4q-D4Z4 chromatin architecture regulates the transcription of muscle atrophic genes in FSHD

Alice Cortesi[1–], Matthieu Pesant[1–], Shruti Sinha[1–], Federica Marasca[1], Eleonora Sala[1], Francesco Gregoretti[2], Laura Antonelli[2], Gennaro Oliva[2], Chiara Chiereghin[3,4], Giulia Soldà[3,4], and Beatrice Bodega[1*].

(1) Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi" (INGM), Milan, Italy

(2) CNR Institute for High Performance Computing and Networking (ICAR), Naples, Italy

(3) Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

(4) Humanitas Clinical and Research Center, Rozzano, Milan, Italy

(□) These authors contributed equally to this work

(*) Correspondence should be addressed to B.B. bodega@ingm.org, phone: +39 02 00660302

Running title: 4q-D4Z4 chromatin architecture is impaired in FSHD

Keywords: DNA repeats, 3D genome organization, D4Z4, FSHD

Abstract

Despite increasing insights in genome structure organization, the role of DNA repetitive elements, accounting for more than two thirds of the human genome, remains elusive. Facioscapulohumeral Dystrophy (FSHD) is associated with deletion of D4Z4 repeat array below 11 units at 4q35.2. It is known that the deletion alters chromatin structure *in cis*, leading to genes upregulation. Here we show a genome-wide role of 4q-D4Z4 array in modulating gene expression via 3D nuclear contacts. We have developed an integrated strategy of 4q-D4Z4 specific 4C-seq and chromatin segmentation analyses, showing that 4q-D4Z4 3D interactome and chromatin states of interacting genes are impaired in FSHD1 condition; in particular, genes which have lost the 4q-D4Z4 interaction and with a more active chromatin state are enriched for muscle atrophy transcriptional signature. Expression level of these genes is restored by the interaction with an ectopic 4q-D4Z4 array, suggesting that the repeat directly modulates the transcription of contacted targets.

Of note, the upregulation of atrophic genes is a common feature of several FSHD1 and FSHD2 patients, indicating that we have identified a core set of deregulated genes involved in FSHD pathophysiology.

Introduction

Among primate specific macrosatellites, D4Z4 is a 3.3 Kb unit tandem repeat duplicated on several chromosomes (Bakker et al. 1995; Deidda et al. 1995; Lyle et al. 1995; Bodega et al. 2006; Bodega et al. 2007), and in particular present as a polymorphic array of 11 to 100-150 copies at 4q35.2 (4q-D4Z4 array) in the general population (Hewitt et al. 1994). Reduction of 4q-D4Z4 array copy number below 11 units is associated with Facioscapulohumeral Dystrophy (FSHD, MIM158900; (van Deutekom et al. 1993)), one of the most common myopathies in humans with an overall prevalence of more than 1:10,000 (Sacconi et al. 2015). FSHD is characterized by progressive, often asymmetric, weakness and wasting of facial (facio), shoulder and upper arm (scapulohumeral) muscles (Tawil and Van Der Maarel 2006), where fiber necrosis and degeneration give rise to muscle atrophy (Sacconi et al. 2015).

FSHD is a genetically variable disorder, mainly transmitted as an autosomal dominant trait, on a specific FSHD-permissive haplotype of Chromosome 4q, namely 4qA (Lemmers et al. 2002; Lemmers et al. 2007). This form accounts for approximately 95% of the cases (FSHD1); however, about 5% of the patients display FSHD lacking D4Z4 array contractions (FSHD2). FSHD2 is caused by mutations in *SMCHD1*, a member of the condensin/cohesin chromatin compaction complexes, that binds to the D4Z4 repeat array (Lemmers et al. 2012). While in healthy individuals the 4q-D4Z4 array is characterized by highly methylated DNA, the contracted allele in FSHD1 and both the 4q-D4Z4 alleles in FSHD2 are hypomethylated (van Overveld et al. 2003; de Greef et al. 2009).

The highly heterogeneous FSHD clinical features suggest a strong epigenetic contribution to the pathology (Tawil et al. 1993; Cabianca and Gabellini 2010; Neguembor and Gabellini 2010; Lanzuolo 2012; Daxinger et al. 2015). It is described that the 4q-D4Z4 array is able to engage short- and long-range genomic contacts with several genes *in cis* (Petrov et al. 2006; Bodega et al. 2009; Himeda et al. 2014; Robin et al. 2015), concomitantly to the Polycomb group (PcG) protein binding and histone deacetylation, resulting in an overall chromatin compaction. Instead, in FSHD1 condition such interactions are lost, with the consequent alteration of the chromatin structure at the FSHD locus, leading to a more active chromatin state, which is responsible for the de-repression of

the genes *in cis* (Gabellini et al. 2002; Jiang et al. 2003; Bodega et al. 2009; Zeng et al. 2009; Cabianca et al. 2012). In particular, one of the major player in FSHD pathogenesis is the transcription factor *DUX4*, encoded from the most telomeric D4Z4 repeat (Gabriels et al. 1999; Dixit et al. 2007; Lemmers et al. 2010); DUX4 is normally silenced in somatic cells (Snider et al. 2010), but it has been found overexpressed in FSHD patients' myotubes, leading to the activation of genes associated with RNA metabolism processes, stem cell and germ-line development, MERVL/HERVL retrotransposons (Geng et al. 2012; Young et al. 2013; Rickard et al. 2015; Hendrickson et al. 2017) and resulting in the induction of toxicity and apoptosis of muscle cells (Bosnakovski et al. 2008; Block et al. 2013).

Besides the established role of 4q-D4Z4 array in modulating the transcription of *in cis* genes, whether the repeat could also directly affect chromatin structure and gene expression of other loci via 3D physical contacts has not been investigated yet. Therefore, we have explored the 4q-D4Z4 chromatin architecture and possible alterations in FSHD.

Results

4q-specific D4Z4 interactome is deregulated in FSHD1 patients

Given the high duplication and sequence similarity of 4q35.2 with multiple regions of the genome, in particular with 10q26.3 (Bodega et al. 2006; Bodega et al. 2007), we designed a 4q-specific 4C-seq (circular chromosome conformation capture sequencing) strategy to investigate its interactome. As 4C viewpoint (VP) we used the region nearby a single sequence length polymorphism (SSLP), present shortly upstream (almost 3.5 Kb) of the first D4Z4 repeat on 4q (4qA and 4qB) and 10q arrays (Lemmers et al. 2007); we performed paired-end sequencing, that allowed to retrieve the information of the SSLP variant (Read 1) and the interacting region (Read 2), assigning with high precision the allele origin of the D4Z4 interactome (Supplemental Fig. S1; Supplemental Table S1; see Methods and Supplemental Methods).

With this approach, we probed the 4q-D4Z4 chromatin conformation in human primary muscle cells from two FSHD1 patients (FSHD1) and two healthy individuals (CN) (Supplemental Table S1), that did not differ for myoblast purity and differentiation efficiency (Supplemental Fig. S2). 4C-seq was performed on myoblasts (MB) to highlight differences that could precede any transcriptional effect in differentiated cells. Comparative analyses of 4C-seq samples showed high level of reproducibility and similarity both at the level of donor origin (fragends read count, CN or FSHD1) (Supplemental Fig. S3) and at the level of viewpoint (called interacting regions, 4q vs 10q) (Supplemental Fig. S4A,B). We identified 4q-D4Z4 specific *cis* interactions with *FRG1*, *ZFP42*, *SORBS2* and *FAM149A* genes (Fig. 1A; Supplemental Fig. S4C-E), as already reported (Bodega et al. 2009; Robin et al. 2015), suggesting that our approach is robust in the detection of 4q-specific D4Z4 interactions.

We retrieved 244 and 258 4q-D4Z4 interacting regions for CN and FSHD1 respectively, and in particular, among them, 175 for CN and 181 for FSHD1 were *trans* interactions. Interestingly, 116 regions interacting in CN were specifically lost in FSHD1 cells and the vast majority (101) were in *trans* (Fig. 1A; Supplemental Table S2).

3D multicolor DNA FISH was performed on the same and additional CN and FSHD1 donor MB to validate 4C results, using a probe on a not duplicated region in 4q35.1 (Supplemental Fig. S5A;

(Tam et al. 2004)), and a probe for a positive (C+) or negative (C-) 4q-D4Z4 interacting region (Fig. 1B; Supplemental Fig. S5B). We developed a novel algorithm (NuCLεD, Nuclear Contacts Locator in 3D, see Supplemental Methods) to automatically detect and localize fluorescent spots in 3D reconstructed nuclei. We observed that 4q/C+ interaction had higher frequency of contacts and higher number of positive interacting nuclei compared to 4q/C- interaction (Fig. 1C,D; Supplemental Table S3), with contact frequencies in the range of those estimated for long range interactions (10-20%, (Finn et al. 2019)). Furthermore, 4q and C+ regions shared the same topological nuclear domain in both CN and FSHD1, whereas 4q and C- did not (Supplemental Fig. S5C,D; Supplemental Table S3). Same results were obtained in CN and FSHD1 myotubes (MT) (Supplemental Fig. S5E-G; Supplemental Table S3). Additionally, with our 4C-seq approach we were also able to retrieve 4q allele specific interactomes (4qA and 4qB), as well as 10q-D4Z4 interactome (Supplemental Fig. S6; Supplemental Table S2; see Supplemental Material).

Overall, the 4q-D4Z4-4C-seq strategy allowed to map genome-wide 4q-D4Z4 contacts and to highlight those deregulated in FSHD1.

Genes that show impaired 4q-D4Z4 interactions and activated chromatin state are enriched for atrophic transcriptional signature

In order to identify novel deregulated genes specific for the FSHD condition, we derived chromatin state changes in FSHD1 cells and intersected with 4q-D4Z4 lost interactions in FSHD1, retrieving genes altered both at structural and chromatin levels.

To define the chromatin state, we generated or used available (ENCODE) ChIP-seq datasets for H3K36me3, H3K4me1, H3K27ac, H3K4me3 and H3K27me3 in CN and FSHD1 MB and MT (Supplemental Fig. S7A,B). The quality of ChIP-seq was validated on the same and additional CN and FSHD1 donors MB and MT (Supplemental Fig. S7C-H). Next, we identified 15 chromatin states using ChromHMM (Ernst and Kellis 2012), that were adopted for further downstream analyses (Fig. 2A; Supplemental Fig. S8A,B). Interestingly, chromatin segmentation analysis revealed transitions at enhancers and promoters distinctive for FSHD1 cells (Supplemental Fig. S8C; see Supplemental Material).

In order to identify the genes that specifically switched to activated or repressed chromatin state in FSHD1, we designed the strategy shown in Fig. 2B. Activated genes were defined as those that showed a transition towards a more active state (considering the coverage of the gene body, promoter and enhancer regions) and repressed genes those that showed an opposite change (Fig. 2B; Supplemental Table S4; see Methods). To verify the reliability of this approach, we inspected the expression level of these genes by analyzing RNA-seq datasets performed on the corresponding cell lines and additional publicly-available RNA-seq datasets (Yao et al. 2014). Notably, the activated or repressed chromatin state switches were associated with higher or lower mRNA expression levels, respectively, in FSHD1 compared to CN (Fig. 2C; Supplemental Table S4).

We next sought genes that had lost the interaction with 4q-D4Z4 and also showed chromatin deregulation in FSHD1. We observed that 28% (450/1614) of genes that had lost contact with 4q-D4Z4 in FSHD1 were mainly activated (FSHD1 lost-activated genes, 71%, 319/450), whereas a minority of them (FSHD1 lost-repressed genes, 29%, 131/450) were repressed (Fig. 3A; Supplemental Fig. S9A-C; Supplemental Table S5; see Supplemental Material). Interestingly, only few of these FSHD1 altered genes were regulated by DUX4 (Supplemental Fig. S9D; Supplemental Fig. S10; see Supplemental Material).

We performed Gene Ontology (GO) analyses on the FSHD1 altered genes and found that FSHD1 lost-activated genes were enriched in GO terms linked to protein catabolic processes and in particular with protein ubiquitination/degradation pathways (Fig. 3B; Supplemental Table S5), that are highly relevant to the FSHD-associated atrophic phenotype (Tawil and Van Der Maarel 2006; Sacconi et al. 2015; Statland and Tawil 2016). Similar analysis on 10q-D4Z4 FSHD1 altered genes did not reveal GO terms related to atrophy (Supplemental Fig. S11A,B).

Indeed, we executed Gene Set Enrichment Analysis (GSEA) and further demonstrated that 4q-D4Z4 specific lost-activated genes in FSHD1 were enriched for genes upregulated in the atrophic condition (Fig. 3C; Supplemental Fig. S11C-E; Supplemental Table S5). Of note, the FSHD1 lost-activated genes included in the atrophic dataset displayed higher expression level both in several FSHD1 (Fig. 3D) as well as FSHD2 (Fig. 3E) RNA-seq datasets, revealing that the epigenetic and

transcriptional deregulation of this core set of genes represents a novel transcriptional signature that is common among different FSHD patients.

4q-D4Z4 lost interacting *FBXO32/ATROGIN1* gene is deregulated in FSHD patients at chromatin and transcriptional level

To finely dissect how the 4q-D4Z4 lost interactions could influence the upregulation of muscle atrophy genes in FSHD1 condition, we further investigated the regulation of *FBXO32* (*ATROGIN1*), that is one of the top-enriched genes identified by GSEA, and also one of the major player in different atrophy related conditions, in human and mouse (Gomes et al. 2001; Lecker et al. 2004; Sandri et al. 2004; Bodine and Baehr 2014). We verified by 3D multicolor DNA FISH the loss of *FBXO32*/4q-D4Z4 interaction on several FSHD1 donors compared to CN, extending our analysis also to FSHD2 (Fig. 4A; Supplemental Fig. S12A,B), and observed a decrease in *FBXO32*/4q-D4Z4 interaction frequency in FSHD1, but also in FSHD2 myoblasts (Fig. 4B,C; Supplemental Table S3). Similar results were obtained in FSHD1 myotubes, although with a smaller difference (Supplemental Fig. S12C-E; Supplemental Table S3).

FBXO32 belongs to the category of activated genes in FSHD1, with the appearance of primed enhancers specifically in FSHD1 condition (Fig. 2B; Supplemental Fig. S13A; Supplemental Table S4). Therefore, we further investigated whether the *FBXO32* locus could display distinct chromatin loops at the level of enhancers-promoter in FSHD1. We performed 4C-seq (Fig. 4D; Supplemental Fig. S13B-D; Supplemental Table S1 and S2), showing interaction peaks between enhancers-promoter with higher normalized 4C reads coverage in FSHD1 (Fig. 4E;). These results were further corroborated in 3C experiments (Fig. 4E), suggesting a strengthening of enhancers-promoter contacts at *FBXO32* locus in FSHD1 cells.

In line with this observation, the binding of RNA Pol II at *FBXO32* promoter and an intragenic region was increased in FSHD1 myoblasts respect to CN (Fig. 4F; Supplemental Fig. S14A) and the *FBXO32* expression was upregulated in several FSHD1 donor muscle cells during differentiation, a trend that is also observed in FSHD2 (Fig. 4G). Finally, the *FBOX32* expression is not dependent by DUX4, as it is not affected by DUX4 overexpression (Supplemental Fig. S14B-D;

see Supplemental Material), and ChIP-seq peaks (Geng et al. 2012) are absent in the *FBXO32* gene region (Supplemental Fig. S14E).

Ectopic 4q-D4Z4 array restores the expression of FSHD1 lost interacting genes

To further investigate whether 4q-D4Z4 array could directly modulate the expression of interacting genes, we transfected CN and FSHD1 myoblasts with a BAC containing at least 15 D4Z4 repeat units (B Bodega, unpublished) (BAC 4q-D4Z4n from 4q35.2 region) in parallel with a control BAC (Ctrl BAC, unrelated and not interacting region); transfection efficiency was comparable among the BACs and ranging around 45% (Supplemental Fig. S14F-H; see Supplemental Methods).

We observed that specifically the ectopic 4q-D4Z4 array was in close spatial proximity to the endogenous 4q region, in 70% of the analyzed nuclei (Fig. 5A,B) and interacted with *FBXO32* with a frequency similar as that of the endogenous locus (roughly 20% of analyzed nuclei, Fig. 5C), indicating that the 4q-D4Z4 BAC occupies the same nuclear topological domain of the endogenous 4q region. We then assessed the effect of 4q-D4Z4 BAC transfection on the expression levels of a subset of genes that had lost 4q-D4Z4 interactions in FSHD1. We observed that the transcription of FSHD1 lost-activated genes was reduced (*FBXO32*, *TRIB3* and *ZNF555*; Fig. 5D), whereas a lost-repressed gene was upregulated (*LZTS3*; Fig. 5E) and no effect was detected for not interacting genes (*FOXO3 and MYOG*; Fig. 5F).

Collectively, these results demonstrate that the 4q-D4Z4 array directly modulates the transcription of its interacting targets, suggesting a simultaneous fine-tuning of genes that occupy the same topological domain.

Discussion

Here, we sought to identify the mechanisms by which the contraction of the tandem repeat D4Z4 on Chromosome 4 contributes to FSHD pathogenesis, using an integrated multi-omics approach (4C-seq, ChIP-seq and RNA-seq).

We found that 4q-D4Z4 interactome is altered in FSHD1 patients. In particular, normal 4q-D4Z4 array contacts several regions in a peripheral nuclear domain, controlling their transcription (Fig. 6A). In FSHD1 patients, the shortened and hypomethylated 4q-D4Z4 array causes an impairment of the chromatin conformation, which results in the loss of contacts with atrophic genes, with their consequent chromatin structure alteration and transcriptional upregulation (Fig. 6B). In this regard, it is already demonstrated that chromatin topological structures predominantly consist of simultaneous multiplex chromatin interactions with high heterogeneity between individual cells (Jiang et al. 2016; Zheng et al. 2019). Indeed, we show that an ectopic wild type 4q-D4Z4 array has the ability to get in close spatial proximity to the endogenous locus, resulting in the restoration of the expression of multiple targets, opening the possibility for further mechanistic studies on the dynamics of 3D interactions.

We propose that the genetic deletion of 4q-D4Z4 array in FSHD1 patients leads to a rewired interactome that may represent an additional component of FSHD pathophysiology.

Since we discovered that the subset of genes losing contact with the 4q-D4Z4 array in FSHD1 mainly show chromatin state switches towards activation, we hypothesize that this might be consistent with a broader derepression occurring at the 4q-D4Z4 array, such as lesser PRC1/2 recruitment together or not with an enhanced activity of Trithorax complex, as already demonstrated *in cis* (Cabianca et al. 2012). Of note, SMCHD1 protein, mutated in FSHD2 patients (Lemmers et al. 2012), is now better characterized and involved in higher order chromatin organization of the inactive X Chromosome (Jansz et al. 2018; Wang et al. 2018). We could hypothesize that this architectural protein could have a central role in regulating 4q-D4Z4 interactions and that its altered function in FSHD1 (due to the contraction and hypomethylation of the array) and FSHD2 patients (due to its mutation) could explain the common atrophic signature. Indeed, FSHD-associated atrophy is one of the main signs of the disease (Tawil and Van Der

Maarel 2006; Lanzuolo 2012; Sacconi et al. 2015), for which a direct link with the genetic defect remained elusive till now.

Importantly, we identified a core set of impaired atrophic genes, which is aberrantly transcribed in FSHD1 muscle cells used in this study and in other FSHD1 cells (Yao et al. 2014). Furthermore, they are also deregulated in FSHD2 muscle cells, indicating that the atrophic signature is a common trait in FSHD pathology. We further investigated *FBXO32* gene regulation, which was one of the top ranking; although it was already described overexpressed in muscle biopsies derived from FSHD1 fetuses and adults (Broucqsault et al. 2013), here we linked its transcriptional deregulation to the reduction in *FBXO32*/4q-D4Z4 interaction. This was predominantly observed in FSHD myoblasts compared to myotubes, in line with previous reports that changes in 3D structure precedes changes in gene expression (Hug et al. 2017; Krijger and de Laat 2017; Cheutin and Cavalli 2018) and already demonstrated also for *FRG1* gene (Bodega et al., 2009).

DNA repetitive elements are involved in a plethora of regulatory mechanisms, such as nuclear structure organization and spatiotemporal gene expression regulation (Gregory 2005; de Laat and Duboule 2013; Bodega and Orlando 2014). Additionally, recent studies have highlighted the contribution of satellite repeats in shaping 3D-genome folding and function, as evidenced for pericentromeric satellites (Politz et al. 2013; Wijchers et al. 2015) and DXZ4 macrosatellite (Giacalone et al. 1992; Rao et al. 2014; Deng et al. 2015; Darrow et al. 2016; Giorgetti et al. 2016). Our study is the first demonstration of a role of DNA repetitive elements in the alteration of genomic architecture in the context of a human genetic disease. It further corroborates the concept that perturbations of the 3D-genome structure are involved in various diseases (Krijger and de Laat 2016; Lupianez et al. 2016), such as cancers (Corces and Corces 2016; Rivera-Reyes et al. 2016; Achinger-Kawecka and Clark 2017) and developmental defects (Woltering et al. 2014; Lupianez et al. 2015; Woltering and Duboule 2015).

Our work highlights a novel role of DNA repeats in orchestrating gene transcription by shaping 3D genomic and chromatin architecture. We propose that perturbation of this DNA repeat-mediated regulatory network may be important in other complex genetic and epigenetic diseases.

Methods

Cell cultures

Although it was not always possible to ascertain the status of the muscle origin used in this study (see Supplemental Table S1, sheet "Cell line information"), the majority of the cells used derived from quadriceps, which in general was asymptomatic. Human primary myoblast cell lines from healthy donors (CN), patients affected by FSHD1 or FSHD2 were obtained from the Telethon BioBank of the C. Besta Neurological Institute, Milan, Italy and the Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA; whereas human immortalized myoblast cell lines from healthy donors and FSHD1 patients were obtained from the University of Massachusetts Medical School Wellstone center for FSH Muscular Dystrophy Research, Wellstone Program & Dept. of Cell & Developmental Biology, Worcester, MA USA. Details of all cell lines are reported in Supplemental Table S1; details on media preparation and FACS analysis for Desmin staining are provided in Supplemental Methods.

4C-seq assay

The 4C assay was performed as previously described (Splinter et al. 2012) with minor modifications. A paired-end 4q-D4Z4-specific 4C-sequencing strategy was developed, where one 4C primer was designed to read the single sequence length polymorphism (SSLP) sequences located shortly upstream (almost 3.5 Kb) of the first D4Z4 repeat on 4q or 10q-D4Z4 arrays (Lemmers et al. 2007) and the second 4C primer reads into the captured sequence ligated to the 'bait' fragment. (Supplemental Fig. S1; Supplemental Table S1). Two donor muscle cell lines of CN (CN-3, CN-4) and FSHD1 (FSHD1-3, FSHD1-4) human primary myoblasts (3.5×10^6 per sample) nuclei were processed. Five biological replicates (start to finish experiments) for each cell line were performed (Supplemental Fig. 3A). For *FBXO32* 4C-seq, we designed specific 4C primers as indicated in Supplemental Table S1. Two donor muscle cell lines of CN (CN-3, CN-4) and FSHD1 (FSHD1-3, FSHD1-4) human primary myoblasts (3.5×10^6 per sample) nuclei were processed. From one to two biological replicates (start to finish experiments) for each cell line were performed. Hind III and Dpn II were used for enzymatic digestions. 4C samples were amplified using the bait

and the SSLP specific primers. 4C sequencing libraries were prepared with 4C templates using the NEBNext Ultra DNA Library Prep Kit for Illumina, according to the manufacturer's protocol, cleaned with Agencourt AMPure XP PCR Purification and sequenced on the Illumina NextSeq 500. For more details, see Supplemental Methods.

4C-Seq analysis

The paired-end 4q-D4Z4-specific 4C-seq reads were de-multiplexed based on the 4C bait reading primer that included the restriction site sequence. All reads were then trimmed and read pairs belonging to 4q-D4Z4, 10q-D4Z4 and 4q alleles were identified using SSLP reading mate (Read 1) (Supplemental Fig. S1; Supplemental Table S1) where no mismatch for the genotype sequence was allowed. Read pairs reading into the captured sequences ligated to the bait (Read 2) from the biological replicates of each donor muscle cell line were then pooled and mapped with Bowtie 2 (Langmead and Salzberg 2012). To find chromosome-wide interacting domains, 4C-ker (Raviram et al. 2016) was used. Reproducibility between donor muscle cell lines and quality of sequencing were assessed using Pearson correlation and cis/overall ratio (see Supplemental Methods). High frequency interactions for each viewpoint were intersected using BEDTools v2.2.4 (Quinlan and Hall 2010) and overlapping regions between donor muscle cell lines (CN-3 vs CN-4 and FSHD1-3 vs FSHD1-4) after removing overhangs were considered as high-confidence interacting domains. The interacting genes were defined as those that fall within the coordinates of these domains. Comparative analyses were performed between the 4q-D4Z4 alleles interactomes and also between the 4q and 10q-D4Z4 interactomes (see Supplemental Methods).

The paired-end *FBXO32* 4C-seq reads were demultiplexed based on the 4C bait reading primer that includes the restriction site sequence. All reads were trimmed and reads from the biological replicates of each donor muscle cell line were pooled and then mapped with Bowtie 2 (Langmead and Salzberg 2012). Reproducibility between donor muscle cell lines and quality of sequencing were assessed using Pearson correlation and cis/overall ratio. *Cis*-interacting domains were identified using 4C-ker (Raviram et al. 2016) and high-confidence interacting domains were

selected. Full lists of interactions are available in Supplemental Table S2. For more details, see Supplemental Methods.

ChIP-seq and ChIP-qPCR experiments

ChIP experiments were performed as previously described (Bodega et al. 2017) with minor modifications. The same donor muscle cell lines used for 4C-seq analysis of CN (CN-3, CN-4) and FSHD1 (FSHD1-3, FSHD1-4) human primary myoblasts and myotubes day 4 (3.5×10^6 per sample) were processed for ChIP-seq analysis. For ChIP-seq and ChIP-qPCR, chromatin was immunoprecipitated with anti-H3K36me3 (ab9050, Abcam), anti-H3K4me1 (07- 436, Millipore), anti-H3K27ac (07-360, Millipore), anti-H3K4me3 (07-473, Millipore) and anti-H3K27me3 (07-449, Millipore); anti-RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody [4H8] (ab5408, Abcam). ChIP sequencing libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina, according to the manufacturer's protocol, cleaned with Agencourt AMPure XP PCR Purification and sequenced on the Illumina NextSeq 500 or Hiseq 2000. For ChIP-qPCR experiments, qRT-PCR analysis was performed on a StepOnePlus Real-Time PCR System, using power SYBR Green q-PCR master mix. The relative enrichment obtained by using all the antibodies was quantified after normalization for input chromatin. Primers used are reported in Supplemental Table S6. For more details, see Supplemental Methods.

ChIP-seq analysis

We generated ChIP-seq datasets for CN myoblasts (MB) and myotubes (MT) day 4 for the following histone marks: H3K36me3, H3K4me3 and H3K27me3. H3K36me3, H3K4me1, H3K27ac, H3K4me3 and H3K27me3 datasets were generated for FSHD1 myoblasts and myotubes day 4. The following already published ChIP-seq datasets from ENCODE were used: H3K4me1 of human skeletal myoblasts (ENCSR000ANI), H3K27ac of human skeletal myoblasts (ENCSR000ANF), H3K4me1 of human skeletal myotubes (ENCSR000ANX) and H3K27ac of human skeletal myotubes (ENCSR000ANV). Reads were mapped with Bowtie 2 (Langmead and Salzberg 2012) on quality-checked (FastQC v0.11.2) and trimmed reads (trimmomatic v0.32; (Bolger et al. 2014)).

For visualization of ChIP-seq tracks of independent samples, reads were normalized using bins per million mapped reads (BPM), same as TPM in RNA-seq, and to ensure fair comparison between all datasets, were further normalized to the respective input to produce coverage files reporting the $\log_2$ ratio of normalized read number between samples and inputs using bamCompare module. Quality and reproducibility assessment were done using deepTools2 package. For more details, see Supplemental Methods. Details on DUX4 ChIP-seq analysis is provided in Supplemental Methods.

Chromatin state analysis

We used ChromHMM (Ernst and Kellis 2012) with default parameters to derive genome-wide chromatin states maps of CN and FSHD1 myoblasts and myotubes. We used the 5 histone marks H3K36me3, H3K4me1, H3K27ac, H3K4me3 and H3K27me3, as well as the respective input files, and binarized the data with BinarizeBed. We chose 15 states as the optimal number according to the maximal informative annotated genomic features and minimal redundancy. Subsequent functional annotations were attributed to each state choosing names and a color code for visualization according to the Roadmap Epigenomics Consortium nomenclature (Roadmap Epigenomics Consortium et al. 2015). Total number of derived chromatin features was similar between the samples (CN MB: 503,507; FSHD1 MB: 565,653; CN MT: 609,058; FSHD1 MT: 581,946). We performed overlap enrichment of the 15 chromatin states with known genome organization features (Supplementary Fig. S8A,B) and intersected chromatin states and genes bodies retrieved from GENCODE version 19 using BEDTools v2.2.4 (Quinlan and Hall 2010). Calculations of pairwise Jaccard were performed with BEDTools.

Chromatin state switches analysis

In order to define whether CN and FSHD1 cells showed differences at gene chromatin state level, we took CN data as reference to search for specific switches in FSHD1. We intersected chromatin states retrieved from gene bodies in CN MB, CN MT, FSHD1 MB and FSHD1 MT. We postulated that a given state in a particular condition (CN MB or MT) should intersect another state in the

other condition (FSHD1 MB or MT) in a reciprocal manner. We thus performed BEDTools intersect using –f .60 –r thus requiring that at least 60% of a state in CN MB or MT recovered a state in FSHD1 MB or MT in a reciprocal manner. In this way, identical states in the CN versus FSHD1 comparison (conserved states in the CN/FSHD1 comparison) as well as different states in the CN versus FSHD1 comparison (switching states in the CN/FSHD1 comparison) were retrieved. We focused on chromatin state switches between conditions (Supplemental Table S4) and added directionality to the chromatin state switches (that we chose to be active or repressive switches). We grouped the states into 3 main categories: promoters, enhancers and enhancer priming. The states involved in each group, as well as the definition of the directional switches they are involved in are summarized in Fig. 2B. For each gene, we also summarized all chromatin states expressed as a percentage of coverage across the gene body and defined the state with the highest coverage as being the "major state" for a given gene. We grouped those major states into the 2 categories of active and repressed as indicated in Fig. 2B. To obtain the genes showing directional switches, genes activated should display one of the following features: i) major state transition from repressed to active ii) major active state with at least one additional chromatin state switch towards activation. On the contrary, genes repressed should display either i) transition from an active to a repressed major state ii) major repressive state with at least one additional chromatin state switch towards repression as defined in Fig. 2B.

RNA-seq assay and data analysis

RNA-seq studies were performed on the same donor muscle cell lines used for 4C-seq and ChIP-seq analyses of CN (CN-3, CN-4) and FSHD1 (FSHD1-3, FSHD1-4) human primary myoblasts and myotubes day 4. Briefly, total RNA was isolated using the miRNA Tissue kit on an automated Maxwell RSC extractor, following the manufacturer's instructions. RNA integrity was assessed on TapeStation. Subsequently, RNA for each donor muscle cell line was used to generate single-end 75-bp sequencing libraries with the TruSeq Stranded mRNA Library Prep Kit, according to the manufacturer's protocol. Sequencing was performed on a NextSeq500.

In addition, published RNA-seq dataset GSE56787 (Yao et al. 2014) consisting of human primary healthy, as well as FSHD myoblasts/myotubes obtained from the University of Rochester bio-repository (http://www.urmc.rochester.edu/fields-center) were also analyzed.

All fastq files were analyzed with FastQC v0.11.2. Adapters were removed and trimming was performed with Trimmomatic with standard parameters. Reads mapping to the reference genome GRCh37/hg19 was performed with STAR 2.3.0e (Dobin et al. 2013). The reference annotation used was GENCODEv19 and normalized FPKM (fragments per kilobase of transcript per million mapped reads) values were obtained with Cuffdiff (Trapnell et al. 2013). Normalized FPKM were $\log_2$ transformed and a value of 1 was added to all FPKM values to finally obtain $\log_2$ (1+ FPKM) values used in downstream analyses (Supplemental Table S4; Supplemental Table S5). For more details, see Supplemental Methods.

Gene Ontology analysis

Gene Ontology analysis was performed on protein-coding genes and retrieved from different analysis with the Cytoscape v3.2.0 (Shannon et al. 2003) plug-in ClueGO v2.1.5 (Bindea et al. 2009). Statistically enriched Biological Processes (updated on 04/18/2016) were functionally grouped according to their k-score, and the most significant GO term of each group was used as summarizing GO term for the group. Full lists of GO terms and associated genes are available in Supplemental Table S5.

Gene Set Enrichment Analysis (GSEA)

GSEA was performed as described in (Subramanian et al. 2005). The gene set was represented by the 319 lost-activated genes or by the 131 lost-repressed genes (genes from Fig. 3A; Supplemental Table S5). We tested if those genes were significantly enriched in a gene expression dataset associated with a skeletal muscle atrophic condition (disuse muscle atrophy, GSE21496; Supplemental Table S5; (Reich et al. 2010)). We also tested the association of our gene sets with a gene expression dataset from skeletal muscle hypertrophy (GSE12474; Supplemental Table S5; (Goto et al. 2011)). GSEA was performed on those datasets with the

ranking metric Signal2noise with 1,000 phenotype permutations for statistical assessment of enrichment.

Three-dimensional multicolor DNA FISH

To produce probes for 3D multicolor DNA FISH we used the following BAC DNA clones (BACPAC Resources Program, CHORI): CH16-77M12 (D4Z4, containing at least 15 units of D4Z4 repeat, B Bodega, unpublished and (Cabianca et al. 2012)), RP11-279K24 (4q), RP11-846C19 (C-), RP11-115K4 (C+), RP11-288G11 (10q26.3) and RP11-174I12 (*FBXO32*). Probes used for 3D multicolor DNA FISH in transfection experiments presented in Fig. 5A-C were produced from PCR designed on the pTARBAC6 backbone of the transfected CH16-291A23 BAC (BAC 4q-D4Z4n), on the pBACe3.6 backbone of the transfected RP11-2A16 BAC (Ctrl BAC), on a 35 Kb genomic region of 4q35.1 (4q) and on a 35 Kb genomic region of *FBXO32*. Primers used are reported in Supplemental Table S6. 1-3 µg of BAC DNA or pooled PCR products were labelled with bio-dUTP, dig-dUTP or cy3-dUTP through nick translation. The 3D multicolor DNA FISH assay was performed accordingly to (Cremer et al. 2008) with minor adaptations. One to three donor muscle cell lines of CN and FSHD human primary myoblasts or myotubes day 4 were processed for each experiment. An Eclipse Ti-E (Nikon Instruments) microscope was used to scan the nuclei, with an axial distance between 0.2-0.25 µm consecutive sections. In order to automatically analyze 3D multicolor DNA FISH in fluorescence cell image z-stacks, we developed a tool in MATLAB. The tool, that we named NuCLεD (Nuclear Contacts Locator in 3D), is capable to automatically detect and localize fluorescent 3D spots in cell image stacks. Measurements retrieved are shown in Supplemental Table S3. Details on 3D multicolor DNA FISH protocol and NuCLεD algorithm description are provided in Supplemental Methods.

Chromatin conformation capture (3C)

The 3C assay was performed as previously described (Cortesi and Bodega 2016) with minor adaptations. Two donor muscle cell lines of CN (CN-3, CN-4) and FSHD1 (FSHD1-3, FSHD1-4) human primary myoblasts (3.5 x 10^6 per sample) nuclei were processed. One to two biological

457 replicates for each cell line was done. Digestion was performed using Hind III. A reference
458 template was generated by digesting, mixing and ligating a BAC covering the genomic region of
459 interest (*FBXO32* region, RP11-174I12). 3C templates and the reference template were used to
460 perform PCR analysis with DreamTaq DNA Polymerase using primers designed around the Hind
461 III restriction sites present at *FBXO32* region and, as bait primer, the same of *FBXO32*-4C
462 (Supplemental Table S6) on a Veriti 96-Well Thermal Cycler. The PCR products were
463 densitometrically quantified using the ImageJ software. Data are presented as the ratio of
464 amplification obtained with 3C templates in respect to the reference template. For more details,
465 see Supplemental Methods.
466
467 **BAC transfection**
468 BAC transfections were performed accordingly to (Montigny et al. 2003) with minor adaptations.
469 CN and FSHD1 human primary and immortalized myoblasts were plated. The following day, BAC
470 DNA (RP11-2A16, as control BAC, representative of an unrelated and not interacting genomic
471 region, Chr 17q21.33, and CH16-291A23, containing at least 15 units of D4Z4 repeat, B Bodega,
472 unpublished and (Cabianca et al. 2012)) were diluted in Opti-MEM with the addition of P3000
473 Reagent. Lipofectamine 3000 Reagent were diluted in Opti-MEM. After 5 min BAC DNA (plus
474 P3000) and Lipofectamine preparations were gently mixed and incubated for 20 min at room
475 temperature. Transfection complexes were then added to the cells and incubated at 37 °C for 48 h.
476 The primer pairs used for PCR or qRT-PCR amplifications are shown in Supplemental Table S6.
477 For more details on transfection efficiency and DNA extraction, see Supplemental Methods.
478
479 **Statistics and Bioinformatics**
480 To determine the significance between two groups, we used Wilcoxon matched-pairs signed rank
481 test, Student's *t*-test or Fisher's exact test, as reported in Figure legends; exact *P* values and exact
482 types of tests used are specified in Figure legends. For correlation analysis, we used Pearson
483 correlation; the exact values are specified in the figures. Multiple comparisons were done by two-

484 way ANOVA followed by Bonferroni post-test correction; exact *P* values and types of tests used
485 are specified in Figure legends.
486 For all statistical tests, the 0.05 level of confidence was accepted for statistical significance;
487 statistical significance is denoted by asterisks in figures, where * represent p-value <0.05, **
488 represents <0.01, *** represents <0.001 and **** represents <0.0001.
489 All reads were assessed for quality using FastQC and processed using Trimmomatic. They were
490 aligned to the human genome (hg19) using either Bowtie 2 (Langmead and Salzberg 2012) or
491 STAR 2.3.0e (Dobin et al. 2013). Aligning to GRCh38 is expected to provide similar results, as only
492 a small number of bases change genome-wide with the major difference between the releases is in
493 centromere assembly (Guo et al. 2017), which is not the focus of our study.
494

495 **Data access**

496 Circular chromosome conformation capture and sequencing data (4C-seq), chromatin
497 immunoprecipitation and RNA sequencing data (ChIP-seq and RNA-seq) for the human samples
498 have been submitted to the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/)
499 under the accession number SRP117155.
500

501 **Acknowledgments**

502 We thank C. Lanzuolo, D. Gabellini, V. Ranzani, M. R. Matarazzo and E. Battaglioli for their
503 stimulating discussions and constructive criticisms of this manuscript. We gratefully acknowledge
504 M. Mora from the Telethon BioBank of the C. Besta Neurological Institute, R. Tawil from the Fields
505 Center for FSHD of the Rochester Medical Center Dept. of Neurology and J. Chen from the
506 University of Massachusetts Medical School Wellstone center for FSH Muscular Dystrophy
507 Research, Wellstone Program & Dept. of Cell & Developmental Biology for providing the human
508 myoblasts. The authors acknowledge the scientific and technical assistance of the INGM Imaging
509 Facility (Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi" (INGM), Milan,
510 Italy), in particular C. Cordiglieri, for assistance during 3D multicolor DNA FISH images acquisition.

pCMV-Mock and pCMV-DUX4 plasmids was a kind gift of D. Gabellini. This work has been supported by the following grants to B.B.: EPIGEN Italian flagship program, Association Française contre les Myopathies (AFM-Telethon, grant nr 18754) and Giovani Ricercatori, Italian Ministry of Health (GR-2011-02349383).

Author contributions

A.C. and M.P. and S.S, designed and performed experiments, analyzed the data and wrote the manuscript. F.M and E.S. performed experiments. F.G., L.A., G.O. developed the new algorithm for 3D multicolor DNA FISH image analysis. C.C. and G.S. performed RNA-seq library preparation, NGS data processing and sequencing. B.B. conceived this study, designed experiments, analyzed the data and wrote the manuscript.

Disclosure declaration

The authors declare no competing financial interests.

Figure Legends

Figure 1. 4q-D4Z4 specific 4C-seq highlights FSHD1 impaired interactome

(*A*) (left) Circos plot depicting *cis* and *trans* 4q-D4Z4 interactions in CN (CN-3, CN-4) myoblasts called by 4C-ker. Common interactions with FSHD1 (FSHD1-3, FSHD1-4) myoblasts are in grey, whereas interactions specifically lost in FSHD1 are highlighted in light blue. (right) Zoomed-in circos plot representation of common (grey) and FSHD1 lost (light blue) *cis* interactions on Chr 4. Gene are indicated for a region extending up to 4 Mb from the VP. Black triangles in circos plots depict the VP localization. (*B*) Representative nuclei of 3D multicolor DNA FISH using probes mapping to 4q35.1 region (4q, green), a 4q-D4Z4 positive interacting region (8q24.3, C+, red) and a 4q-D4Z4 not interacting region (3q11.2, C-, magenta) and in CN (CN-1, CN-2, CN-3, CN-4) and FSHD1 (FSHD1-1, FSHD1-2, FSHD1-3, FSHD1-4) myoblasts. Nuclei are counterstained with DAPI (blue). All images at 63X magnification. Scale bar=5 µm. (*C*) Cumulative frequency distributions of distances (below 1.5 µm) between 4q and C+ and between 4q and C- in CN (dark and light grey; left) and FSHD1 (dark and light blue; right) myoblasts. n=1,296 (CN 4q/C+), 1,708 (CN 4q/C-), 884 (FSHD1 4q/C+) and 1,128 (FSHD1 4q/C-). *P* values were calculated by unpaired one-tailed *t*-test with confidence interval of 99%. Asterisks represent statistical *P* values; for 4q/C+ vs 4q/C- in CN and FSHD1 $p<0.0001$. (*D*) Percentage of nuclei positive for the interactions (under the cut-off of 1.5 µm). n= 427 (CN 4q/C-), 324 (CN 4q/C+), 282 (FSHD1 4q/C-) and 221 (FSHD1 4q/C+). *P* values were calculated by fisher's exact one-sided test with confidence interval of 99%. Asterisks represent statistical *P* values; for 4q/C- vs. 4q/C+ in CN $p<0.0001$; for 4q/C- vs. 4q/C+ in FSHD1 $p=0.0046$.

Figure 2. Chromatin segmentation analysis revealed chromatin state switches consistent with transcriptional changes in FSHD1 muscle cells

(*A*) ChromHMM 15-state model obtained with ChIP-seq datasets for H3K36me3, H3K4me1, H3K27ac, H3K4me3 and H3K27me3. Heatmaps display histone marks emission probabilities and transition probabilities between chromatin states. (*B*) Schematic representation of the strategy used to assign genes as activated or repressed in FSHD1. (*C*) Expression levels from RNA-seq

datasets for FSHD1 activated and repressed genes in MB (left) and MT (right), in CN (CN-3, CN-4 and Yao's datasets C20, C21, C22) and FSHD1 (FSHD1-3, FSHD1-4 and Yao's datasets F4, F6) (Yao et al. 2014). Box & whiskers plots show the median of matched expression values of each gene for CN and FSHD1 and whiskers extend to the 5-95 percentiles. *P* values were calculated by paired two-tailed Wilcoxon matched-pairs signed rank test with confidence interval of 99%. Asterisks represent statistical *P* values; for CN vs. FSHD1 activated in MB $p<0.0001$; for CN vs. FSHD1 repressed in MB $p<0.0001$; for CN vs. FSHD1 repressed in MT $p<0.0001$.

Figure 3. Genes which have lost the interaction with 4q-D4Z4 have a more active chromatin state and are enriched for muscle atrophy signature in FSHD muscle cells

(*A*) Flowchart of filtering steps to identify FSHD1 altered genes. Genes within lost 4q-D4Z4 interactions were filtered as activated (red) or repressed (blue) in FSHD1. (*B*) Gene Ontology analysis (Biological Processes) of FSHD1 lost-activated and repressed genes. Bars correspond to $-\log_{10}$ of the *P* value. (*C*) Gene Set Enrichment Analysis (GSEA) results of the 319 FSHD1 lost-activated genes performed on expression data from unloading-induced muscle atrophy subjects (Reich et al. 2010). Genes upregulated in atrophic condition are depicted in red whereas genes not enriched are depicted in blue. NES, Normalized Enrichment Score. (*D*) Expression levels from RNA-seq datasets for atrophic genes (Reich et al. 2010), in CN (CN-3, CN-4 and Yao's datasets C20, C21, C22) and FSHD1 (FSHD1-3, FSHD1-4 and Yao's datasets F4, F6) (Yao et al. 2014) MB and MT. Box & whiskers plots show the median of matched expression values of each gene for CN and FSHD1 and whiskers extend to the 5-95 percentiles. *P* values were calculated by paired two-tailed *t*-test with confidence interval of 99%. Asterisks represent statistical *P* values; for CN vs. FSHD1 in MB $p=0.0099$. (*E*) Expression levels from RNA-seq datasets for atrophic genes (Reich et al. 2010), in CN (CN-3, CN-4 and Yao's datasets C20, C21, C22) and FSHD2 (Yao's datasets F12, F14, F20) (Yao et al. 2014) MB and MT. Box & whiskers plots show the median of matched expression values of each gene for CN and FSHD2 and whiskers extend to the 5-95 percentiles. *P* values were calculated by paired two-tailed *t*-test with confidence interval of 99%. Asterisks

represent statistical *P* values; for CN vs. FSHD2 in MB p=0.0251; for CN vs. FSHD2 in MT p=0.0041.

Figure 4. *FBXO32* gene has a deregulated chromatin structure and it is overexpressed in FSHD1 and FSHD2 muscle cells

(*A*) Representative nuclei of 3D multicolor DNA FISH using probes mapping to 4q35.1 region (4q, green) and *FBXO32* (red) in CN (CN-1, CN-3, CN-4), FSHD1 (FSHD1-1, FSHD1-3, FSHD1-4) and FSHD2 (FSHD2-1, FSHD2-2) myoblasts. Nuclei are counterstained with DAPI (blue). All images at 63X magnification. Scale bar=5 µm. (*B*) Cumulative frequency distribution of distances (below 1.5 µm) between 4q and *FBXO32* in CN (grey), FSHD1 (blue) and FSHD2 (dark blue) myoblasts. n= 3,652 (CN), 2,464 (FSHD1) and 1,020 (FSHD2). *P* values were calculated by unpaired one-tailed *t*-test with confidence interval of 99%. Asterisks represent statistical *P* values; for CN vs. FSHD1 p=0.0473; for CN vs. FSHD2 p=0.0036. (*C*) Percentage of nuclei positive for the interactions (under the cut-off of 1.5 µm). n= 913 (CN), 616 (FSHD1) and 255 (FSHD2). (*D*) 4C normalized coverage tracks at the *FBXO32* locus for *FBXO32*-4C VP in CN (CN-3, CN-4; grey) and FSHD1 (FSHD1-3, FSHD1-4; blue). (*E*) (top) Schematic representation of the *FBXO32* locus and Hind III sites. (middle) Chart showing the frequencies of 3C interaction between *FBXO32* promoter and the indicated Hind III restriction sites (sites 4-32), using the same bait of the 4C VP (light gray vertical bar) in CN (grey) and FSHD (blue). n=3 (CN) and 3 (FSHD1). S.e.m. is indicated. *P* values were calculated by two-way ANOVA followed by Bonferroni post-test correction. Asterisks represent statistical *P* values; for P19, P25 and P30 CN vs. FSHD1 p<0.001; for P32 CN vs. FSHD1 p<0.01. (bottom) 4C normalized coverage tracks as well as ChromHMM chromatin states tracks at the *FBXO32* locus for *FBXO32*-4C VP in CN (grey) and FSHD1 (blue). The arrow represents the promoter region; enhancers are highlighted in yellow. (*F*) Bar plot showing enrichment of RNA Pol II at *FBXO32* promoter (left) and an intragenic region (right) assessed by ChIP-qPCR experiment in CN (grey) and FSHD1 (blue) myoblasts. Results are presented as % of input. n=2 CN (CN-3, CN-4) and 2 FSHD1 (FSHD3, FSHD1-4). S.e.m. is indicated. *P* values were calculated by unpaired one-tailed *t*-test with confidence interval of 99%. Dots represent the values of each

replicate; asterisks represent statistical *P* values; for *FBXO32* intragenic region CN vs. FSHD1 p=0.0050. (*G*) Expression levels of *FBXO32* gene during CN (grey), FSHD1 (blue) and FSHD2 (dark blue) differentiation (MB, myoblasts, MT2, myotubes day 2, MT4, myotubes day 4, MT6, myotubes day 6). Data were normalized on *GAPDH* expression and on MB. n=4 CN (CN-1, CN-2, CN-3, CN-4), 4 FSHD1 (FSHD1-1, FSHD1-2, FSHD1-3, FSHD1-4) and 2 FSHD2 (FSHD2-1, FSHD2-2). S.e.m. is indicated. *P* values were calculated by two-way ANOVA followed by Bonferroni post-test correction. Dots represent the values of each replicate; asterisks represent statistical *P* values; for MT4, CN vs. FSHD1 p<0.0290 and CN vs. FSHD2 p<0.0001.

Figure 5. Ectopic 4q-D4Z4 array restores the expression of FSHD1 lost interacting genes

(*A*) (top) Representation of the BAC containing 4q upstream region and D4Z4 array (at least 15 D4Z4 units, B Bodega, unpublished (BAC 4q-D4Z4n)). (bottom) Representative nucleus of 3D multicolor DNA FISH using probes for the transfected BAC backbone (red) and 4q35.1 region (4q, green) in myoblasts transfected with BAC 4q-D4Z4n. Nuclei are counterstained with DAPI (blue). All images at 63X magnification. Scale bar=5 µm. n, number of nuclei analyzed. (*B*) Representative nucleus of 3D multicolor DNA FISH using probes for the transfected BAC backbone (red) and 4q35.1 region (4q, green) in myoblasts transfected with Ctrl BAC (RP11-2A16, representative of an unrelated and not interacting genomic region, Chr 17q21.33). Nuclei are counterstained with DAPI (blue). All images at 63X magnification. Scale bar=5 µm. n, number of nuclei analyzed. (*C*) Representative nucleus of 3D multicolor DNA FISH using probes for the transfected BAC backbone (red) and *FBXO32* region (*FBXO32*, light blue) in myoblasts transfected with BAC 4q-D4Z4n. Nuclei are counterstained with DAPI (blue). All images at 63X magnification. Scale bar=5 µm. n, number of nuclei analyzed. (*D*) Bar plots showing expression levels of *FBXO32*, *TRIB3* and *ZNF555* (FSHD1 lost-activated genes) in CN (grey) and FSHD1 (blue) myoblasts transfected with Ctrl BAC and BAC 4q-D4Z4n. Data were normalized on *GAPDH* expression. n=at least 3 (with the exception of *TRIB3* and *ZNF555* CN Ctrl BAC, n=2). S.e.m. is indicated. *P* values were calculated by paired one-tailed *t*-test with confidence interval of 99%. Dots represent the values of each replicate; asterisks represent statistical *P* values; for *FBXO32*

637 Ctrl BAC vs. BAC 4q-D4Z4n in CN *p*=0.0182; for *FBXO32* Ctrl BAC vs BAC 4q-D4Z4n in FSHD1
638 *p*=0.0073; for *TRIB3* Ctrl BAC vs BAC 4q-D4Z4n in FSHD1 *p*=0.0281. (*E*) Bar plot showing
639 expression levels of *LZTS3* (FSHD1 lost-repressed gene) in CN (grey) and FSHD1 (blue)
640 myoblasts transfected with Ctrl BAC and BAC 4q-D4Z4n. Data were normalized on *GAPDH*
641 expression. n=3 (with the exception of CN Ctrl BAC, n=2). S.e.m. is indicated. *P* value was
642 calculated by paired one-tailed *t*-test with confidence interval of 99%. Dots represent the values of
643 each replicate; asterisks represent statistical *P* values; for Ctrl BAC vs. BAC 4q-D4Z4n in FSHD1
644 *p*=0.0296. (*F*) Bar plots showing expression levels of *FOXO3* and *MYOG* (not interacting genes) in
645 CN (grey) and FSHD1 (blue) myoblasts transfected with Ctrl BAC and BAC 4q-D4Z4n. Data were
646 normalized on *GAPDH* expression. n=at least 3. S.e.m. is indicated. Dots represent the values of
647 each replicate.
648
649 **Figure 6. Model of 4q-D4Z4 mediated regulation of atrophic genes transcription**
650 (*A*) 4q-D4Z4 array is interacting with a subset of atrophic genes, organizing their chromatin
651 structure and keeping on hold their transcription in healthy donor muscle cells. (*B*) In FSHD1
652 patients' muscle cells, the deleted and hypomethylated 4q-D4Z4 array causes an impairment of
653 D4Z4 interactome leading to a chromatin switch towards an active state (mainly enhancer and
654 promoter regions), which in turn results in the transcriptional upregulation of the atrophic genes.
655

References

Achinger-Kawecka J, Clark SJ. 2017. Disruption of the 3D cancer genome blueprint. *Epigenomics* **9**: 47-55.

Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, van der Wielen M, Rasmussen K, Frants RR. 1995. The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve Suppl*: S39-44.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**: 1091-1093.

Block GJ, Narayanan D, Amell AM, Petek LM, Davidson KC, Bird TD, Tawil R, Moon RT, Miller DG. 2013. Wnt/beta-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum Mol Genet* **22**: 4661-4672.

Bodega B, Cardone MF, Muller S, Neusser M, Orzan F, Rossi E, Battaglioli E, Marozzi A, Riva P, Rocchi M et al. 2007. Evolutionary genomic remodelling of the human 4q subtelomere (4q35.2). *BMC Evol Biol* **7**: 39.

Bodega B, Cardone MF, Rocchi M, Meneveri R, Marozzi A, Ginelli E. 2006. The boundary of macaque rDNA is constituted by low-copy sequences conserved during evolution. *Genomics* **88**: 564-571.

Bodega B, Marasca F, Ranzani V, Cherubini A, Della Valle F, Neguembor MV, Wassef M, Zippo A, Lanzuolo C, Pagani M et al. 2017. A cytosolic Ezh1 isoform modulates a PRC2-Ezh1 epigenetic adaptive response in postmitotic cells. *Nat Struct Mol Biol* doi:10.1038/nsmb.3392.

Bodega B, Orlando V. 2014. Repetitive elements dynamics in cell identity programming, maintenance and disease. *Curr Opin Cell Biol* **31**: 67-73.

Bodega B, Ramirez GD, Grasser F, Cheli S, Brunelli S, Mora M, Meneveri R, Marozzi A, Mueller S, Battaglioli E et al. 2009. 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.

Bodine SC, Baehr LM. 2014. Skeletal muscle atrophy and the E3 ubiquitin ligases MuRF1 and MAFbx/atrogin-1. *Am J Physiol Endocrinol Metab* **307**: E469-484.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Bosnakovski D, Xu Z, Gang EJ, Galindo CL, Liu M, Simsek T, Garner HR, Agha-Mohammadi S, Tassin A, Coppee F et al. 2008. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* **27**: 2766-2779.

Broucqsault N, Morere J, Gaillard MC, Dumonceaux J, Torrents J, Salort-Campana E, Maues De Paula A, Bartoli M, Fernandez C, Chesnais AL 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.

695 Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D. 2012. A long ncRNA
696 links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular
697 dystrophy. *Cell* **149**: 819-831.

698 Cabianca DS, Gabellini D. 2010. The cell biology of disease: FSHD: copy number variations on the
699 theme of muscular dystrophy. *J Cell Biol* **191**: 1049-1060.

700 Cheutin T, Cavalli G. 2018. Loss of PRC1 induces higher-order opening of Hox loci independently
701 of transcription during Drosophila embryogenesis. *Nat Commun* **9**: 3898.

702 Corces MR, Corces VG. 2016. The three-dimensional cancer genome. *Curr Opin Genet Dev* **36**: 1-
703 7.

704 Cortesi A, Bodega B. 2016. Chromosome Conformation Capture in Primary Human Cells. *Methods*
705 *Mol Biol* **1480**: 213-221.

706 Cremer M, Grasser F, Lanctot C, Muller S, Neusser M, Zinner R, Solovei I, Cremer T. 2008.
707 Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes.
708 *Methods Mol Biol* **463**: 205-239.

709 Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang SC, Sanborn
710 AL, Machol I, Shamim M et al. 2016. Deletion of DXZ4 on the human inactive X
711 chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* **113**:
712 E4504-4512.

713 Daxinger L, Tapscott SJ, van der Maarel SM. 2015. Genetic and epigenetic contributors to FSHD.
714 *Curr Opin Genet Dev* **33**: 56-61.

715 de Greef JC, Lemmers RJ, van Engelen BG, Sacconi S, Venance SL, Frants RR, Tawil R, van der
716 Maarel SM. 2009. Common epigenetic changes of D4Z4 in contraction-dependent and
717 contraction-independent FSHD. *Hum Mutat* **30**: 1449-1459.

718 de Laat W, Duboule D. 2013. Topology of mammalian developmental enhancers and their
719 regulatory landscapes. *Nature* **502**: 499-506.

720 Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, Felicetti L. 1995. Physical mapping evidence
721 for a duplicated region on chromosome 10qter showing high homology with the
722 facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* **3**:
723 155-167.

724 Deng X, Ma W, Ramani V, Hill A, Yang F, Ay F, Berletch JB, Blau CA, Shendure J, Duan Z et al.
725 2015. Bipartite structure of the inactive mouse X chromosome. *Genome Biol* **16**: 152.

726 Dixit M, Ansseau E, Tassin A, Winokur S, Shi R, Qian H, Sauvage S, Matteotti C, van Acker AM,
727 Leo O et al. 2007. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy,
728 encodes a transcriptional activator of PITX1. *Proc Natl Acad Sci U S A* **104**: 18157-18162.

729 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
730 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

731 Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization.
732 *Nat Methods* **9**: 215-216.

733 Finn EH, Pegoraro G, Brandao HB, Valton AL, Oomen ME, Dekker J, Mirny L, Misteli T. 2019.
734 Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**:
735 1502-1515 e1510.

736 Gabellini D, Green MR, Tupler R. 2002. Inappropriate gene activation in FSHD: a repressor
737 complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110**: 339-348.

738 Gabriels J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, Padberg GW,
739 Frants RR, Hewitt JE, Collen D et al. 1999. Nucleotide sequence of the partially deleted
740 D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element.
741 *Gene* **236**: 25-32.

742 Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, van der Maarel SM, Ruzzo WL,
743 Gentleman RC, Tawil R et al. 2012. DUX4 activates germline genes, retroelements, and
744 immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell* **22**: 38-51.

745 Giacalone J, Friedes J, Francke U. 1992. A novel GC-rich human macrosatellite VNTR in Xq24 is
746 differentially methylated on active and inactive X chromosomes. *Nat Genet* **1**: 137-143.

747 Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY, Heard E
748 et al. 2016. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**:
749 575-579.

750 Gomes MD, Lecker SH, Jagoe RT, Navon A, Goldberg AL. 2001. Atrogin-1, a muscle-specific F-
751 box protein highly expressed during muscle atrophy. *Proc Natl Acad Sci U S A* **98**: 14440-
752 14445.

753 Goto K, Oda H, Kondo H, Igaki M, Suzuki A, Tsuchiya S, Murase T, Hase T, Fujiya H, Matsumoto I
754 et al. 2011. Responses of muscle mass, strength and gene transcripts to long-term heat
755 stress in healthy human subjects. *Eur J Appl Physiol* **111**: 17-27.

756 Gregory TR. 2005. Synergy between sequence and size in large-scale genomics. *Nat Rev Genet*
757 **6**: 699-708.

758 Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. 2017. Improvements and impacts of GRCh38
759 human reference on high throughput sequencing data analysis. *Genomics* **109**: 83-90.

760 Hendrickson PG, Dorais JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, Weaver BD, Pflueger C,
761 Emery BR, Wilcox AL et al. 2017. Conserved roles of mouse DUX and human DUX4 in
762 activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**: 925-
763 934.

764 Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, van Deutekom JC, Francis F,
765 Sharpe PT, Hofker M et al. 1994. Analysis of the tandem repeat locus D4Z4 associated
766 with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* **3**: 1287-1295.

767 Himeda CL, Debarnot C, Homma S, Beermann ML, Miller JB, Jones PL, Jones TI. 2014. Myogenic
768 enhancers regulate expression of the facioscapulohumeral muscular dystrophy-associated
769 DUX4 gene. *Mol Cell Biol* **34**: 1942-1955.

770 Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during
771 Zygotic Genome Activation Independent of Transcription. *Cell* **169**: 216-228 e219.

772 Jansz N, Keniry A, Trussart M, Bildsoe H, Beck T, Tonks ID, Mould AW, Hickey P, Breslin K,
773 Iminitoff M et al. 2018. Smchd1 regulates long-range chromatin interactions on the inactive
774 X chromosome and at Hox clusters. *Nat Struct Mol Biol* **25**: 766-777.

775 Jiang G, Yang F, van Overveld PG, Vedanarayanan V, van der Maarel S, Ehrlich M. 2003. Testing
776 the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by

analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol Genet* **12**: 2909-2921.

Jiang T, Raviram R, Snetkova V, Rocha PP, Proudhon C, Badri S, Bonneau R, Skok JA, Kluger Y. 2016. Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions. *Nucleic Acids Res* **44**: 8714-8725.

Krijger PH, de Laat W. 2016. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* **17**: 771-782.

Krijger PHL, de Laat W. 2017. Can We Just Say: Transcription Second? *Cell* **169**: 184-185.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Lanzuolo C. 2012. Epigenetic alterations in muscular disorders. *Comp Funct Genomics* **2012**: 256892.

Lecker SH, Jagoe RT, Gilbert A, Gomes M, Baracos V, Bailey J, Price SR, Mitch WE, Goldberg AL. 2004. Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. *FASEB J* **18**: 39-51.

Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. 2002. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat Genet* **32**: 235-236.

Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR 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.

Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camano P, Dauwerse JG, Snider L, Straasheijm KR, van Ommen GJ, Padberg GW et al. 2010. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**: 1650-1653.

Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, de Knijff P, Padberg GW, Frants RR, van der Maarel SM. 2007. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am J Hum Genet* **81**: 884-894.

Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R et al. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**: 1012-1025.

Lupianez DG, Spielmann M, Mundlos S. 2016. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**: 225-237.

Lyle R, Wright TJ, Clark LN, Hewitt JE. 1995. The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* **28**: 389-397.

Montigny WJ, Phelps SF, Illenye S, Heintz NH. 2003. Parameters influencing high-efficiency transfection of bacterial artificial chromosomes into cultured mammalian cells. *Biotechniques* **35**: 796-807.

817 Neguembor MV, Gabellini D. 2010. In junk we trust: repetitive DNA, epigenetics and
818 facioscapulohumeral muscular dystrophy. *Epigenomics* **2**: 271-287.

819 Petrov A, Pirozhkova I, Carnac G, Laoudj D, Lipinski M, Vassetzky YS. 2006. Chromatin loop
820 domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients
821 versus normal human myoblasts. *Proc Natl Acad Sci U S A* **103**: 6982-6987.

822 Politz JC, Scalzo D, Groudine M. 2013. Something silent this way forms: the functional
823 organization of the repressive nuclear compartment. *Annu Rev Cell Dev Biol* **29**: 241-270.

824 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
825 *Bioinformatics* **26**: 841-842.

826 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol
827 I, Omer AD, Lander ES et al. 2014. A 3D map of the human genome at kilobase resolution
828 reveals principles of chromatin looping. *Cell* **159**: 1665-1680.

829 Raviram R, Rocha PP, Muller CL, Miraldi ER, Badri S, Fu Y, Swanzey E, Proudhon C, Snetkova V,
830 Bonneau R et al. 2016. 4C-ker: A Method to Reproducibly Identify Genome-Wide
831 Interactions Captured by 4C-Seq Experiments. *PLoS Comput Biol* **12**: e1004780.

832 Reich KA, Chen YW, Thompson PD, Hoffman EP, Clarkson PM. 2010. Forty-eight hours of
833 unloading and 24 h of reloading lead to changes in global gene expression patterns related
834 to ubiquitination and oxidative stress in humans. *J Appl Physiol (1985)* **109**: 1404-1415.

835 Rickard AM, Petek LM, Miller DG. 2015. Endogenous DUX4 expression in FSHD myotubes is
836 sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum*
837 *Mol Genet* **24**: 5901-5914.

838 Rivera-Reyes A, Hayer KE, Bassing CH. 2016. Genomic Alterations of Non-Coding Regions
839 Underlie Human Cancer: Lessons from T-ALL. *Trends Mol Med* **22**: 1035-1046.

840 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-
841 Moussavi A, Kheradpour P, Zhang Z, Wang J et al. 2015. Integrative analysis of 111
842 reference human epigenomes. *Nature* **518**: 317-330.

843 Robin JD, Ludlow AT, Batten K, Gaillard MC, Stadler G, Magdinier F, Wright WE, Shay JW. 2015.
844 SORBS2 transcription is activated by telomere position effect-over long distance upon
845 telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy.
846 *Genome Res* **25**: 1781-1790.

847 Sacconi S, Salviati L, Desnuelle C. 2015. Facioscapulohumeral muscular dystrophy. *Biochim*
848 *Biophys Acta* **1852**: 607-614.

849 Sandri M, Sandri C, Gilbert A, Skurk C, Calabria E, Picard A, Walsh K, Schiaffino S, Lecker SH,
850 Goldberg AL. 2004. Foxo transcription factors induce the atrophy-related ubiquitin ligase
851 atrogin-1 and cause skeletal muscle atrophy. *Cell* **117**: 399-412.

852 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker
853 T. 2003. Cytoscape: a software environment for integrated models of biomolecular
854 interaction networks. *Genome Res* **13**: 2498-2504.

855 Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der
856 Maarel SM, Tapscott SJ et al. 2010. Facioscapulohumeral dystrophy: incomplete
857 suppression of a retrotransposed gene. *PLoS Genet* **6**: e1001181.

Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W. 2012. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**: 221-230.

Statland JM, Tawil R. 2016. Facioscapulohumeral Muscular Dystrophy. *Continuum (Minneap Minn)* **22**: 1916-1931.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**: 15545-15550.

Tam R, Smith KP, Lawrence JB. 2004. The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J Cell Biol* **167**: 269-279.

Tawil R, Storvick D, Feasby TE, Weiffenbach B, Griggs RC. 1993. Extreme variability of expression in monozygotic twins with FSH muscular dystrophy. *Neurology* **43**: 345-348.

Tawil R, Van Der Maarel SM. 2006. Facioscapulohumeral muscular dystrophy. *Muscle Nerve* **34**: 1-15.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**: 46-53.

van Deutekom JC, Wijmenga C, van Tienhoven EA, Gruter AM, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR. 1993. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet* **2**: 2037-2042.

van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. 2003. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* **35**: 315-317.

Wang CY, Jegu T, Chu HP, Oh HJ, Lee JT. 2018. SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* **174**: 406-421 e425.

Wijchers PJ, Geeven G, Eyres M, Bergsma AJ, Janssen M, Verstegen M, Zhu Y, Schell Y, Vermeulen C, de Wit E et al. 2015. Characterization and dynamics of pericentromere-associated domains in mice. *Genome Res* **25**: 958-969.

Woltering JM, Duboule D. 2015. Tetrapod axial evolution and developmental constraints; Empirical underpinning by a mouse model. *Mech Dev* **138 Pt 2**: 64-72.

Woltering JM, Noordermeer D, Leleu M, Duboule D. 2014. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* **12**: e1001773.

Yao Z, Snider L, Balog J, Lemmers RJ, Van Der Maarel SM, Tawil R, Tapscott SJ. 2014. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet* **23**: 5342-5352.

Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, Balog J, Tawil R, van der Maarel SM, Tapscott SJ. 2013. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet* **9**: e1003947.

898 Zeng W, de Greef JC, Chen YY, Chien R, Kong X, Gregson HC, Winokur ST, Pyle A, Robertson
899 KD, Schmiesing JA et al. 2009. Specific loss of histone H3 lysine 9 trimethylation and
900 HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral
901 dystrophy (FSHD). *PLoS Genet* **5**: e1000559.

902 Zheng M, Tian SZ, Capurso D, Kim M, Maurya R, Lee B, Piecuch E, Gong L, Zhu JJ, Li Z et al.
903 2019. Multiplex chromatin interactions with single-molecule precision. *Nature* **566**: 558-562.

904