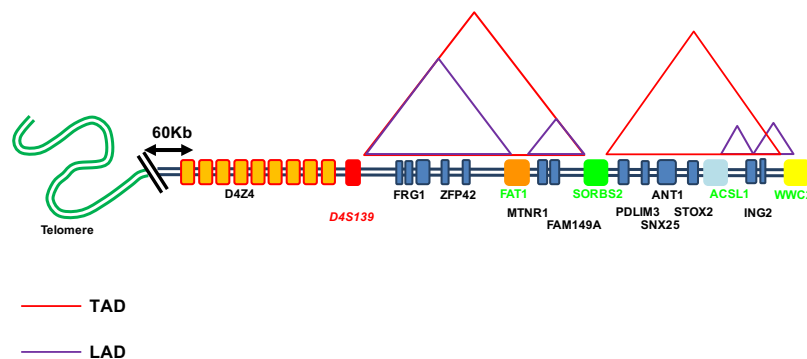
**Exploring the 3D organization of the 4q35 locus.** Given the peculiar organization of the 4q35 region and the topological role played by D4Z4 in healthy and diseased cells<sup>12,13,16</sup>, we aimed at deciphering the higher-order changes of this locus in FSHD1 and 2 patients compared to healthy donors. To this aim, we first analyzed datasets available in the literature<sup>6</sup>. Indeed, in the past decade, a wealth of chromatin techniques coupled to deep-sequencing technologies have emerged and multiple databases for 3D DNA organization are available<sup>11,38,39</sup>. From these, we retrieved 4q35 locus interaction maps (183–191 Mb of Chr. 4) generated from 3 different cell lines and 1 muscle biopsy (Supplementary Fig. 1). Amongst these maps, two achieved a Kb resolution (GM12878, IMR90) while the remaining two barely achieved a Mb resolution (STL002, *Psoas* muscle biopsy), a poor resolution mostly due to the depth of sequencing, number of assays performed and nature of the samples.

Altogether these maps obtained from different cell types show a conserved domain organization for the most subtelomeric part of the locus up to a region located between the *FAT1* and *SORBS2* genes (187 and 186 Mb of Chr.4; respectively), two genes previously implicated in FSHD<sup>16,40,41</sup>. When observed at a Mb level, these maps suggest that the chromatin organization and domains (TADs) are set, independently of cell types and tissue of origin, *i.e.*, conserved between lymphocytes, fibroblasts and muscle (*Psoas*).

Given the involvement of genomic rearrangement of this subtelomeric region in FSHD, we next set out to investigate the 4q35 chromatin organization by 3D DNA FISH using selected probes in well-defined cells (Table 1). Unlike deep sequencing techniques such as chromatin conformation capture (3C, HiC) or replication

Sample	Status	4q	10q	Age	Sex	hIPSCs	SMCHD1
A1	Control	>10RU	NA	Foetus	♀	Available	WT
A2 (ID: AG08498)	Control	>10RU	NA	1	♂		WT
A3	Control	>20RU	NA	21	♂		WT
N (ID: 12759)	FSHD1	7RU qA > 10RU qB	NA	51	♀	Available	WT
M (ID: 12566)	FSHD1	3.5RU qA 22RU qB	NA	54	♀		WT
F (ID: 120723)	FSHD1	6RU qA > 10RU qB	NA	28	♀		WT
T (ID: TaIF)	FSHD1	2RU qA > 10RU qB	NA	13	♂		WT
H	FSHD1	8RU qA 8RU qA	NA	29	♂		WT
G (ID: 11440)	FSHD2	12RU qA 32RU qA	14UR qA 17UR qA	37	♂		c.2338 + 4A > G intron 18 r.2261_2337del; p.S754*
S (ID: 15166)	FSHD2	12RU qA 23RU qB	13UR qB 20UR qB	67	♀		c.4614_4615 insTATAATA; r.4614_4615 insTATAATA; p.A1539Yfs*4

**Table 1.** Summary of cells (fibroblasts from skin biopsies) used in the study with their respective associated status and number of D4Z4 repeats (RU) at the 4q and 10q locus along with mutation found in *SMCHD1*. FSHD individuals were diagnosed by clinical examination by neurologists with clinical expertise in the evaluation of neuromuscular diseases. The genetic status was confirmed by genetic testing (Southern Blot, DNA combing). Gender and age at biopsy are indicated as well as the corresponding identification number (ID) of samples used elsewhere<sup>56</sup>.



**Figure 1.** Organization of the 4q35 region. Graphical representation of the subtelomeric region of chromosome 4 (4q35 locus) from the telomere to the *WWC2* gene. Using available data from the literature (*i.e.*, HiC, ChIA-PET) we report the detected TADs (Red) and LADs (purple) found and preserved across cell types.

timing assays, this low-resolution approach with a limitation of resolution of approximately 300Kb has the unique advantage of allowing the analysis of each individual allele at a single cell level. To further investigate whether D4Z4 array shortening below the 11-units threshold is associated to 4q35 topological changes, we used primary fibroblasts isolated from skin biopsies from five different FSHD1 patients, one of which presenting a homozygous D4Z4 contraction (*i.e.*, 8 UR) along with fibroblasts isolated from two FSHD2 individuals (*i.e.*, >10 UR) carrier of a mutation in *SMCHD1* (Table 1) or healthy donors.

As no clear TADs were found in more centromeric regions and given the putative interactions between the FSHD locus and the nuclear lamina or matrix<sup>12,16–18,37</sup>, we focused our attention on the different regions separated by LADs<sup>22</sup> and the conserved TAD revealed by available deep sequencing datasets<sup>11,39,42</sup> (Fig. 1). The 4q35 locus contains four LADs. The most distal one (190–191 Mb of Chr 4) is located upstream of the D4Z4 array and overlaps with a large gene-desert between the region containing the *ZFP42* and *FAT1* genes. The second LAD (187–187.7 Mb) separates the *FAT1* gene from the *SORBS2* region and encompasses several genes (including *MTNR1A* and *FAM149A*). The third (184.8–185.4 Mb) and fourth (185.6–185.8 Mb) LADs are situated in a gene-poor region, suggesting the existence of two loops anchored at the nuclear periphery between the *ACSL1* and *WWC2*-containing domains<sup>43</sup>. Interestingly, the *WWC2* region correspond to a TAD boundary in at least two data sets (GM12878, IMR90; respectively)<sup>6</sup>. To understand higher-order 3D folding patterns of the 4q35 region in the different contexts, we selected five probes encompassing the whole 4q35 regions and the different putative topological domains (Fig. 1).

**3D genome organization of the 4q35 locus is modified in FSHD.** Chromatin Conformation Capture based analysis (chromatin studies in cell populations, *e.g.* HiC) reveal a global conservation of genomic domains, regardless of the cell types. Considering the cell-by-cell resolution of FISH-based techniques and taking advantage of the conserved chromatin topology between cell types and tissues, we investigated the three-dimensional

organization of the 4q35 region of each individual allele *i.e.*, healthy and diseased alleles in primary fibroblasts from FSHD patients and healthy donors<sup>42,44</sup>. FSHD patients have all been carefully diagnosed at the clinical level by experienced neurologists and at the molecular level by DNA combing<sup>45,46</sup> (Table 1). For FSHD1, we used cases with a range of contracted D4Z4 arrays restricted to the 4qA-type haplotype (2, 4, 6, 7 and 8 Repeated Units (RU); respectively) representative of the heterogeneity seen in FSHD1. This group includes cells from a patient carrying two contracted 4qA allele (H). Likewise, D4Z4 repeats were >10RU in FSHD2 individuals who are carriers of a mutation in the *SMCHD1* gene.

We performed 3D DNA FISH in FSHD1 (n = 5), FSHD2 (n = 2) and control (n = 3) cells, using for each assay two sets of probes (Figs 2 and 3, Supplementary Figs 2–5). One probe was common between all experiments and hybridizes with the D4S139 minisatellite (red) proximal to the D4Z4 repetitive array and corresponding to the most proximal region specific to the 4q35 locus (Fig. 2A). Rationale for the choice of remaining probes is described above and targeted the *FAT1*, *SORBS2*, *ASCL1* and *WWC2* regions, respectively (green). After hybridization and confocal imaging, stacks of nuclei were processed using the IMARIS software (Bitplane, AG) for 3D reconstruction and analysis (n > 30 nuclei per sample and per condition; *i.e.*, 60 alleles) (Fig. 2B, Supplementary Fig. 2). Three Dimensions computational modeling allows one to retrieve accurate values from confocal imaging such as volume of nuclei and probes, spatial distribution, position (relative to the nuclear envelope) and distances between probes (Fig. 2C). We compiled results for each assay (4 assays; >300 nuclei analyzed per assay) and asked whether global 4q35 folding is identical between cells from controls and patients and between FSHD1 and FSHD2.

First, all signals were found at the nuclear periphery regardless of disease status and D4Z4 contractions (average distance from nuclear periphery < 0.3  $\mu\text{m}$ ; Figs 1 and 2) indicating that the size of the D4Z4 array does not modify the radial position of the locus within the nucleus. Next, we evaluated if signals between probes were separated or adjacent (noted sep or adj; respectively). This binary decision (sep, adj) was based on physical observations of separated or interacting objects (determined by IMARIS) and measurement of distances between the gravity center of the probe signal after 3D reconstruction (Supplementary Fig. 2). Overall, distances associated to adjacent and separated events were significantly different between conditions ( $p < 0.0001$ ; Mann-Whitney,  $\alpha = 0.05$ ). All assays considered (n = 1400 nuclei; 2800 signals) we measured a mean distance of 0.73  $\mu\text{m}$  ( $\pm\text{SD} = 0.23 \mu\text{m}$ ) for adjacent signals and 3.92  $\mu\text{m}$  ( $\pm\text{SD} = 1.43 \mu\text{m}$ ) for separated signals.

In control cells, the mean distance between the D4S139 region and *FAT1* was significantly higher than in FSHD cells (Fig. 2C, Mean  $\pm$  SD; CTRL = 2.26  $\pm$  SD = 0.98  $\mu\text{m}$  and FSHD = 0.73  $\pm$  SD = 0.24  $\mu\text{m}$ ;  $p < 0.0001$ ; Mann-Whitney,  $\alpha = 0.05$ ). Accordingly, signals were mostly separated in all control cells (90%) and adjacent in FSHD cells (Fig. 2D, FSHD1 = 75%; FSHD2 = 84%). Since both alleles behaved similarly (95%) regardless of number of repeats, our data suggest a modulation of the chromatin structure independent of the D4Z4 array (Fig. 2E).

Strikingly, regarding *SORBS2*, we observed a partial loss of interaction restricted to FSHD1 cells. Indeed, when considering FSHD1 cells carrying only one contracted allele (Table 1, N, M, F and T), 90% of cells displayed a partial loss of interaction where only one out of two signals, *i.e.*, allele, were separated (Fig. 3A). Accordingly, in FSHD1 cells carrying two contracted alleles (Table 1, H), the interaction was completely lost (Fig. 3,  $p < 0.001$ ; Chi-Square test,  $\alpha = 0.05$ ) suggesting that this interaction is alleviated when the number of D4Z4 repeats is decreased as previously observed<sup>16</sup>. In FSHD2 cells (Table 1, G and S), the interaction was preserved but a subset of cells (24%) displayed a complete loss of interaction (for both alleles) as opposed to a smaller subset in control cells (10%). This suggest a dynamic interaction influenced by other parameters (*e.g.*, *SMCHD1* status, DNA methylation, telomere length).

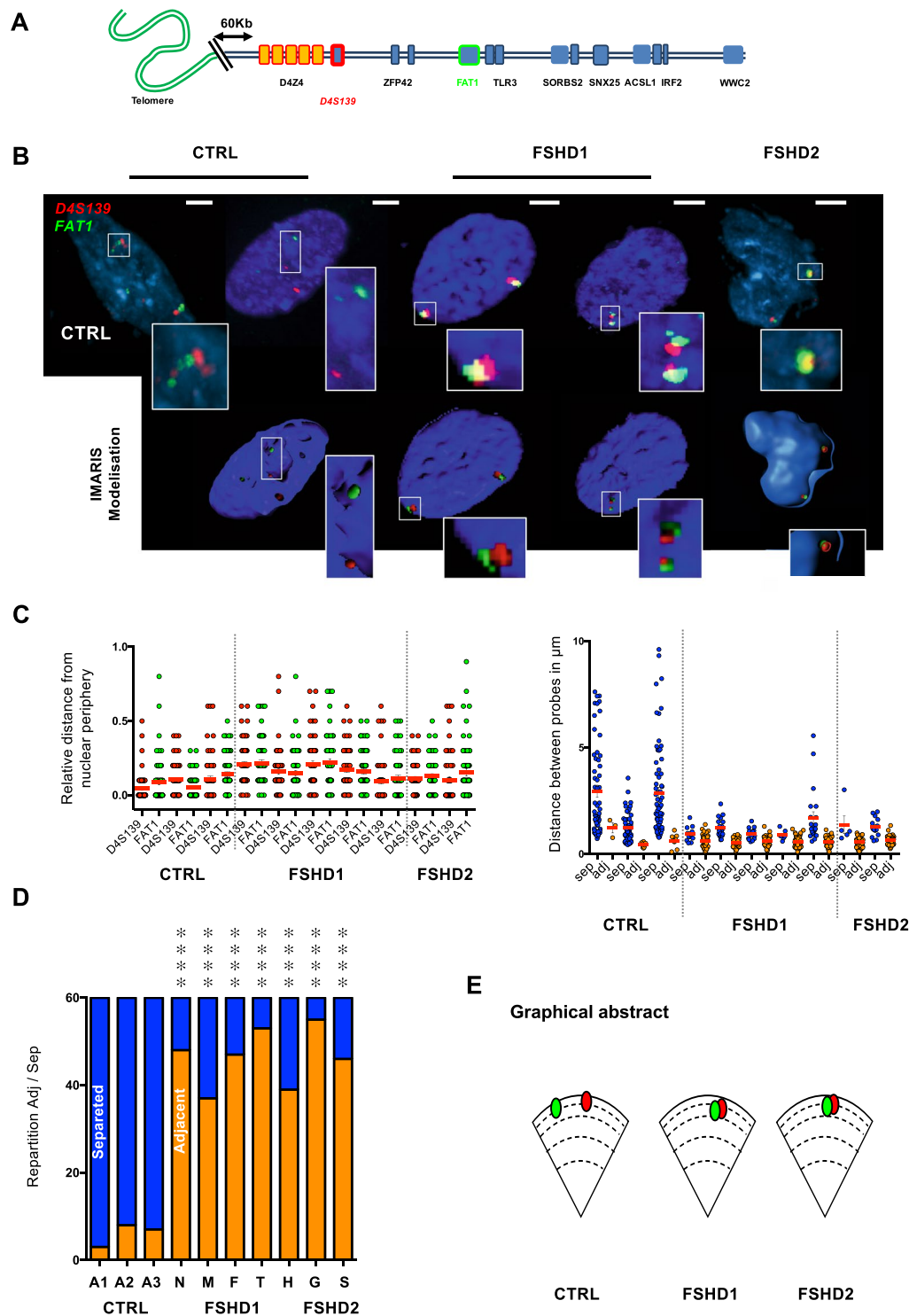
Next, regarding the *ASCL1* region, no difference was found between samples considering the D4S139 and *ASCL1* probes (Mean  $\pm$  SD; CTRL = 3.91  $\pm$  SD = 1.38  $\mu\text{m}$  and FSHD = 4.1  $\pm$  SD = 1.7  $\mu\text{m}$ ;  $p = 0.97$ ; Mann-Whitney,  $\alpha = 0.05$ ), for which signals were separated in all situations, further suggesting that compaction of chromatin was similar between all samples analyzed.

Last, we looked at the *WWC2* region and found a majority of adjacent signals (85%) with a probe targeting the D4S139 region, in control cells (Mean  $\pm$  SD; CTRL = 0.66  $\pm$  SD = 0.22  $\mu\text{m}$ ). Oppositely, probe signals were separated in most FSHD cells (62%) regardless of the number of D4Z4 repeats.

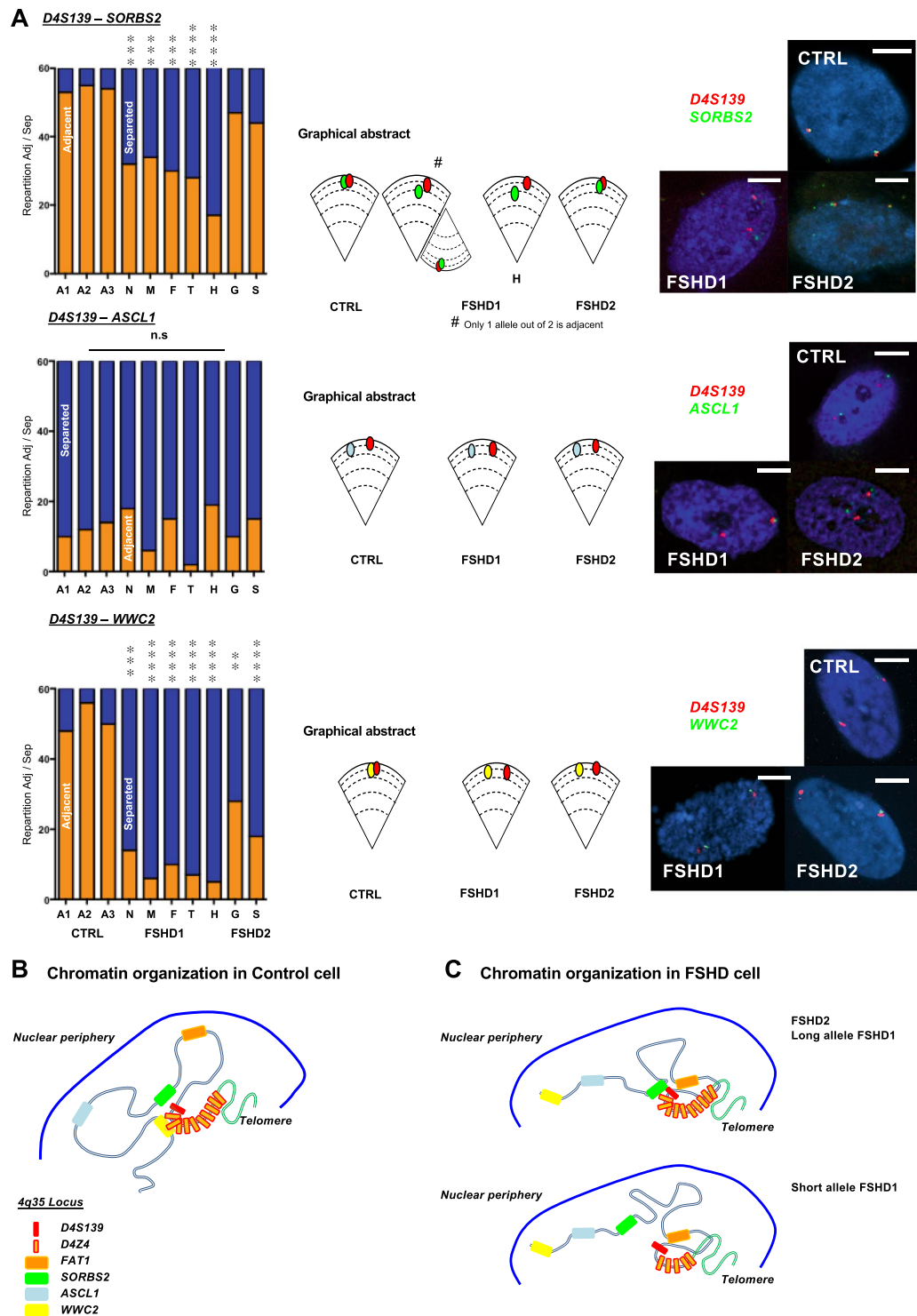
Altogether, our data show that in primary FSHD fibroblast, the D4S139 region no longer interacts with *WWC2* but with *FAT1* ( $p < 0.001$  for FSHD cases; Chi-Square test,  $\alpha = 0.05$ ). In control cells, the D4S139 region is localized in the vicinity of the *SORBS2* and *WWC2* regions, but separated from *FAT1* and *ASCL1* (all in different LADs), a finding reminiscent of the TAD identified in the literature (Fig. 1, Supplementary Fig. 1)<sup>6</sup>. In FSHD cells, interactions were identical between FSHD1 and 2 in all conditions but one (D4S139-*SORBS2*; Fig. 3A). This suggests that the topological organization of this locus is orchestrated by epigenetic features associated with the D4Z4 array, either in *-cis* (number of repeats for FSHD1) or *-trans* (*SMCHD1* status in FSHD2) and its peripheral subnuclear localization. Taken together our results detail the spatial organization of the 4q35 locus within the nucleus and reveal the existence of new topological domains that depend on the residual number of D4Z4 macrosatellite repeats (Fig. 3B,C).

### Changes in chromatin organization of the 4q35 locus correlate with modulation in gene expression.

To further investigate the consequences triggered by changes in DNA folding within the 4q35 region, we examined the level of expression of different genes located within or outside the modulated chromatin loops by RT-qPCR (Fig. 4). Hence, we exploited a restricted gene expression analysis to decipher the emergence of additional territories by analyzing expression of 14 genes located at the 4q35 locus in our fibroblasts (Fig. 4). In our cellular model, we first determined the level of transcription as the ratio between the gene of interest and house-keeping genes for normalization (Supplementary Fig. 6). Next, we analyzed genes localized in loops present in control cells (Fig. 4A) but absent in FSHD cells and *vice versa*. Interestingly, expression of most genes within the



**Figure 2.** The distance between the distal 4q35 region and the *FAT1* gene is decreased in cells (fibroblasts) from FSHD1 and 2 patients. **(A)** Graphical representation of the subtelomeric region of chromosome 4, from the telomere (green) to the *WWC2* gene (blue square) located at a distance of 7 Mb from the D4Z4 array. 3D FISH was performed using two sets of probes corresponding either to the D4S139 variable number tandem minisatellite (VNTR; red) or the *FAT1* region (green). **(B)** Representative pictures of FISH images in primary fibroblasts from healthy donors (control) or FSHD1 and FSHD2 patients and 3D reconstruction using IMARIS. Scale is indicated by a white bar (=2 $\mu\text{m}$ ). **(C)** Quantification of FISH signal was done using IMARIS. We measured the distance of each probe to the nuclear periphery (left) and distance between the probes (right). At least 30 nuclei were analyzed (60 alleles). For each cell, we evaluated if signals were separated or adjacent (sep, adj; respectively). **(D)** Summary of events and associated Chi-Square test (n = 30 cells per sample; 60 alleles). Frequency of adjacent signals is increased in all FSHD cells. \*\*\* $p < 0.001$ . **(E)** Summary of the average distances between probes and their respective distances to the periphery in each sample. Signals are mostly localized at the periphery regardless of disease status.



**Figure 3.** Distinct long distance interactions at the 4q35 region in controls and FSHD cells (fibroblasts). (A) Quantifications of colocalization and nuclear distribution of 3D DNA FISH assay within the 4q35 locus. We report the frequency of association between the D4S139 region (red; as in Fig. 2) and either the SORBS2, ASCL1 or WWC2 region (green, blue, yellow; respectively) in primary fibroblasts from healthy donors, FSHD1 or FSHD2 patients. A Chi-Square test was performed for each assay to determine statistical differences ( $n = 30$  cells per sample, 60 alleles). (B,C) Graphical representation of the 4q35 chromatin organization as revealed by 3D DNA FISH in control cells (B) and FSHD1 and 2 cells (C).  $**p < 0.01$ ;  $***p < 0.005$ ;  $****p < 0.001$ .

last 7 Mb of chromosome 4q was significantly modulated (global upregulation; Supplementary Fig. 6, Fig. 4B; e.g., excluding FAT1 and STOX2,  $p > 0.05$ ; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ), with a high variability between samples, including between individuals with the same genotype (Fig. 4B).

First, considering the D4S139-*FAT1* loop (Fig. 4A), we detected a significant upregulation of the *FRG1* and *ZFP42* genes in all FSHD cells regardless of the D4Z4 status (FSHD1 and FSHD2) ( $p < 0.05$ ; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ , Fig. 4B). *DUX4* was significantly upregulated only in FSHD1 fibroblasts ( $p = 0.0057$ ; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ) compared to control cells, but not detectable in the FSHD2 cells tested here.

Next, given the strong interactions involving the D4S139 probe and other loci (*FAT1*, *SORBS2*; Figs 2 and 3), we deduced a global interaction between the three regions. Indeed, we found an interaction between D4S139 and *FAT1* for both alleles in most FSHD cells ( $\pm 77.5\%$ ; Fig. 2) and between D4S139 and *SORBS2* for one allele in FSHD1 cells ( $\pm 90\%$ ) and both alleles in most FSHD2 cells (76%; Fig. 3). This highly suggests an interaction between *SORBS2* and *FAT1* in FSHD2 cells and an interaction involving the healthy non-contracted allele in FSHD1 cells. Expression of genes located within this newly formed loop was significantly upregulated not only in FSHD cells presenting the interaction but across all samples (*MTNR1A*, *FAM149A*). This suggest other modulators of expression, including a possible unidentified loop<sup>47</sup>. Strikingly, expression of *SORBS2* was directly correlated to the formation of the loop. Only cells where the D4S139-*SORBS2* interaction was not detected presented upregulation of its expression (FSHD1;  $p < 0.001$ ; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ , Fig. 4B). Of note, the *PDLIM3* gene, which shares a bidirectional promoter with *SORBS2*, displayed a downregulated expression in all FSHD cells.

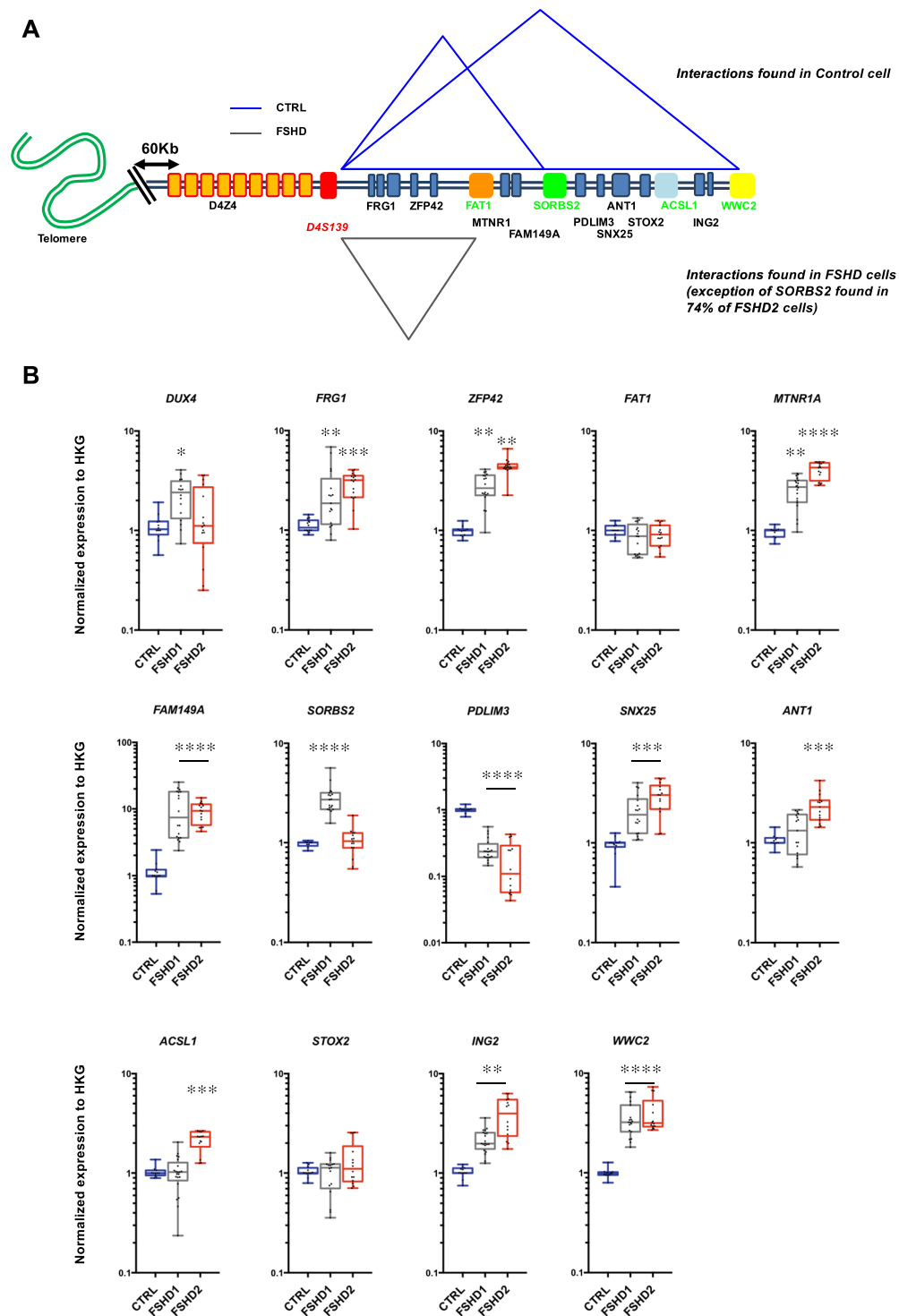
Our set of 3D FISH assay did not allow us to detect other chromatin structures beyond the *SORBS2* locus in FSHD cells, whereas control cells showed a D4S139-*WWC2* interaction. Expression of genes beyond the *SORBS2* locus displayed mixed results. The *ING2* and *WWC2* gene were upregulated in FSHD cells ( $p < 0.001$ ; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ) while *ANT1* and *ACSL1* were upregulated only in FSHD2 cells. Altogether, we found that in FSHD cells, levels of expression were globally increased beyond the *SORBS2* locus, advocating for a looser chromatin structure possibly permissive for transcription.

**4q35 genes are differentially expressed in muscle biopsies.** In primary fibroblasts, we found that the modification of the 4q35 chromatin landscape correlates with changes in transcription of genes located within the modified domains. In a second step, we thus consolidated our findings by analyzing gene expression directly in muscle biopsies isolated from individuals diagnosed with either FSHD1 or FSHD2. Hence, for confronting our *in vitro* data in primary fibroblasts with tissues affected in this pathology (*i.e.*, skeletal muscle), we used a collection of biopsies from control and affected individuals (FSHD1,  $n = 7$ ; FSHD2,  $n = 3$ ; Control,  $n = 7$ ; Table 2). As found by others, *DUX4* and *FRG1* expression were upregulated in FSHD biopsies (Fig. 5;  $p < 0.005$ ; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ )<sup>27,48,49</sup>. Additionally, *FAM149A* and *SORBS2* were exclusively upregulated in FSHD1 ( $p < 0.001$ , respectively; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ); *FAT1* and *ASCL1* in FSHD2 ( $p = 0.023$  and  $p = 0.006$ , respectively; Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ).

If one considers that chromatin territories are conserved among cell types, our results strongly suggest a direct correlation between 3D organization of the 4q35 locus (*e.g.*, no interaction in FSHD1) and *SORBS2* or *WWC2* expression in FSHD. Moreover, the intriguing upregulated expression of *ASCL1* and *FAT1* restricted to FSHD2 biopsies advocates for a more complex regulation possibly involving *SMCHD1* as described for other genomic regions<sup>33,50</sup>, or the existence additional DNA interaction yet to be characterized<sup>16</sup>.

**The 4q35 chromatin organization is set to a new conformation in hiPSC.** Earlier works have shown that the position of single gene loci and entire chromosomes is heritable after cell division<sup>51</sup>. TADs are largely maintained through development as TAD boundaries tend to be 65% similar among cell types<sup>42,52,53</sup>. However, HiC data revealed massive transformation in chromosome topology after reprogramming with many interactions present in the cells of origin being progressively erased from early to late passage<sup>54</sup>. Furthermore, the nuclear lamina undergoes profound changes since A-type Lamins present in the somatic cell-of-origin are not expressed in pluripotent cells<sup>55</sup>. Our recent findings showed that D4Z4 methylation is increased upon reprogramming in cells from FSHD1 patients<sup>56</sup> suggesting chromatin changes at the 4q35 locus in pluripotent cells. We therefore investigated whether chromatin folding and 3D organization was inherited or modified after reprogramming by taking advantage of the chromatin organization in FSHD1 cells carrying both a contracted and non-contracted *D4Z4* allele (N, Table 1), thus displaying the 3D organization found across all FSHD cells (FSHD1 and FSHD2). We restrained our 3D DNA FISH analysis to hiPSCs generated from the corresponding fibroblasts (A and N; respectively). hiPSCs cells were characterized and described elsewhere<sup>56,57</sup>. We analyzed the 4q35 conformation using probes for the D4S139 region together with probes targeting the *FAT1*, *SORBS2*, *ACSL1* and *WWC2* loci (Fig. 6). Remarkably, no difference was found between control and FSHD1 hiPSCs. Signals were all localized at the nuclear periphery and only one long-distance interaction was preserved after reprogramming between D4S139 and *SORBS2* (Fig. 6B), consistent with the main TAD found in other cell types (Supplementary Fig. 1). Indeed, the longest interaction observed in control cells encompassing D4S139 and the *WWC2* region (7 Mb) was lost in all hiPSCs. This observation suggests a loop established *de novo* in control but not FSHD cells, reinforcing the hypothesis of a disturbed development in FSHD cells<sup>48,58,59</sup>. In parallel, we tested gene expression by RT-qPCRs (Fig. 7). Our data revealed no significant difference between conditions, except for the *DUX4* gene upregulated in FSHD1 hiPSCs ( $p = 0.0130$ ; Mann Whitney test;  $\alpha = 0.05$ ).

In conclusion, as seen by others, *DUX4* expression was independent of the centromeric 4q35 chromatin organization, but might be under the influence of other chromatin features such as changes in DNA methylation, attachment to the nuclear matrix or telomere length<sup>28,29,37</sup>. Overall, our data suggest that the 4q35 3D organization is modified in differentiated cells (fibroblasts) in pathological conditions, *i.e.*, upon shortening of the *D4Z4* array or in cells carrying a mutation in *SMCHD1*. At the pluripotent state, cells display a domain encompassing the most subtelomeric part of the locus and the *SORBS2* region, which appears to be the most conserved domain among cell types as also revealed by HiC data available in the literature<sup>6,42</sup> (Supplementary Fig. 1). Thus, upon reprogramming,



**Figure 4.** 3D genome organization within the 4q35 locus and gene expression in FSHD cells (fibroblasts). Schematic representation of interactions described by 3D FISH experiments (A) and associated gene expression (B). Gene expression quantified by RT-qPCR in control (n = 3), FSHD1 (n = 5) and FSHD2 (n = 2) cells. Each measure represents the average fold-change expression of six independent assays (biological triplicate in technical RT duplicate) normalized to three housekeeping genes (HKG: *HPRT*, *PPIA* and *GAPDH*;  $\Delta\Delta C_t$  method). Differences compared to the control cells are represented as boxplot with associated statistical significance (Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ) along with means  $\pm$  SEM. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.005$ ; \*\*\*\* $p < 0.001$ .

cells undergo a profound reorganization of the 4q35 chromatin structure with an increased level of methylation in FSHD1 cells<sup>56</sup> but also, as shown here, erasure of long-distance interactions indicating a dynamic organization of the 4q35 locus depending on the number of D4Z4 macrosatellite repeats and its epigenetic status.



Sample	Status	Origin	Age	Sex
1	Control	Quad	77	♂
2	FSHD1	Quad	Fetus	♀
3	Control	Quad	Fetus	♂
4	Control	Quad	Fetus	♂
5	FSHD1	Quad	61	♂
6	FSHD2	Quad	57	♂
7	FSHD2	Quad	45	♂
8	FSHD2	Quad	42	♂
9	FSHD1	Quad	54	♂
10	Control	Quad	52	♂
11	Control	Quad	Fetus	♂
12	FSHD1	Quad	61	♂
13	Control	Quad	69	♂
14	Control	Quad	71	♀
15	FSHD1	Quad	Fetus	♂
16	FSHD1	Quad	Fetus	♀
17	FSHD1	Quad	Fetus	♀

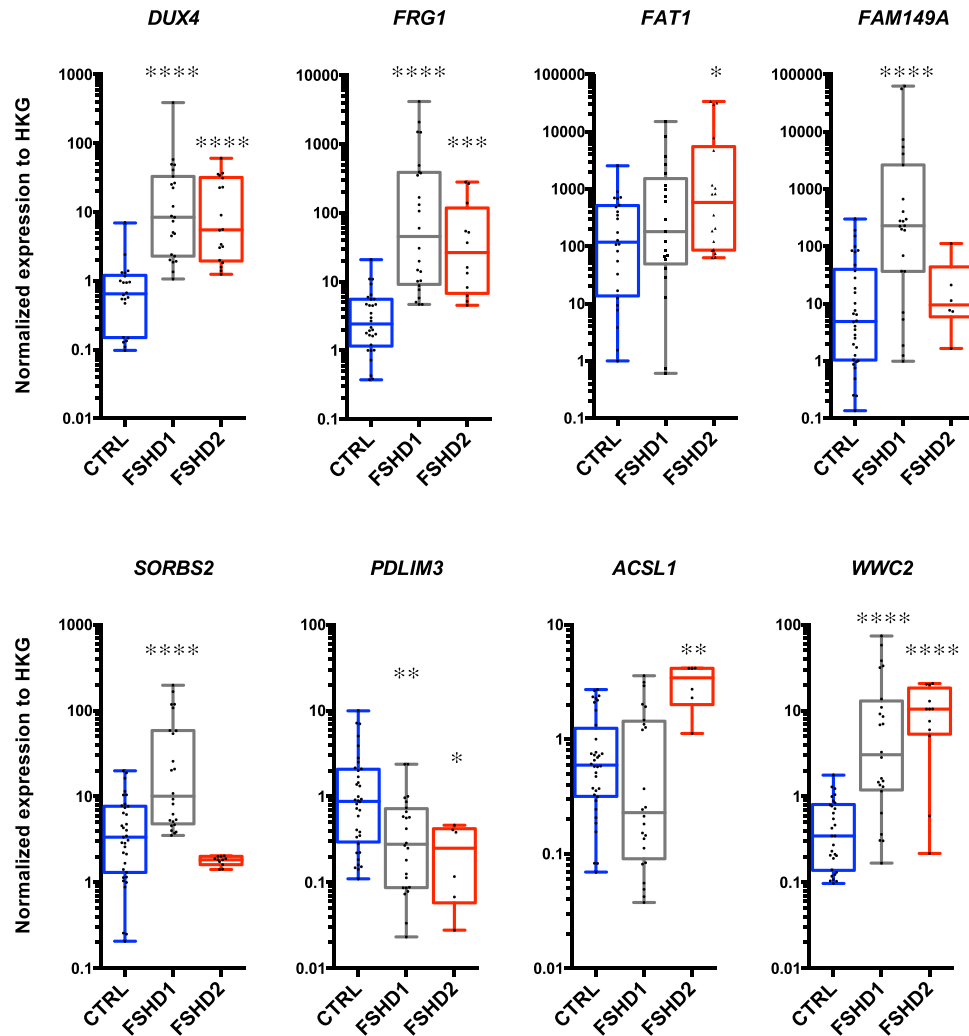
**Table 2.** Summary of muscle biopsies used in the study with their respective status. FSHD individuals were first diagnosed by a clinical examination (except for fetuses obtained after medical abortion) and further confirmed by genetic testing (Southern Blot, DNA combing). Muscle origin, gender and age at biopsy are reported.

## Discussion

In this study, we investigated the organization of the 4q35 locus linked to Facio-Scapulo Humeral Dystrophy in different contexts. Overall, our data show important 3D changes within the 4q35 locus due but not limited to D4Z4 array contraction. We observed differences in chromatin organization between cells from controls and individuals with clinical signs of the disease (e.g., FSHD1, FSHD2). While one interaction was dependent of the D4Z4 array (e.g., D4S139-SORBS2), most interactions are independent of the number of D4Z4 repeats (FSHD2 cases, Figs 2 and 3). This chromatin landscape is profoundly modified upon reprogramming with erasure of the parent-of-origin architectural organization and establishment of a novel pluripotent-specific profile identical between controls and FSHD1 cells (Fig. 6). In somatic cells, modified long-distance interactions correlate with different expression signatures. This observation was validated in muscle biopsies (Fig. 5) and advocates for a conserved topology among cell types and a role for D4Z4 or SMCHD1 in this higher-order organization, with consequences on gene expression. We also observed that expression of *DUX4* was independent of changes occurring in the chromatin organization of more centromeric territories, hence confirming findings of others<sup>28–30,35,36</sup>.

Importantly, our work based on confocal imaging, which provides chromatin analysis at the single cell and single allele level provides a precise representation and frequency of events across multiple samples and for each individual allele within a cell. This technical design limits overrepresentation of discrete events that might bias data interpretation as sometimes observed using 3C techniques<sup>36</sup>. Likewise, previous work based on DNA replication timing experiments revealed high similarities between samples (97.3%) and cell types (>80%) without any difference for the 4q35 locus between samples<sup>60</sup>. Those conflicting observations (interactions vs. no interactions) could arise from difficulties inherent to the analysis of subtelomeric regions in general (i.e., GC-rich, late replication), insufficient depth of sequencing and incomplete assembly of subtelomeric sequences<sup>61–63</sup>. Altogether, these works using sequencing techniques highlight the challenge in analyzing the 4q35 region as illustrated by our recent findings describing complex *cis* rearrangements in a subset of patients<sup>45</sup> and raise further questions on the peculiar organization of the 4q35 loci in FSHD, where one needs to consider the specificity of each allele. By using only samples from well characterized cells with regard to number of repeats (analyzed by Southern blot or DNA combing)<sup>45</sup>, these concerns were carefully considered in the design of our study.

Boundary elements or insulators are critical for the establishment or maintenance of the genome architecture. Boundary elements bound by the CTCF protein concur to the formation of TADs and mediate long-distance interactions. At the 4q35 region, we have described D4Z4 as a genomic element acting as a CTCF and A-type Lamins dependent insulator<sup>13</sup> and able to tether a telomere at the nuclear periphery<sup>12</sup>. We have also identified additional sequences with insulator activity along the 4q35 region, which are not able to tether the 4q35 to the nuclear periphery (Morere and Magdinier, unpublished data). This indicates a key role for D4Z4 in the positioning of this region within the nuclear space but also in the establishment of long distance interactions. In agreement with this hypothesis, we previously showed that in the context of a short D4Z4 array and short telomeres, disruption of 4q35 chromatin loops modulate gene expression but more strikingly splicing of genes such as *SORBS2* or *FAM149A*<sup>16</sup>. Using different cells, our present work confirms the existence of these interactions (Figs 3, 5) with a loop encompassing the *SORBS2* region and the D4Z4 array, conserved in cells with long D4Z4 array. Interestingly, this loop loosened for one out of the two alleles in FSHD1 fibroblasts is reset in hiPSCs after reprogramming and telomerase reactivation. Strikingly, the other 3D structures were completely lost (e.g., D4S139-*FAT1* restricted to FSHD; D4S139-*WWC2* limited to controls) reinforcing the importance of the telomere vicinity in the regulation of this locus<sup>12,16,23</sup>. Interestingly, these loops do not depend on *SMCHD1* gene

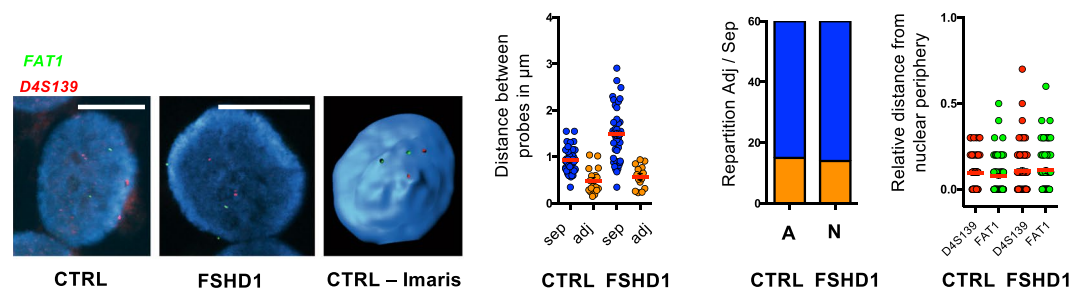
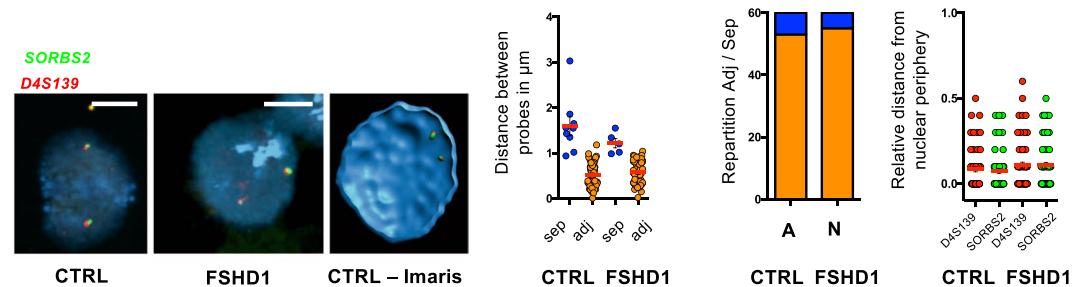
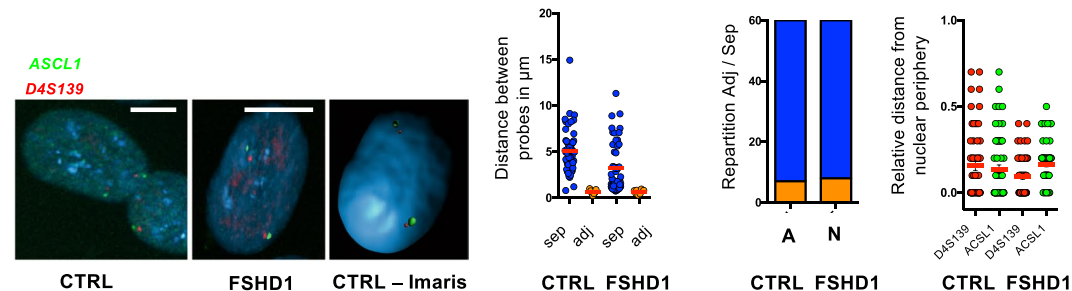
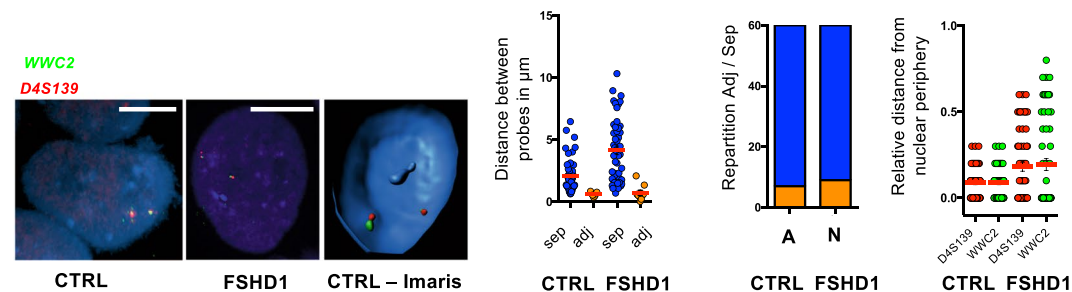


**Figure 5.** Expression of genes of the 4q35 region in controls and FSHD biopsies (muscle). Gene expression quantified by RT-qPCR in control (n = 7, Blue); FSHD1 (n = 7, Grey) and FSHD2 (n = 3, Red) muscle biopsies (*quadriceps femoris*). Each measure represents the average fold-change expression of six independent assays (RT triplicate, technical duplicate) normalized to three housekeeping genes (HKG: *HPRT*, *PPIA* and *GAPDH*;  $\Delta\Delta\text{CT}$  method). Differences from the control group are represented as boxplots with associated statistical significance (Kruskal-Wallis multiple comparison test;  $\alpha = 0.05$ ) along with means  $\pm$  SEM. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.005$ ; \*\*\*\* $p < 0.001$ .

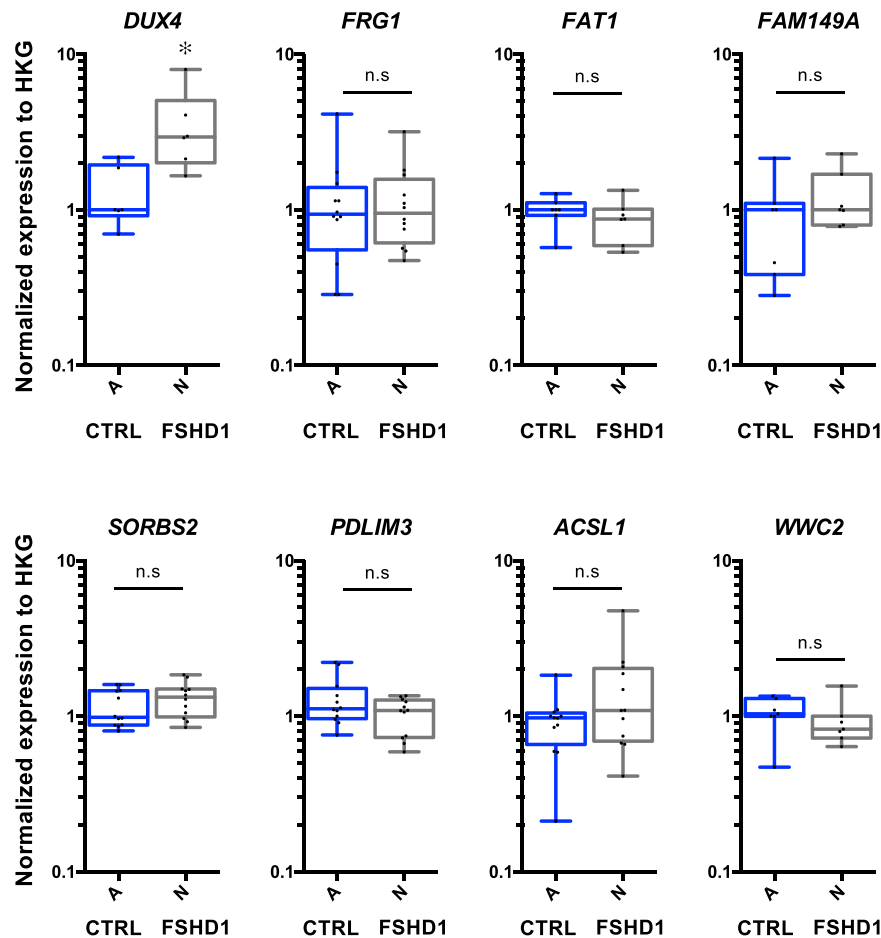
dosage, since long distance interactions are maintained in most FSHD2 cells (Fig. 3B; 76%) carrying a heterozygote truncating mutation. This further argues in favor of a pleiotropic role for this chromatin factor with a role in D4Z4 regulation which likely differs from its topological role in the regulation of *Hox* genes or formation of A and B compartments in X inactivation in the mouse<sup>33,50</sup>.

At the 4q35 locus, the territory identified by 3D FISH is reminiscent of the TAD identified in different cell types<sup>6</sup>, advocating for a pivotal role in the setting of boundaries within the 4q35 region. Further, by investigating hiPSCs which do not express A-type Lamins<sup>55</sup>, we observed that FSHD1 cells did not exhibit the same chromatin organization as somatic cells of origin with a loss in the 3D Chromatin structure present before reprogramming. This intriguing result reveals that tethering to the nuclear periphery of the 4q35 region is not alleviated by absence of A-type Lamins as previously observed using artificially fragmented telomeres<sup>12</sup>. In this telomerase positive context, the short D4Z4 array (*e.g.*, exhibiting an un-masked CTCF site) harbors a peripheral nuclear position but could be untangled with an abutting long telomere, hence masking the CTCF site and modification of long distance interactions. The total topological re-set towards a new organization common between controls and diseased cells suggests a model where differences in chromatin topology would arise during or after differentiation. Hence, one could hypothesize that the FSHD chromatin landscape is not inherited but progressively installed upon differentiation.

Based on our observations, we propose here a complex orchestration of the 4q35 locus folding upon differentiation from a pluripotent state (*i.e.*, hiPSC) with clear differences between control cells and cells from patients affected with FSHD. Combined with previous findings<sup>12,13,16,23</sup>, defining the dynamic 3D organization of this

**A** *D4S139 – FAT1***B** *D4S139 – SORBS2***C** *D4S139 – ACSL1***D** *D4S139 – WWC2*

**Figure 6.** Topology of the 4q35 region is identical in control and FSHD pluripotent stem cells. Three-Dimension DNA FISH assay within the 4q35 locus in human Induced Pluripotent Stem Cells (hiPSC) generated from control and FSHD1 primary fibroblasts (A, N; respectively - see Table 1). 3D FISH was done using two sets of probes corresponding either to the D4S139 region (red) or different loci along the 4q35 chromosome (green, *FAT1*, *SORBS2*, *ACSL1* or *WWC2*; respectively). Representative pictures and 3D reconstruction using IMARIS along with their associated quantifications are presented. We measured the distance of each probe to the nuclear periphery and distance between the probes. For each cell, we evaluated if signals are separated or adjacent (sep, adj; respectively). Scattergram distribution of events and associated Chi-Square test ( $n = 30$  cells per sample, 60 alleles) are shown. Scale is indicated by a white bar ( $=4\mu\text{m}$ ). No statistical differences were found between cells. (A) Localization between the D4S139 region and the *FAT1* gene. (B) Localization between the D4S139 region and the *SORBS2* gene. (C) Localization between the D4S139 region and the *ACSL1* gene. (D) Localization between the D4S139 region and the *WWC2* gene.



**Figure 7.** 4q35 gene expression is not modulated in FSHD hiPSCs compared to controls. Gene expression quantified by RT-qPCR in control FSHD1 hiPSC (A, N; respectively - see Table 1). Each measure represents the average fold-change expression of six independent assays (Biological triplicate in technical RT duplicate) normalized to three housekeeping genes (HKG: *HPRT*, *PPIA* and *GAPDH*;  $\Delta\Delta C_t$  method). Differences relative to control cells are represented as boxplots with associated statistical significance (Mann Whitney;  $\alpha = 0.05$ ) along with Means  $\pm$  SEM. \* $p < 0.05$ .

locus in different genetic contexts will provide further insights into the understanding of its regulation and its impact on the transcriptional signature of 4q35 genes during development and differentiation towards the muscle lineage. Furthermore, our results confirm a role of the D4Z4 macrosatellite in the organization of chromatin in human cells and formation of long-distance loops.

## Materials and Methods

The present study, protocols and experiments were approved, performed and validated following the scientific guidelines provided by the Agence National pour la Recherche (ANR); INSERM (U1251 laboratory) and AFM. Research was approved by the biomedicine agency (PFS13-006).

**Study samples.** All individuals have provided written informed consent for the collection of samples and subsequent analysis for medical research. The study was done in accordance with the Declaration of Helsinki and samples were obtained from the Centre de Ressources Biologiques (CRB) Hôpital de la Timone Enfant, Marseille (CRB AP-HM 0COPE05E001). Controls are randomly selected individuals or patient's relatives, selected in the same age range and sex representation as patients. Controls are neither carrier of any genetic mutation nor affected by any constitutive pathology. Samples are listed in Tables 1 and 2.

**Reprogramming of human fibroblasts and production of hiPSCs.** Human iPSCs were generated as described elsewhere<sup>57,58</sup>. HiPSCs colonies were picked about 4 to 6 weeks after reprogramming based on their ES cell-like morphology. Colonies were grown and expanded in mTeSR<sup>TM</sup> 1 medium (StemCells) on dishes coated with Matrigel<sup>TM</sup> (Corning, cat. No. 354277). The different clones used were fully characterized using classical procedures<sup>56,57</sup>.

**RNA extraction and quality control.** Total RNA was extracted using the RNeasy kit (Qiagen) following manufacturer's instructions. Quality, quantification and sizing of total RNA was evaluated using the RNA 6000 Pico assay (Agilent Technologies Ref. 5067-1513) on an Agilent 2100 Bioanalyzer system.

**Quantitative RT-PCR.** Reverse transcription of 500 ng of total RNA was performed using the Superscript II or IV kit and a mixture of random hexamers and oligo dT following manufacturer's instructions at 42 °C for 50 minutes followed by inactivation at 70 °C for 15 minutes (Life Technologies). Primers were designed using Primer Blast and Primer3 (See list attached below). Real-time PCR amplification was performed on a LightCycler 480 (Roche) using the SYBR green master mix. All PCR were performed using a standardized protocol and data were analyzed with the Lightcycler 480 software version 1.5.0.39 (Roche). Primer efficiency was determined by absolute quantification using a standard curve. For each sample, fold-change was obtained by comparative quantification and normalization to expression of the *HPRT*, *GAPDH* and *PPIA* (*DUX4* and 4q35 genes) housekeeping genes used as standard. Data are expressed as means  $\pm$  SEM.

Gene	Forward Primer	Reverse Primer
<i>HPRT</i>	tgatagatccattcctatgactgtaga	caagacattcttccagttaaagtgtg
<i>GAPDH</i>	agccacatcgctcagacac	gcccatacaccacaaatcc
<i>PPIA</i>	atgctggaccacaacaaat	tctttcacttgcacaaacacc
<i>DUX4</i>	cccaggtaccagcagacc	tccaggagatgtaactctaacca
<i>FRG1</i>	ttgttggaatctggtggaca	gcccacaacaagtccatct
<i>FAT1</i>	atggagggtcgattcacg	cattagatggctctggcg
<i>FAM149A</i>	ccgcccagattcatcactt	gctgaagcggtttcttga
<i>SORBS2</i>	aaccttaagcaagaggaggcgag	ttcccctgtagcctagctcagc
<i>PDLIM3</i>	ggtgctgtggtgaagcgcg	gccctctggggccttctgtc
<i>ASCL1</i>	cctgggattcacttctcagggga	gtgcaagccctgttctgtctgt
<i>WWC2</i>	tgacaatatggcagttgcccca	tcactgtcactccgatttaacctgc

**3D Fish.** Cells were grown on 4-well slides (Millipore) or 4-well PCA slides (hIPSCs; Sarstedt), fixed in 4% paraformaldehyde and treated as described<sup>12</sup> to maintain the 3D structure of the cells. All probes were denatured at 80  $\pm$  1 °C for 5 minutes before hybridization. Nuclei were counterstained with DAPI (Sigma) diluted to 1 mg/ml in PBS and mounted in Vectashield (Vector Laboratories). Probes were either purchased (RainbowFISH, Cytocell) or produced by nick translation using after-mentioned BAC clones (CHORI) as template: RP11-521G19 (D4S139); RP11-33M11 (*FAT1*); RP11-159A22 (*SORBS2*); RP11-374K13 (*ASCL1*) and RP11-451F20 (*WWC2*), respectively.

Images were acquired using a confocal scanning laser system from Zeiss (Germany). A 63x Plan-APOCHROMAT, oil immersion, NA 1.40 objective (Zeiss) was used to record optical sections at intervals of 0.24  $\mu$ m. The pinhole was set the closest to 1 Airy with optical slices in all wavelengths with identical thickness (0.4  $\mu$ m). Generated.lsm files with a voxel size of 0.1  $\mu$ m  $\times$  0.1  $\mu$ m  $\times$  0.24  $\mu$ m were processed using the IMARIS software (Bitplane, AG). After 3D analysis, we compared the mean volume ratio of nuclei and of the probe signals. Values were compared using a non-parametric Kruskal-Wallis test; population of separated or adjacent signals were compared to the control condition using a Chi-square test.

## References

1. Robin, J. D. & Magdinier, F. Physiological and Pathological Aging Affects Chromatin Dynamics, Structure and Function at the Nuclear Edge. *Front Genet* **7**, 153 (2016).
2. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144**, 3646–3658 (2017).
3. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics* **2**, 292–301 (2001).
4. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20**, 2223–2237 (2006).
5. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
6. Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 151 (2018).
7. Jerković, I. *et al.* Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF. *PLOS Genetics* **13**, e1006567 (2017).
8. Stavreva, D. A. *et al.* Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. *Genome Res.* **25**, 845–857 (2015).
9. Robin, J. D. & Magdinier, F. In *Handbook of Epigenetics* 73–92, <https://doi.org/10.1016/B978-0-12-805388-1.00006-7> (Elsevier, 2017).
10. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
11. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
12. Ottaviani, A. *et al.* Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. *EMBO J.* **28**, 2428–2436 (2009).
13. Ottaviani, A., Schluth-Bolard, C., Gilson, E. & Magdinier, F. D4Z4 as a prototype of CTCF and lamins-dependent insulator in human cells. *Nucleus* **1**, 30–36 (2010).

14. Zullo, J. M. *et al.* DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149**, 1474–1487 (2012).
15. Sun, J. H. *et al.* Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* **175**, 224–238.e15 (2018).
16. Robin, J. D. *et al.* SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy. *Genome Res.* **25**, 1781–1790 (2015).
17. Tam, R., Smith, K. P. & Lawrence, J. B. The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J. Cell Biol.* **167**, 269–279 (2004).
18. Masny, P. S. *et al.* Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum. Mol. Genet.* **13**, 1857–1871 (2004).
19. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
20. Ueda, J. *et al.* Heterochromatin dynamics during the differentiation process revealed by the DNA methylation reporter mouse, MethylRO. *Stem Cell Reports* **2**, 910–924 (2014).
21. Gonzalez-Suarez, I., Redwood, A. B. & Gonzalo, S. Loss of A-type lamins and genomic instability. *Cell Cycle* **8**, 3860–3865 (2009).
22. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
23. Arnoult, N. *et al.* Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization. *PLOS Genetics* **6**, e1000920 (2010).
24. Tsumagari, K. *et al.* Epigenetics of a tandem DNA repeat: chromatin DNaseI sensitivity and opposite methylation changes in cancers. *Nucleic Acids Res.* **36**, 2196–2207 (2008).
25. Sarfarazi, M. *et al.* Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: combined analysis of an international consortium. *Am. J. Hum. Genet.* **51**, 396–403 (1992).
26. Wijmenga, C. *et al.* Physical mapping and YAC-cloning connects four genetically distinct 4qter loci (D4S163, D4S139, D4F35S1 and D4F104S1) in the FSHD gene-region. *Hum. Mol. Genet.* **2**, 1667–1672 (1993).
27. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–1653 (2010).
28. Gaillard, M.-C. *et al.* Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology* **83**, 733–742 (2014).
29. Stadler, G. *et al.* Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nature Structural & Molecular Biology* **20**, 671–678 (2013).
30. Jones, T. I. *et al.* Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. *Clin Epigenetics* **7**, 37 (2015).
31. Lemmers, R. J. L. F. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
32. Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
33. Jansz, N. *et al.* SmcHD1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nature Structural & Molecular Biology* **25**, 766–777 (2018).
34. Scionti, I. *et al.* Large-scale population analysis challenges the current criteria for the molecular diagnosis of facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **90**, 628–635 (2012).
35. Bodega, B. *et al.* Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biol.* **7**, 41 (2009).
36. Pirozhkova, I. *et al.* A Functional Role for 4qA/B in the Structural Rearrangement of the 4q35 Region and in the Regulation of FRG1 and ANT1 in Facioscapulohumeral Dystrophy. *PLoS ONE* **3**, e3389 (2008).
37. Petrov, A. *et al.* Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. *Proceedings of the National Academy of Sciences* **103**, 6982–6987 (2006).
38. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
39. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports* **17**, 2042–2059 (2016).
40. Puppo, F. *et al.* Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. *Hum. Mutat.* **36**, 443–453 (2015).
41. Caruso, N. *et al.* Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. *PLOS Genetics* **9**, e1003550 (2013).
42. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **162**, 687–688 (2015).
43. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
44. Meaburn, K. J., Gudla, P. R., Khan, S., Lockett, S. J. & Misteli, T. Disease-specific gene repositioning in breast cancer. *J. Cell Biol.* **187**, 801–812 (2009).
45. Nguyen, K. *et al.* Molecular combing reveals complex 4q35 rearrangements in Facioscapulohumeral dystrophy. *Hum. Mutat.* **38**, 1432–1441 (2017).
46. Nguyen, K. *et al.* Molecular combing reveals allelic combinations in facioscapulohumeral dystrophy. *Ann. Neurol.* **70**, 627–633 (2011).
47. Ghirlando, R. *et al.* Chromatin domains, insulators, and the regulation of gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1819**, 644–651 (2012).
48. Brouqsault, N. *et al.* 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 (2013).
49. Gabellini, D., Green, M. R. & Tupler, R. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110**, 339–348 (2002).
50. Wang, C.-Y., Jégu, T., Chu, H.-P., Oh, H. J. & Lee, J. T. SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* **174**, 406–421.e25 (2018).
51. Strickfaden, H., Zunhammer, A., van Koningsbruggen, S., Köhler, D. & Cremer, T. 4D chromatin dynamics in cycling cells: Theodor Boveri's hypotheses revisited. *Nucleus* **1**, 284–297 (2010).
52. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10**, 1297–1309 (2015).
53. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
54. Gonzales, K. A. U. & Ng, H.-H. Looping around Reprogramming: The Topological Memory of Induced Pluripotency. *Cell Stem Cell* **18**, 557–559 (2016).
55. Constantinescu, D., Gray, H. L., Sammak, P. J., Schatten, G. P. & Csoka, A. B. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells* **24**, 474–474 (2006).

56. Dion, C. *et al.* SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. *Nucleic Acids Res.* **40**, 663 (2019).
57. Badja, C. *et al.* Efficient and cost-effective generation of mature neurons from human induced pluripotent stem cells. *Stem Cells Transl Med* **3**, 1467–1472 (2014).
58. Ferreboeuf, M. *et al.* DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet.* **23**, 171–181 (2014).
59. Banerji, C. R. S. *et al.* PAX7 target genes are globally repressed in facioscapulohumeral muscular dystrophy skeletal muscle. *Nat Commun* **8**, 2152 (2017).
60. Pope, B. D. *et al.* DNA replication timing is maintained genome-wide in primary human myoblasts independent of D4Z4 contraction in FSH muscular dystrophy. *PLoS ONE* **6**, e27413 (2011).
61. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17**, 215 (2016).
62. Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* **14**, 673–678 (2017).
63. Sims, D., Sudbery, I., Iltot, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* **15**, 121–132 (2014).

## Acknowledgements

We are indebted and thank all patients for participating in this study. We acknowledge the Marseille Stem Cell core facility for technical contribution. This study was funded by “Association Française contre les Myopathies” (AFM; MNHDecrypt and TRIM-RD), Agence Nationale de la Recherche (ANR-13-BSV1-0001). CD and JM were recipient of a fellowship from the French Ministry of Education and FSH Society. CL is the recipient of a fellowship from the French Ministry of Education. MCG was the recipient of a fellowship from the FSH Society. CB was the recipient of fellowship from the Algerian Ministry of Education and Fondation pour la Recherche Médicale.

## Author Contributions

Drs. Marie-Cécile Gaillard, Natacha Broucqsault, Camille Dion, Cherif Badja, Julia Morere, Stéphane Roche and miss Camille Laberthonnière set up the experimental conditions, performed the experiments analyzed and organized the data. These authors report no disclosure. Dr. Karine Nguyen provided and processed patients’ samples. Dr. Nguyen reports no disclosure. Dr Jérôme D. Robin: Contributed to the design and coordination of the project, analyzed the data, performed the statistical analysis and wrote the paper. Dr. Robin reports no disclosure. Dr. Frederique Magdinier: designed and coordinated the project, provided funding, performed the statistical analysis, analyzed the data and wrote the paper. Dr. Magdinier reports no disclosure.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-46861-x>.

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019