SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dottorato di Ricerca in / PhD program IN TRANSLATIONAL AND MOLECULAR MEDICINE
UNIVERSITY OF MILANO-BICOCCA
SCHOOL OF MEDICINE AND SCHOOL OF SCIENCE

Ciclo / Cycle XXIX

Chromatin landscape of D4Z4 repeat interactome reveals a muscle atrophy signature in Facioscapulohumeral Dystrophy

Cognome / Surname Cortesi Nome / Name Alice

Matricola / Registration number 787776

Tutore / Tutor: Dr. Beatrice Bodega

Coordinatore / Coordinator: Prof. Andrea Biondi

ANNO ACCADEMICO / ACADEMIC YEAR 2015-2016

Table of contents

Chapter 2 “Chromosome Conformation Capture in primary human cells”

Chapter 3 “Chromatin landscape of D4Z4 repeat interactome reveals a muscle atrophy signature in Facioscapulohumeral Dystrophy”

Chapter 4 Summary, conclusions and future perspectives

Chapter 1 General introduction

1.1 Epigenetics and Epigenome: concepts and definitions

Genetic information in *H. sapiens* is contained and distributed among 23 couples of chromosomes in a variety of different combination between protein coding genes and regulatory elements; this constitutes the basic information for the establishment of cellular identity and the regulation of processes such as development, differentiation, homeostasis, etc. *H. sapiens* retains at least of 200 of morphological, structural and functional different cell types (phenotypes); although they share the same genome they all differ because of a different epigenome. In 1940 Conrad Waddington defined epigenetics as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being”. Today we defined the epigenetics as each hereditary information not directly linked to the sequence of DNA, as the sum of phenotypic variations not directly correlated to genotypic differences (“Epigenetics” Allis C.D., Jenuwein T., Reinberg D. Cold Spring Harbor Laboratory Press, Cap3). Indeed gene expression is orchestrated in a cell specific manner through different levels of epigenetic regulation, that go beyond the canonical interplay between DNA sequences and transcription factors and imply DNA methylation, histone variants, histone modifications and chromatin remodeling, finally including nuclear architecture, and the presence of non coding, regulative RNAs with multiple functions that establish in

concert a specific epigenome [1-5] (Figure 1). The epigenome contributes to the quality, stability and heritability of cell specific transcriptional programs that undergo profound changes during development, cell differentiation and/or metabolic switches.

Damage or perturbation of epigenetic components may lead to deviations from a determined cellular program, resulting in severe developmental disorders, tumor progression [6,7] and could explain human complex diseases [8].

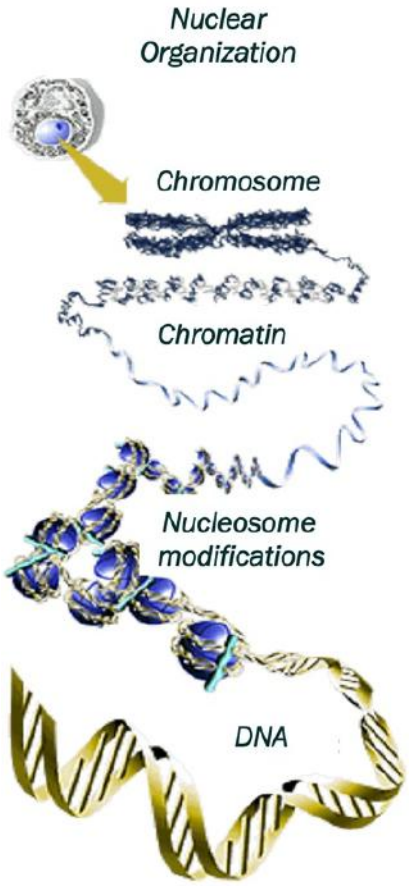


Figure 1. Genome is highly organized through different levels of epigenetic regulation.

Epigenetic regulation consists of changes in chromatin structure mediated, primarily, by DNA methylation, secondarily by histone variants substitution into nucleosomes and posttranslational modification of histone tails and thirdly by higher order chromatin arrangement, responsible for nuclear architecture and 3D genome organization. To various extents, all these levels of organization contribute to the stability and heritability of transcription programs and define what is meant as the epigenomic level of gene regulation.

1.1.1 Genetic information is highly organized into chromatin, a platform for epigenetic regulation

An extraordinary combination of mechanisms act on chromatin regulating the huge amount of information stored in the genome; it is possible to discriminate between three hierarchically interconnected levels at which epigenetic control of the genome takes place.

The first level of regulatory mechanism is found on the DNA molecule. Specific DNA sequences mediate the targeting of nuclear factors that regulate the transcriptional state.

The second regulatory level includes marks such as methylation of the DNA itself, covalent post-translational modifications (PTMs) of the core histones, histone variants and other chromatin-associated proteins that “read” or “write” previous marks, changing the functional state of chromatin in response to environmental challenges [9,1]. Various studies report somewhat different classifications of chromatin types, mostly depending on the parameters used in the computational analysis, but the general consensus is that repressive chromatin accounts for Polycomb-bound euchromatin, heterochromatin and a chromatin state that has no strong enrichment for any of the specific factors or marks used for mapping [10-12], in contrast, there are various types of active or open chromatin, and it has proven more difficult to rigorously classify them, probably because the classification depends on

the number of factors that are used for mapping. However, at least four types of open chromatin can be distinguished with some certainty, encompassing ‘enhancers’, ‘promoters’, ‘transcribed regions’ and ‘regions bound by chromatin insulator proteins’ [13].

A third regulatory level is the three-dimensional (3D) organization of chromatin in the nucleus that reflects basic nuclear metabolisms [5]. As an additional and interconnected level, today is becoming more and more clear the regulatory function of non-coding transcripts both in epigenetic marks deposition and 3D structure organization [14-17].

Genomes are spatially arranged at several, hierarchical levels in the 3D space of the cell nucleus, starting with the folding of the chromatin fiber into higher-order structures, the formation of loops over a wide range of genomic distances and the formation of chromosome domains, (sub-megabase topologically associating domains (TADs) [18-21] and multi-megabase active and inactive compartments [22]), culminating in their aggregation to form chromosome territories [23].

The eukaryotic genome is packaged in a highly organized fashion to fit within the spatial constraints of the cell nucleus and at the same time allow for access of regulatory factors to the underlying sequences for nuclear metabolisms [24]. The first layer of packaging involves the winding of 147 base pairs of DNA around an octamer of core histone proteins (H3-H4 tetramers and two dimers of H2-H2B) to form a nucleosome

[25]. Core nucleosomes are linked by linker DNA [26,25] and H1 histones.

A string of nucleosomes presents itself as an 11 nm fiber. The next layer of packaging involves the helical stacking of nucleosomes to form a chromatin fiber with a diameter of ~30 nm [27]. This in turn is highly organized in loops that defined functional and topological domains [2,28,5,29].

There are four types of loops that have direct functional consequences for transcription [30].

The first type joins the 5′ end of transcribed genes with the transcription termination site.

The second type of regulatory loop brings distant enhancers in contact with promoters.

A third type of looped transcriptional regulation is Polycomb-dependent repression via looping of regions containing Polycomb response elements to reach distal gene promoters.

A fourth type of looping interactions involves insulator-binding proteins, such as CTCF, cohesin and insulator-binding proteins that are present in insects but not in mammals [31,23].

Moreover that loops display properties of being both dynamic and relatively static in different contexts, indicating that their formation can be independently regulated [32].

At a bigger scale discrete chromatin domains (so-called topological domains) are folded; the domain size depends on the chromosomal region, the cell type and the species, spanning few tens of kilobases to several megabases

(averaging ~100 kb in flies and ~1 Mb in humans) [33,10,11] [34,12,13,35].

Domains of active “open” and inactive “closed” chromatin appears to be partitioned into separate sub-nuclear domains [22,36].

Finally these discrete chromatin domains constitute chromosome territories [23].

The nuclear localization also influences gene expression, regulating its access to specific machinery responsible for specific functions, such as transcription or replication [37,38]. In addition, due to its highly dynamic nature, the genome moves in the nucleus driving specific genomic regions toward nuclear compartments defined by a high concentration of specific factors and substrates that facilitate more efficient biological reactions [39]. This constant motion plays also a role in coordinating the expression of coregulated genes, separated by longer chromosomal regions or located on different chromosomes [40,8].

Going more deeply in the 3D scenario, although individual chromosomes occupy distinct territories, they show substantial intermingling allowing for interchromosomal contacts [41,36]. In particular, both active and inactive domains undergo long-range interactions. Moreover, whereas loops lead to juxtaposition of genome regions on the same chromosome, functional interactions between distinct chromosomes are also emerging as prominent functional regulators [23].

Indeed, with the advent of new technologies that detect a broader spectrum of chromatin contacts, interactions in trans are being increasingly appreciated, although their functions remain largely obscure [42,32].

The idea that the 3D genome topology is in some way functional to the regulation of nuclear activities takes place and a lot of work is done in this direction [43].

Rather, a full knowledge of genome function in vivo requires investigation and understanding of the three-dimensional (3D) folding and spatial organization of chromosomes in the nucleus [44]. Two experimental approaches that have been extensively used for this purpose are DNA fluorescent in situ hybridization (DNA FISH) [45] and chromosome conformation capture (3C) and its derivatives [46]. In the past, DNA FISH was the method of choice for investigations of the 3D structure of the genome [47]. FISH is visually compelling but generally limited to looking at the locations of a few specific targets in a few hundred cells, although recent probe developments based on massively parallel custom oligonucleotide synthesis have expanded the scope and scale of sequences that can be analyzed in each hybridization reaction [48]. DNA FISH has nevertheless allowed many fundamental discoveries to be made, such as the existence of chromosomal territories [49] and the dynamic repositioning of genomic loci with respect to nuclear compartments (such as the nuclear periphery) during differentiation [50].

The recent advent of 3C-based approaches (such as circularized chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C), and Hi-C [46] has revolutionized the field of nuclear organization, enabling the detection of physical proximity between multiple genomic loci (and eventually across an entire genome) simultaneously [47]. First study, using 3C to explore the three-dimensional organization of chromosomes at high resolution, describes intrachromosomal interactions between telomeres as well as interchromosomal interactions between centromeres and between homologous chromosomes in yeast [51]. 3C method was applied to analyze physical connections between genes and *cis* enhancers (mouse and human β-*globin* locus model [52-54]). Then, highly specific associations between loci located on separate chromosomes are described. These *trans*-interactions can be between a distant enhancer with different target genes (olfactory receptor genes [55]). In other cases, *trans*-interactions appear to play a role in a higher level of gene control to coordinately regulate multiple loci with a set of both intrachromosomal and interchromosomal interactions (T helper 2 cytokine locus [56]), or providing additional levels of gene regulation by allowing combinatorial association of genes and sets of regulatory elements (imprinted loci [57]). It is also reported that specific DNA binding sites could mediate the formation of this topologically complex structure (Polycomb Response Elements in Drosophila [58]). Finally, it is

demonstrated that 3D structure could also have a role in developmental processes (mammalian X-chromosome inactivation [59,60]).

1.2 Repetitive elements and their role in epigenetics

In the last decade, thanks to the advent of high throughput next generation sequencing, the genomic sequences of *Homo sapiens* and several model organisms became available and it was surprisingly found that the number of protein-coding genes does not correlate with organism complexity [61]. Comparative genomic studies have revealed that the proportion of the genome occupied by genes decreases as biological complexity increases [62]. But more interestingly, the non-protein coding component of the genomic DNA, and in particular repetitive elements, represent a progressively larger proportion of the genome in organisms with increasing complexity, suggesting that it might significantly contribute to higher eukaryotes sophistication [63]. Recent estimations indicate that repetitive sequences could account for up to 66–69% of the human genome [64], while genes occupy less than 2% [65].

Although this, the repetitive fraction of the genome is largely ignored, but there is increasing evidence of the peculiar functions of the repeated (epi)genome [66].

Indeed, the role of DNA repeats in chromosome structural organization, gene regulation, genome integrity, and evolution has been described [67,65,68-70].

DNA repeats can be also transcribed, frequently in a cell and tissue-specific fashion, moreover there is a large proportion of capped-transcripts initiating from repetitive units. It has been

suggested that these can provide regulatory elements to protein-coding genes, such as alternative promoters, exons, or polyadenylation sites, and ncRNAs, thus significantly expanding the regulatory capability of higher eukaryote genomes [71,72] [73,74]. Moreover, binding sites for important regulatory factors such as CTCF or TP53 are often associated with genomic repeats [71,72,75,76].

Repetitive elements can either mobilize or rearrange in somatic tissues, thus providing an unexpected dynamic dimension to the normal physiology of the soma, but also contributing to the etiopathogenesis of diseases [77,69,70].

Repetitive elements could be classified as widely interspersed repeats or they can be located one next to another to form tandem repeats. Repeats can range in size from 1 to 2 bases to millions of bases and might comprise just two copies or millions of copies [78-82].

1.2.2 Interspersed repeats

Interspersed repeats, also called transposable elements (TEs), are the results of ancient or present activity of mobile genetic elements [66].

Intriguingly, the proportion of the genome occupied by transposable elements (TEs), increases as biological complexity increases [62] , meanings that they have a significant role in evolution and in generating genetic diversity

[83] .

Being significant contributors to the copy number variation present in humans, mobile elements are also an important source of genetic variation [84-88].

TE could be divided in DNA transposons and retro-transposons, on the base of their transposition mechanism (Table 1). These elements can mediate their own mobilization respectively by a cut-and-paste mechanism or by a copy-and-paste process taking advantage of RNA as intermediate [89].

Approximately 3% of a typical mammalian genome is made of DNA transposons [65]; however, with the exception of some bat species, DNA transposons no longer mobilize in mammals [90,91].

In contrast, retro-transposons comprise more than 40% of a typical mammalian genome and are still active in most mammalian species [92,65,93].

DNA transposons generally move by a cut-and-paste mechanism in which the transposon is excised from one location and reintegrated elsewhere. Most DNA transposons move through a non-replicative mechanism. DNA transposons consist of a transposase gene that is flanked by two *Terminal Inverted Repeats* (TIRs). The transposase recognizes these TIRs to perform the excision of the transposon DNA body, which is inserted into a new genomic location. Upon insertion, target site DNA is duplicated, resulting in *Target Site Duplications* (TSDs), which represent a unique hallmark for

each DNA transposon [94].

Retro-transposons are named autonomous elements, when they contain the activities necessary for their retrotransposition, in particular, a reverse transcriptase (RT) activity, which reverse-transcribes and integrates the TE transcript into a new genomic location. In addition to the autonomous retro-transposons, there are a large number of non autonomous retro-transposons in mammalian genomes. These elements do not encode any proteins. Therefore, they require activities encoded by other autonomous repeats for their mobility [95].

Retro-transposons are composed of long terminal repeat (LTR) and non-LTR containing elements. The LTR retro-transposons are endogenous retroviruses (ERVs) that have lost the ability to go outside the host cell due to a non-functional envelope gene. Non-LTR retro-transposons are represented principally by long interspersed elements (LINEs) and short interspersed elements (SINEs) [66].

LTR retro-transposons are similar to retroviruses in terms of their structure and mechanism of retrotransposition and are hence often called endogenous retroviruses (ERVs; [96]). Full length ERVs are flanked by LTRs that promote the transcription and maturation of ERV RNAs, and they also contain functional gag and pol genes, which encode structural proteins and enzymes involved in retrotransposition. However, ERVs often lack a functional env gene, which encodes the envelope protein that retroviruses typically use to exit cells [97]. Furthermore,

recombination between LTRs occurs frequently, deleting the intervening internal ERV sequence and generating only LTRs [98]. The mobilization of active ERVs involves an RNA intermediate and a copy-and-paste mechanism that is similar to the initial steps of retroviral infection. ERVs generally no longer mobilize in humans [96] [99] .

Mammalian non-LTR retro-transposons are exemplified by long interspersed element class 1 (LINE-1) retro-transposons [100].

In humans, the most important LINE is the RNA polymerase II transcribed LINE-1 (L1). L1 is the only element able to encode the proteins required for mobilization [66]. Notably, active LINE-1 elements encode two protein products termed open reading frame (ORF) 1 and ORF2 that are strictly required for LINE-1 mobilization [101]. While ORF1 encodes an RNA binding protein with nucleic acid chaperone activity [102,103], ORF2 codes for a protein with endonuclease and reverse transcriptase activity [104,105] . Hence, these are the only known autonomously active human retro-transposons [92,106] [93].

LINE-1 retro-transposons make up 17% of the human genome and, although most are molecular fossils that have lost their ability to move, 80-100 copies of LINE-1 retain retrotransposition potential [107,84].

L1s are also responsible for the mobilization of the non-autonomous Alus, SVAs and processed pseudogenes (cellular mRNAs that become substrates of the reverse transcriptases

and are inserted into the genome)[66] .

Mammalian genomes also contain numerous short interspersed element (SINE) non-LTR retro-transposons, exemplified by the RNA polymerase III transcribed Alus and SVA (SINE-VNTR-Alu) in the human genome [65]. SINEs are non autonomous retro-transposons that use LINE-1 proteins *in trans* to mobilize [108-111]. Non-LTR retro-transposons also move by a copy-and-paste mechanism, but one that is fundamentally different from that used by LTR retro-transposons [93] .

LINE-1 elements are thought to mobilize during at least two different developmental contexts: the early embryo and the developing/adult brain [92] [106] [93]. By contrast, the expression and mobilization of TEs in other somatic cells in humans appears to be low.

L1 mobilization has been associated with brain cell development, where the occurrence of L1 retrotransposition in adult cells has been suggested to contribute to neuronal somatic diversification [112].

Retrotransposition, with only few exceptions such as V(D)J recombination [113], is an almost unique source of somatic genetic mosaicism, leading not only to heritable genetic variation but also to intra-individual variability.

Mobile elements can display differential activity in different tissues of the soma, suggesting that every individual is a genetic mosaic variegated by the differential insertion of mobile elements [112,114]. This represents a revolutionary concept

that is changing the view of this class of repetitive elements [83].

Additionally, mobilization of L1 repeats has been associated with both physiological and pathological processes and is regulated by DNA methylation [114].

When expressed, TEs can affect developmental processes either via their gene products, which can influence the behavior of host cells, or through new insertions that cause genetic changes in the host genome. New TE insertions into or near genes can act as insertional mutagens in mammalian genomes and interfere with gene function [92,115,106,116,93].

Such insertions can, for example, introduce actively transcribing promoters into genes and cause transcriptional interference.

TE promoters can drive the expression of novel transcripts that encompass part of the coding region. The co-option of TE-derived sequences as gene promoters can allow a gene to be expressed in new cell types or contexts and can generate truncated or extended protein products, potentially allowing host genes to acquire new functions [100].

Indeed, LTR sequences frequently act as promoters for host genes [117]. ERVs are also able to drive host gene expression in differentiating somatic tissues. Variable epigenetic silencing of ERV-derived alternative promoters in somatic tissues can also contribute to the regulation of host genes. Thus, mechanisms that regulate ERV expression are able to impact ERV promoter-driven host gene expression and the

development of somatic tissues.

SINE and LINE-1 TEs can also act as alternative promoters to drive the expression of host genes.

The exaptation of TEs as alternative promoters thus appears to be a relatively simple way to alter the pattern or level of expression of host genes during development.

In addition to acting as promoters that drive the expression of alternative isoforms of host genes, transcription factor-binding sites within TEs can act as host gene enhancers in specific tissues or developmental contexts [100].

They can also induce premature termination of transcription via the incorporation of TE-derived polyadenylation sites [118]. In addition, inefficient transcriptional elongation through the AT-rich LINE-1 sequence can modulate gene expression levels [119]. TE insertions can also introduce TE-derived splice acceptor or donor sites that alter splicing, generating non-functional or nonsense transcripts [120], or can be incorporated into mRNAs and introduce frame shifts or premature termination codons.

TEs can also have an impact on mammalian development through their proteins becoming domesticated, i.e. performing functions for the host organism. The human genome contains around 50 genes that are probably domesticated TEs [65].

TEs often insert into introns or untranslated regions of genes, and this can sporadically result in the exonization of TEs.

Indeed, exonized TEs can expand the mammalian

transcriptome and proteome but can also be used to fine-tune gene regulation [121]. The accumulation of TE-derived sequences in cellular mRNAs can lead to their differential regulation due to TE control mechanisms targeting these TE portions or structures [115,122]. Notably, several classes of TEs have been found inserted within mammalian RNAs [121,123,124] and it is likely that their presence impacts gene regulation and function by providing or interfering with regulatory elements in those RNAs.

Another way that TEs diversify the mammalian transcriptome is by generating long non-coding RNAs (lncRNAs) from their promoters. Remarkably, almost two-thirds of all known human lncRNAs contain TE fragments in their sequences [125,126]. LncRNAs are more abundant than known genes and are involved in multiple biological processes including gene regulation, maintenance of nuclear architecture and splicing [127-129]. Intriguingly unexpected roles for multiple human TE-derived lncRNAs in regulating pluripotency have also recently been described [130-132]. Multiple mechanisms may be involved, and some of these TE-derived lncRNAs represent transcripts originating from ERV elements acting as enhancers or promoters in these cells [130]. However, lncRNAs derived from HERV-H ERVs in human ESCs appear to act *in trans* via physical association with chromatin modifiers [131-133].

TEs can also influence developmental processes *in trans* through the LINE-1-dependent generation of processed

pseudogenes [134]. Although LINE-1-encoded proteins tend to bind to their encoding mRNA *in cis* [134,135], they occasionally bind to cellular host mRNAs *in trans* and catalyze their insertion into the genome as processed pseudogenes. Although most inserted processed pseudogenes lack a functional promoter upon insertion, a promoter can evolve, be captured by a new TE insertion, or be generated by recombination, subsequent diversification of key developmental regulators [100].

Finally, ongoing LINE-1 retrotransposition itself can generate new genes by a mechanism termed exon shuffling [136]. Exon shuffling occurs when an active LINE-1 within a gene retrotransposes to a new genomic location and delivers nearby coding sequences to the new locus. Indeed, LINE-1-mediated exon shuffling has probably increased the repertoire of the human proteome but, as a result of frequent 5′ truncation during retrotransposition [137,77], its overall contribution to the human genome remains elusive.

In addition to their more direct roles in regulating host gene expression, TEs can influence the organization of mammalian chromosomes [100].

Due to their nature, mobile elements have the potential to affect common diseases, through structural variation, deregulated transcriptional activity or epigenetic effects [66]. However, it should be noted that many of these mechanisms can also potentially confer new properties and functions to a host gene rather than simply inactivate it [100].

These data imply that there has been significant activity of TEs during evolution but that most of the genetic changes caused by TEs are not detrimental and moreover indicate their inherent potential to create and diversify biological processes, as proposed 60 years ago by McClintock, Britten and Davidson [138,139]. The proposal that TEs have a present-day function in host genomes to provide *cis*-regulatory elements that co-ordinate the expression of groups of genes [139].

Furthermore, recent findings showing that TEs mobilize much more frequently in development than previously anticipated suggest that these sequences may have additional present-day functions in host genomes (recently reviewed in [92,106,93].

Repeat type				Estimated number of copies	Average length	Mobility	Estimated % genome coverage	
Interspersed	Retrotransposons	LTR	LTR (Long terminal repeat) or ERV (Endogenous retroviruses) (MaLR, ERV, ERV1, ERV-K, ERV-L, etc.)	200,000	6–11 kb	Autonomous retrotransposition (retroviral-like)	8%	42%
		Non-LTR	LINE (Long interspersed element) (L1, L2, CR1, etc.)	500,000	6 kb	Autonomous retrotransposition	20%	
			SINE (Short interspersed element) (Alu, MIR, etc.)	1,000,000	0.3 kb	L-1 dependent Retrotransposition	13%	
			SVA SINE-R/VNTR/Alu	2700	2–5 kb	L-1 dependent Retrotransposition		
	DNA transposons		DNA transposons (MER1, MER2, Mariner, Merlin, etc.)	300,000	1–3 kb	inert	2–3%	

Number of copies and genome coverage are estimated values based on current genome coverage.

Table 1. Major features of the most represented interspersed repetitive elements in the human genome. [66]

1.2.2 Tandem repeats

Tandem repeats constitute a large portion of the human genome, and account for a significant amount of its copy number variation. Although historically relegated as "junk DNA", tandem repeats have taken on a new importance with the realization that their tandem organization provides potentially unique functional characteristics.

Tandemly repeated DNA is organized as multiple copies, arranged in a head to tail pattern to form tandem arrays, and thus represents a distinct type of sequence organization shared by all sequenced genomes [140].

Tandem DNA in the human genome shows a wide range of unit sizes, ranging from microsatellites of a few base pairs to megasatellites of up to several kb [141,142,140] (Table 2).

One of the principal families of DNA tandem repeats in the genome is represented by the alpha satellite DNA of chromosome centromeres [143,144], that are critically important for establishing heterochromatin formation and proper chromosome segregation during cell division (reviewed in [145]). Indeed, maintenance of the structural integrity of centromeres and telomeres is one of the most important functions of tandem repeats [146]. In the human genome, alpha satellite DNA repeat unit consist of 171 bp monomers, which are found in large highly homologous arrays of up to several million bp at the centromeres of all human chromosomes.

These tandem arrays are composed of either diverged monomers with no detectable higher-order structure, or as chromosome-specific higher order repeat units (HORs) characterized by distinct repeating linear arrangements of an integral set of 171 bp monomers [147]. This HOR structures are typical of this type of repeat, as they are important for centromere function [144].

The most abundant tandem repeats after alpha satellite DNA are satellites II and III, localized in the pericentromeric regions of human chromosomes 3, 4, 9, 13, 14, 15, 21, and 22 [148] [149].

Satellite III is composed primarily of the pentameric sequence GAATGn (or CATTCn). This family could have many large arrays at 5 bp and multiples thereof including some repeat units as large as 70 bp, forming arrays of up to ~100 kb [140].

Satellite II is based on highly diverged arrays of GAATG, with 23 bp or 26 bp repeat units and approximate multiples which are identified as HsatII [150].

Moreover there is the beta satellite tandem repeat located on chromosome 9 and on the acrocentric chromosomes (13, 14, 15, 21, and 22). This DNA repeat exists as tandem arrays of diverged approximately equal to 68-base-pair monomer repeat units. The monomer units are organized as distinct subsets, each characterized by a multimeric higher-order repeat unit that is tandemly reiterated and represents a recent unit of amplification [151].

Another abundant type of satellite DNA is the gamma-satellite DNA, a tandem array of 220-bp GC-rich repeating units, usually forming 10- to 200-kb clusters flanked by alpha-satellite DNA (e.g., at 8q11.1) [152]. Gamma-satellite DNA has been identified in the pericentromeric regions of human chromosomes 8, X, and Y [152-154,149].

However, in the human genome, the main groups of tandem repeats are the micro-, mini- or macro-satellites [140]. At a given locus, they are highly polymorphic among individuals and for this reason they are more commonly known as variable number tandem repeats (VNTRs).

Microsatellite DNA or short tandem repeat (STR) are constitute of unit very short in length, approximately 1-13 bp. They present an high level of polymorphism and often used as genetic markers in genetic population studies and also for individual identification.

Minisatellite DNA is represented by monomers variable in length from 6 to 100 bp. One of the most common repeat of this category is the esameric sequence TTAGGG, that is repeated thousands of times at chromosomes extremities, constituting telomeres [66].

Macrosatellites consist of arrays of 2–12kb repeat units, with a number of repeats ranging from a few to over one hundred [140]. They can be either chromosome specific, as DXZ4 at chromosome Xq23 [155] and ZAV at chromosome 9q32 [156] or they can be associated with two or more chromosomal

locations, such as D4Z4, on chromosomes 4q35 and 10q26 [157,158]; and RS447, on 4p15 and 18p23 [159].

They are extremely polymorphic and have recently attracted the attention because of their potential structural and regulatory role in the human genome [66].

Besides their role in evolution [160,143,161], they have been found to be critical in several other processes, including heterochromatin formation, chromosome segregation, [145] and X- chromosome inactivation (XCI) [75]. Moreover, repeat instability is at the basis of a number of diseases [162].

Moreover they could play a role in regulating cellular proliferation through telomeres, that are also involved in senescence mechanisms and DNA damage response. Shorten of telomeres is observed with increasing age in proliferating human tissues [163]. More over telomeric complex allows cells to distinguish random DNA breaks and natural chromosome ends. Whereas broken chromosomes activate DNA damage checkpoints and are repaired, telomeres are not detected as DNA ends. [164-166].

Contraction of a 3.3 kb polymorphic tandem repeat array on chromosome band 4q35 is associated with facioscapulohumeral muscular dystrophy (FSHD) [167-169]. And the large macrosatellite array DXZ4 appears to have a unique function in the process of X chromosome inactivation [75]. Thus, tandem repeats play important functional and evolutionary roles in genome biology.

They are among the most variable loci, experiencing mutations in the number of repeat units that are 100 to 100,000 times more frequent than point mutations [170-173].

In eukaryotes, Tandem Repeats (TRs) located in coding regions and their promoters tend to occur in genes associated with transcriptional regulation, DNA binding, protein-protein binding, and developmental processes [174,175], suggesting a regulatory role for TRs.

In fact, TRs are emerging as good candidates for a type of genomic variation that can directly alter gene expression [176] [177,174,175]. Because gene expression changes might contribute to the fundamental differences between humans and other species [178], it is imperative to study mechanisms that may permit rapid expression changes on short evolutionary time scales [179-182].

Thus, the high incidence of TRs in regulatory regions in some species [175,183] and their high mutability [170-173], suggests that it may be important to study TR variation to understand fundamental differences in gene expression across species and populations. In particular, since TRs constitute 3% of the human genome [65] and are dramatically enriched in promoter regions [174,184], clarifying their functional role may provide important insights for the human biology field [185].

Repeat type			Unit length	Array length	Estimated % genome coverage
Tandem	Satellite	Alpha-Satellite	171 bp	3–5 Mb	22–25%
		Satellite II (HsatII)	23–26 bp or multiple	10–70 kb	
		Satellite III (GAATGn- simple sequence)	5 bp or multiples up to 70 bp	75–100 kb	
		Beta-Satellite	68 bp	2–14.5 kb	
		Gamma-Satellite	220 bp	10–200 kb	
	VNTR (Variable number of tandem repeats)	Microsatellite (Short tandem repeat)	1–13 bp	Hundreds bp	
		Minisatellite (including telomeric repeats)	6–100 bp	1–15 kb or more	
		Macrosatellite	2–12 kb or more	Tens up to hundreds bp	

Genome coverage is estimated on the basis of current genome coverage.

Table 2. Major features of the most represented tamdem repeats in the human genome. [66]

1.2.2.1 D4Z4 macrosatellite

The D4Z4 macrosatellite maps to the subtelomeric region of the chromosome 4 long arm, in 4q35. Each unit is 3.3kb and contain an ORF sequence encoding an homeobox transcription factor, DUX4, and two repetitive elements: LSau, a middle repetitive elemnt associated with heterochromatic regions of the genome [186], and hhspm3, a low-copy GC-rich repeat [187,188] (Figure 2).

This monomer is present in 11 to 100–150 copies in the general population. Interestingly, reduction of D4Z4 copy number below 11 units is associated with FSHD, one of the most important forms of muscular dystrophy [189] [190]. The D4Z4 repeat is not restricted to chromosome 4; D4Z4 belongs to a family of repeats with high sequence identity present also in human chromosomes 10q26, 1p12, and the p-arm of acrocentric chromosomes [188,158]. In particular perfect arrays of D4Z4 units can be detected on chromosome 10q and this results in frequent exchanges between the 4q35 and 10q26 arrays, which share the highest identity [190]. Instead the others localization of D4Z4 are interspersed with beta satellites on many heterochromatic loci [158,191].

Are reported striking differences between 4q/10q D4Z4 and the non-4q/10q D4Z4 homologs on other chromosomes.

Non-4q/10q D4Z4 sequences are highly divergent and variable compared to 4q/10qD4Z4.

The homologs on chromosomes 13, 14, 15, 21, and 22 contain multiple sequences with different patterns and degrees of polymorphism, suggesting that the majority of non-4q/10q D4Z4 homologs are susceptible to frequent mutations.

In particular approximately 80% of the variation in D4Z4 homologs in comparison to the 4q D4Z4 sequence are due to nucleotide polymorphisms with the conversion of C/G into A/T.

More over the repeat exchanges had occurred at much higher frequency among non-4q/10q D4Z4 homologs than between these homologs and 4q/10q D4Z4 [192].

Finally 4q and 10q D4Z4 repeats are unusually conserved compared to non-4q/10q D4Z4 homologs on other chromosomes. In particular the 4q and 10q D4Z4 repeats are highly uniform, at least in the region encompassing the N-terminus of the DUX4 ORF [193]. In contrast, the corresponding regions in the non-4q/10q homologs are highly heterogeneous with frequent interruptions of the DUX4 ORF. The observed hyper variability of the non-4q/10q D4Z4 homologs resulted in alterations of the corresponding amino acid sequence of the putative DUX4 ORF within the sequenced region, frequently introducing a nonsense codon [192].

Genomic analysis of the gorilla locus syntenic to human chromosome 4q35.2 revealed clear differences in organization in comparison with humans. A basic block of short D4Z4 arrays (10–15 repeats) and 4q35.2-specific sequences (i.e. the FRG2 gene and the 13E11 marker) spaced by LINE is repeated at

least twice in gorilla, whereas chimpanzee has a single D4Z4 array linked to a

LINE block, and humans show further remodeling involving the loss of the LINE block. The 4q subtelomere therefore underwent substantial genomic changes during evolution, and its remodelling essentially involved D4Z4 and LINE sequences [191].

This was a recent evolutionary event, whereas the dispersal of the D4Z4 array from the 4q ancestral locus of the macaque has an older evolutionary history, as it took place in an ancestor of the orangutan and preferentially involved acrocentric chromosomes in both chimpanzee and man [194,158,195].

Furthermore, 1qcen, 10qcen and 10qter are human-specific sequence dispersals of D4Z4 repeats.

Both sequence types are prone to unequal homologous recombinations leading to the expansion/contraction of the 4q subtelomere, and this mechanism could account for the expansion of the D4Z4 array and the complete loss of the LINE block in man [191].

In gorilla and chimpanzee cells low acetylation levels were detected indicating the occurrence of a more condensed chromatin structure at 4qter, and this correlates well with their very low expression of FRG1 and FRG2 genes [191]. Within the gorilla and chimpanzee locus syntenic to human chromosome 4q35.2, LINE may represent candidate sequences for chromatin condensation as the silencing of gene expression has been

associated with an epigenetic mechanism mediated by long and short interspersed repeats [196].

Previous studies showed a considerable evolutionary conservation of the nuclear positioning of 4q subtelomere. It is consistently localised at the nuclear periphery in chimpanzee and gorilla fibroblast nuclei as well in human fibroblast nuclei [197-199].

Conversely, although maintaining the preferential association of the 4q35.2 locus with the nuclear periphery, human cells show less chromatin condensation and higher 4q35.2 gene expression [191].

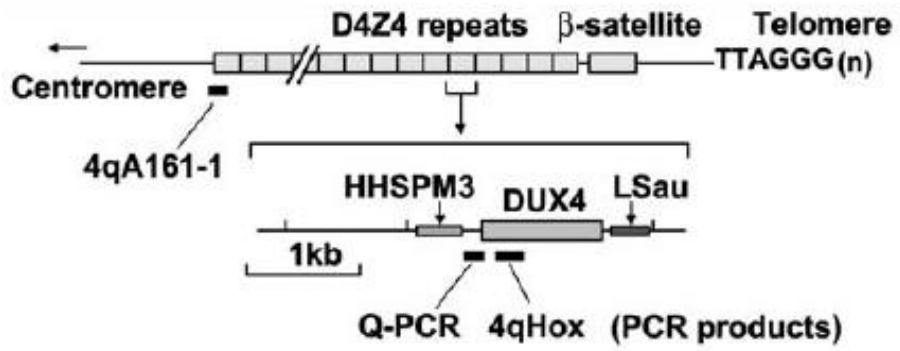


Figure 2. D4Z4 macrosatellite unit composition.

A schematic diagram of the 4qter D4Z4 repeat region and a single D4Z4 repeat. The DUX4 ORF, a GC-rich sequence homologous to the lowcopy repeat HHSPM3 and a middle repetitive element associated with heterochromatic regions of the genome LSau, are shown [193].

1.3 Facioscapulohumeral muscular dystrophy as a complex epigenetic disease

Facioscapulohumeral muscular dystrophy (FSHD) is the most common myopathy found in adults, with an overall incidence of more than 1:14.000. FSHD is a dominant autosomal myopathy that is characterized by progressive, often asymmetric weakness and wasting of facial (facio), shoulder, and upper arm (scapulohumeral) muscles [200,8], but progressing to affect almost all skeletal muscles [201,202]. It is classified among progressive muscular dystrophies, characterized by muscular fiber necrosis and degeneration giving rise to progressive muscular atrophy.

FSHD is a genetically heterogeneous disorder and its genetic bases are unique and involve both genetic and epigenetic alterations [203]. Monozygotic twins with different penetrance of FSHD have been described, suggesting a strong epigenetic contribution to the pathology [8,204,205].

The vast majority of FSHD, indeed, are transmitted as an autosomal dominant trait with the disease locus mapping to the subtelomeric region of chromosome 4q, more specifically at 4q35-qter [206-208]. This form has been termed FSHD1 and accounts for approximately 95% [209]. This chromosomal region lacked classical genes, but contains an array of repeated units of a 3.3 kb macrosatellite repeat ordered head to tail [210], called D4Z4. *De novo* cases accounting for around 25% of

patients [211,212,203]. In approximately half of the *de novo* cases, the contractions occurs somatically, most probably by an intrachromosomal gene conversion mechanism with or without cross over, and leading to somatic mosaicism for the disease allele [202,213].

However some families, with an undistinguishable clinical phenotype display a more complex pattern of inheritance and a distinct genetic defect. This second form is termed FSHD2 [203] and represents about 5% of patients meeting clinical criteria for FSHD [209,214]. This type of FSHD is caused by mutations in SMCHD1, a member of the condensin/cohesin family of chromatin compaction complexes, that binds to the D4Z4 repeat array. In these FSHD2 families, the disease shows a more complex digenic inheritance because the mutation of SMCHD1 on chromosome 18 segregates independently from the FSHD-permissive chromosome 4q haplotype [214]. Interestingly, FSHD2 patients, who phenotypically show FSHD though lacking D4Z4 contractions, display general D4Z4 hypomethylation [215], indicating an important epigenetic condition necessary to develop or generate the disease.

1.3.1 Clinics of FSHD

1.3.1.1 Symptomatology

Disease onset is usually before the second decade [203] and initially affects muscles of the face, shoulder girdle, and upper

arms (mimetic muscles, serratus anterior and rhomboid muscles, and biceps and triceps). Marked involvement of the abdominal and paraspinal musculature can occur, which can cause exaggerated lumbar lordosis or camptocormia (bent spine syndrome). The disease then progresses to involve the lower extremities, typically the distal musculature first (tibialis anterior and gastrocnemius) in a facioscapuloperoneal pattern, then later involving more proximal muscles (quadriceps and hamstrings). [209] In the most severe cases (patients with a more severe infantile presentation and often only 1 to 3 residual D4Z4 units on genetic testing), the muscular degeneration can extend to the pelvic girdle and foot dorsiflexor muscles, thereby affecting the ability of the patient to walk, just in the second decade. However the risk increases with the age and almost 20% of FSHD patients become wheelchair bound after 50 years of age [216].

Restrictive respiratory involvement varies with a range of 0% to 13% reported [217] and principally is caused by loss of core/trunk musculature as opposed to diaphragmatic involvement. In FSHD the respiratory complications typically follow the weakness and are more likely in patients with pelvic girdle weakness who are wheelchair bound or who have prominent paraspinal involvement or kyphoscoliosis.

Largely asymptomatic supraventricular arrhythmias can be found in approximately 5% to 10% of patients [218,219].

Although FSHD is considered as a skeletal muscle specific disease, extra muscular involvement has been reported. FSHD is associated with retinal vasculopathy, a blood vessel disorder of the retina, in 60% of cases [220,200,221,222] and sensorineural hearing loss in 75% of affected individuals [222-225]. Mental retardation and epilepsy are also present in FSHD patients more frequently than in healthy people [219,225-230]

Most FSHD patients report their first symptoms during the second or third decade of their life; however, the age of onset can vary from infancy to age 50 [167,230].

Earlier onset is associated with more rapid progression and higher severity [231].

One of the most intriguing clinical observations in FSHD is the marked inter- and intrafamilial variability in disease onset, progression and in the clinical presentation [232]. While one-fifth of FSHD1 patients will become wheelchair dependent, an equal proportion of FSHD1 mutation carriers remain asymptomatic through their lives [203,202].

Interestingly, the FSHD phenotype is gender dependent. Typically, males are more severely affected, whereas females can develop a milder or asymptomatic form of the disease [233,234] and display later onset and a slower progression.

The main characteristics of muscular involvement in FSHD are asymmetry, selectivity, and the progression profile. Muscle impairment in FSHD is often asymmetric: muscles on one side

of the body appear much more compromised than on the other side.

FSHD2 patients do not seem to display a significantly different clinical phenotype compared to FSHD1, in terms of the pattern of muscular weakness and the frequency of extra muscular involvement except for retinal telangectasia that has never been described in these patients [235].

To date, no early onset FSHD2 patient has been reported in literature, besides a patient carrying both FSHD1 and FSHD2 who began to develop symptoms in early childhood [236].

1.3.1.2 Diagnosis

The clinical diagnosis of FSHD is confirmed by molecular analysis, bypassing the need for muscle biopsy in most cases. Standard molecular testing for FSHD1 demonstrates the presence of a contraction of the D4Z4 repeated units in one copy of 4q35.

The molecular diagnosis of FSHD1 is performed by a classical technique, consisting of a Southern blot [29,33] using the p13E-11 probe, on linear gel electrophoresis (LGE), which recognizes a region proximal to the D4Z4 locus on chromosome 4, after a double digestion of the DNA with EcoRI and BlnI enzymes and yields a sensitivity of 93% and specificity of 98% [217]. Normal individuals usually have fragments of more than 38 kb, while

patients with FSHD1 have fragments between 10 kb and 38 kb (corresponding to an estimated 1 to 10 residual D4Z4 units).

The use of pulse field gel electrophoresis (PFGE) is highly recommended, as well as 4qA and 4qB determination. In dubious cases, it is important to search for rearrangements between chromosomes 10 and 4 and proximal deletions including the p13E-11 probe region by using specific protocols, including additional restriction enzyme digestion combined or not with additional probes [203].

If these analysis result negative (ie, no contraction of D4Z4 repeats) then determination of the presence of at least one A allele and very low methylation (less than 20%) confirms a diagnosis of FSHD2 [209].

1.3.2 FSHD2 form of disease

There is a less common form of FSHD, and these individuals diagnosed with FSHD2 (i.e. about 5% of the FSHD patients) display clinical features that are identical to those observed in FSHD1 patients, but present different molecular defect [235].

The apparent mode of inheritance is variable; while the majority of cases are apparently sporadic, dominant and recessive patterns have also been described [235,237].

FSHD2 patients carry smaller but normally sized D4Z4 repeat arrays (i.e. 8–20 units) [235], but show a strong reduction of D4Z4 DNA methylation on 4q35, suggesting impaired

epigenetic silencing of the repeat units, similar to the contracted D4Z4 repeat arrays in FSHD1 [214,202] (Figure 3). Therefore FSHD1 and FSHD2 show similar epigenetic changes at D4Z4 and share a common disease pathway [238].

FSHD2 patients carry at least one somatic DUX4 polyadenylation signal-containing chromosome 4, and, like FSHD1 patients, display a variegated DUX4 expression pattern in muscle cell cultures [239].

In FSHD2 patients were also identified mutations in the Structural Maintenance of Chromosomes Hinge Domain Containing 1 (SMCHD1) gene [214], on chromosome 18. Several studies have since identified disease-causing variants in SMCHD1 that include splice site, insertion-deletion, missense and nonsense variants [240,236,241-245]. SMCHD1 mutations are now presumed to account for 85% of FSHD2 cases [246].

Therefore FSHD2 show a digenic inheritance: haploinsufficiency of SMCHD1 and the 4qA haplotype [202,214]. SMCHD1 belongs to the ubiquitous SMC gene superfamily that form essential components of the cohesin/condensin protein complexes. In muscle cells SMCHD1 has been reported to directly bind to the D4Z4 repeat array, suggesting that it is required to maintain a repressive chromatin structure in somatic cells [214,247-249].

In addition to its involvement in silencing of the D4Z4 repeat array, in mammals SMCHD1 has been implicated in several alternative silencing pathways. These include X-chromosome

inactivation, silencing of monoallelicly-expressed genes, and telomere silencing suggesting an important role for components of the repeat silencing machinery in FSHD [250-253].

However, the precise molecular mechanism of SMCHD1 function at D4Z4 remains to be determined.

Intriguingly, SMCHD1 has recently also been identified as a modifier of disease severity, when mutated, in a growing number of FSHD families with moderately contracted D4Z4 alleles (7–10 D4Z4 units) [243]. These patients are considered FSHD1 + 2 patients [236].

1.3.3 Molecular basis of the pathology

1.3.3.1 D4Z4 array contraction on 4q35

Genetically, the disease is not caused by classical mutations in a protein coding gene. Rather, it is associated with reduction in the copy number of the 3.3 kb macrosatellite D4Z4 repeat, mapping to the sub-telomeric region of human chromosome 4 long arm (4q35) on a specific FSHD-permissive haplotype of chromosome 4q [254].

In the general population, this repeat array is extremely polymorphic ranging from 11 to 150 copies [255,256]. Most patients with FSHD present a partial deletion of the D4Z4 array, which leaves 1–10 units on the affected allele [190] (Figure 3).

In addition to polymorphism associated with D4Z4 repeat number, a recent study on single-nucleotide polymorphisms

localized on the D4Z4 locus allowed the identification of nine different haplotypes [257] in the general population. Only a few of them, all 4qA types, are associated with the disease. These variants differ for the presence of a β satellite repeat immediately distal to the D4Z4 array on 4qA allele [254,258]. The most frequent among these “permissive” haplotypes is 4A161 [259-263].

D4Z4 repeat arrays are not restricted to chromosome 4q, but homologous sequences have been identified on many chromosomes [188].

Due to a duplication event, a highly similar and equally polymorphic repeat array localizes to the subtelomere of chromosome 10q [157]. In particular, the subtelomere of chromosome 10q is almost identical to the region in 4q containing D4Z4 repeats, containing highly homologous and equally polymorphic repeat arrays [264,157] and additionally extending to 45 kb proximal of D4Z4 array and 15–25 kb distal [265].

However, chromosomes 10 with less than 11 repeat units does not cause FSHD1 [266], suggesting that the chromatin environment associated with chromosome 4q could contribute to FSHD development.

Although a linear negative correlation between repeat size, age of onset and clinical severity has not been observed, some findings indicated that patients carrying the lower number of

repeats result in earlier disease onset and enhanced severity [267-271].

Population studies have indicated that sizes between 8 and 11 units represent a grey zone where, depending on the epigenetic state of the D4Z4 repeat array, disease penetrance is incomplete [260,272,240] and those carriers may be paucisymptomatic or asymptomatic.

Interestingly, at least one D4Z4 unit is necessary to develop FSHD, as monosomy of 4q does not cause the disease [273], suggesting a gain of function effect

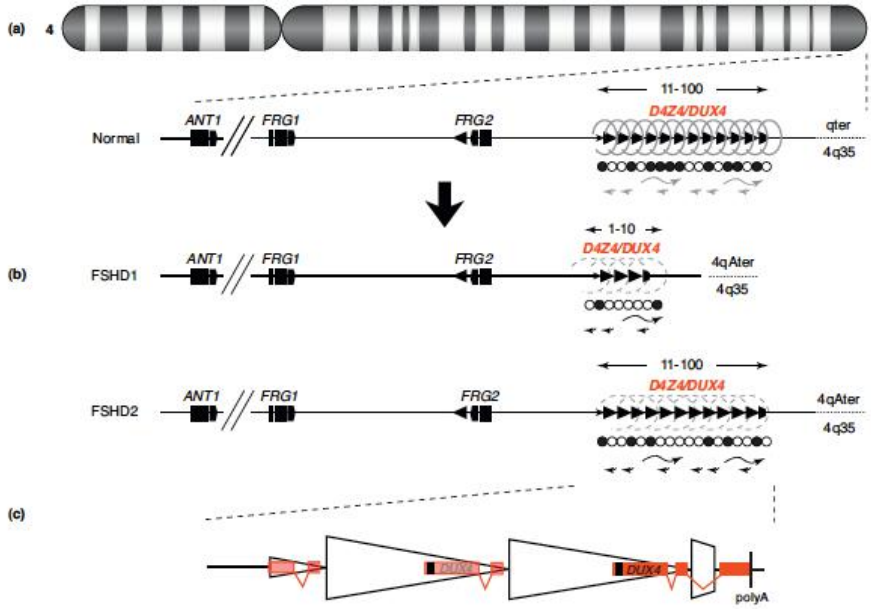


Figure 3. Schematic of the FSHD locus.

(a) The D4Z4 repeat (triangles) is located in the subtelomere of chromosome 4q and can vary between 11 and 100 copies in the unaffected population. This repeat structure has a closed chromatin structure characterized by heterochromatic histone modifications (dense springs), high DNA methylation levels (closed circles) and complex bidirectional transcriptional activity (gray arrows). Candidate genes DUX4, FRG2, FRG1 and ANT1 are indicated. (b) In patients with FSHD, the chromatin structure of D4Z4 adopts a more open configuration (open springs and open circles) leading to inefficient transcriptional repression (black arrows) of the D4Z4 repeat. (c) The DUX4 gene is located within each D4Z4 unit. On permissive chromosomes, the last copy of the DUX4 genes splices to a third exon located in the region immediately flanking the repeat and stabilizing the transcript owing to the presence of a polyadenylation (polyA) signal [238].

1.3.3.2 Changes in chromatin structure at the FSHD locus

There is a general consensus in the field in supporting the view that epigenetic mechanisms are important players in FSHD. Several FSHD clinical features, such as the variability in severity and rate of progression, the gender bias in penetrance, the asymmetric muscle wasting and the discordance of the disease in monozygotic twins, strongly suggest the involvement of epigenetic factors [63,230,238]

Increasing evidence suggested that, in patients, chromatin conformation of FSHD locus is altered at multiple levels, from DNA methylation, through histone modifications up to higher-order chromosome structures, resulting in perturbation of heterochromatic gene silencing in the 4q35.

D4Z4 units have a high content of highly methylated GC-dinucleotides, having characteristics of CpG islands [255], as well as two dispersed repeat elements (LSau and hhspm3), characteristic of heterochromatic, nonexpressed regions of the human genome [274]. Notably, the area occupied by D4Z4 repeats in healthy subjects represents one of the largest GC-rich regions of the human genome. [275].

At sizes >10 units the array adopts a repressed chromatin structure in somatic cells as evidenced by high levels of CpG methylation and the presence of repressive histone modifications [193,276]. D4Z4 contractions to a size between 1 and 10 units are associated with partial relaxation of the D4Z4

chromatin structure in somatic cells (mostly determined by D4Z4 hypomethylation) [215,277]. Interestingly, FSHD2 patients, who phenotypically show FSHD though lacking D4Z4 contractions, display general D4Z4 hypomethylation [215], indicating that this is an important epigenetic condition necessary to develop or generate the disease.

The D4Z4 repeat array is enriched of two repressive marks: trimethylation of lysine 9 or 27 of histone H3 (H3K9me3 and H3K27me3, resp.). The first, generally associated with constitutive heterochromatin, is deposited by the histone methyltransferase SUV39 and is responsible for HP1 repressor recruitment [193]. H3K27me3 is characteristic of facultative heterochromatin, is deposited by the PRC2 subunit EZH2, and in turn recruits PRC1 and PRC2 to establish transcriptionally repressed domains.

Indeed, D4Z4 repeat is indicated as a novel Polycomb target [278] able to initiate de novo PcG recruitment to ectopic sites and mediate copy number-dependent repression of gene expression, typical features of *Drosophila* PREs [279,278]. In fact each D4Z4 unit is extremely GC rich, containing a sequence nearly identical to the consensus motif of *Drosophila* PREs and several putative DNA binding sites for factors which are Polycomb recruiters in *Drosophila*, such as YY1 and GAGA factor [280-282,279,278]. The region is also bound by proteins associated to Polycomb recruitment in mammals like Jarid2 [283-287] or homologs of PcG recruiters in *Drosophila* (YY1,

HMGB2, c-Krox/Th-POK; vertebrate fly homologs Pho, Dsp1, GAGA factor, respectively) [281,282,279,288,289]. Finally, the repeats array also shows enrichment for the Polycomb-associated histone variant macroH2A [290].

In addition to the complex heterochromatic features found at D4Z4 locus, there has been shown the presence of euchromatin histone marks associated with transcriptional activation in the first proximal D4Z4 unit of the array, such as acetylation of histone H4 and dimethylation of Lys 4 of histone H3 [291,193]. This could reflect the complexity of bidirectional transcriptional activity at the locus and could suggest the potential presence of noncoding RNA that further regulate the transcription.

The mechanism proposed is that in healthy subjects, the presence of many D4Z4 units results in extensive PcG binding, DNA methylation, histone de-acetylation and chromatin compaction leading to a repressive chromatin organization.

In FSHD patients, reduction of D4Z4 copy number results in a critical reduction of PcG binding and a reduced spreading of the PcG histone mark H3K27me3 on the contracted 4q35 allele [278]. Interestingly, also H3K9me3 is lost in FSHD patients, preventing the binding of D4Z4 to the heterochromatin-binding protein HP1γ and the sister chromatid cohesion complex, cohesin. Overall the reduction of both these two repressive histone marks lead to the de-repression of 4q35 locus [230,193].

Moreover this creates the epigenetic environment permissive for the transcription of an activatory, long ncRNA, called DBE-T, originating proximally to and covering part of the repeat array. DBE-T is produced solely in FSHD patients and is required for opening up the 4q35 chromatin and for de-repression of 4q35 genes. Mechanistically, we discovered that DBE-T is a chromatin-associated RNA that functions *in cis*, by directly recruiting and binding the TrxG protein ASH1L to the 4q35 locus. Ash1L recruitment is associated with the accumulation of H3K36me2 at the FSHD locus. This leads to a structural and epigenetic remodeling of the FSHD locus, toward a more active chromatin state, which is responsible for the de-repression of 4q35 genes [275,66,278] (Figure 4).

However, D4Z4 array has a complex transcriptional profile that includes sense and antisense transcripts and RNA processing [292].

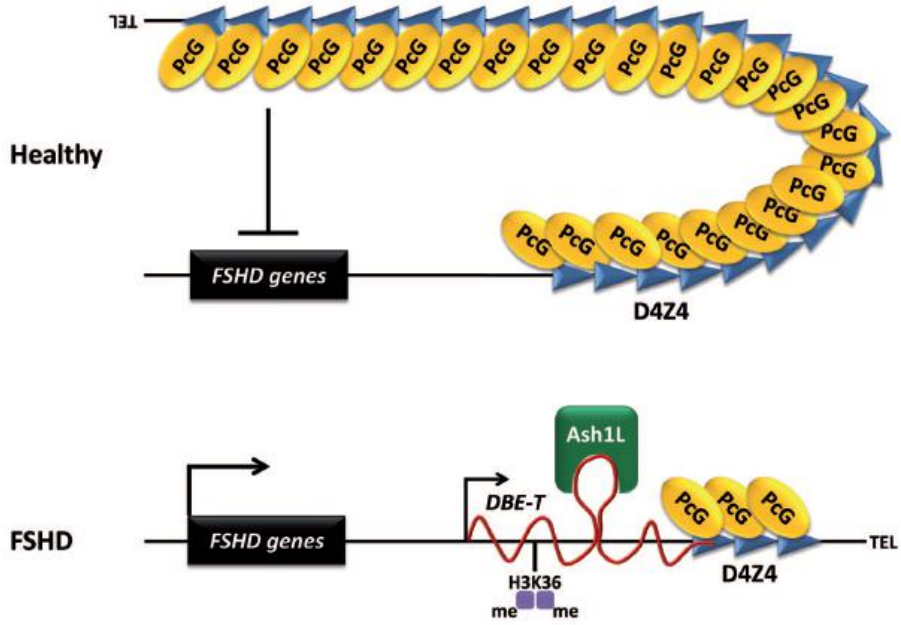


Figure 4. Model of DBE-T mediated de-repression of 4q35 genes in FSHD.

In healthy subjects, the D4Z4 repeat array carrying many units displays extensive binding of PcG proteins leading to repression of the 4q35 locus. In FSHD patients, reduction in D4Z4 copy number critically diminishes PcG binding and silencing, allowing for transcription of the lncRNA DBE-T to occur. DBE-T functions in cis to promote opening of chromatin structure and de-repression of 4q35 genes through direct binding and recruitment of the TrxG protein Ash1L, which drives H3K36me2 at the FSHD locus [275].

1.3.3.3 Nuclear localization of 4q subtelomere

Finally we have to consider that epigenetic chromatin regulation depends also on appropriate intranuclear positioning. Indeed regulatory proteins are spatially clustered in specific territories and the position of chromosomal region in the nucleus influences its transcriptional activity. The 4q subtelomere is preferentially localized in the nuclear periphery in both controls and FSHD patients [198,197], and this localization is also evolutionary conserved [191]. Moreover it has been shown that the nuclear periphery localization in controls and FSHD1 cells is directed by different sequences, proximal or within D4Z4 repeat, respectively.

More deeply a sequence 215 kb proximal to the repeat array shows a stronger localization to the nuclear rim than D4Z4 in healthy subjects, suggesting that a region proximal to D4Z4, and not the repeat array itself, directs the 4q telomere to the periphery [198]. Moreover an 80-bp sequence is identified within D4Z4 unit that tethers the subtelomere in the nuclear periphery in a CTCF and Lamin-A-dependent manner [293]. This property is lost upon D4Z4 multimerization.

Thus, it appears that in healthy subjects, multiple copies of D4Z4 are located near the nuclear periphery due to a 4q-specific signal proximal to D4Z4, whereas in FSHD patients the perinuclear location is mediated by D4Z4 [293].

CTCF is a multifunctional DNA-binding protein that is important for transcriptional regulation, chromatin insulation, and chromatin organization [294]. It is also reported that the same 80-bp D4Z4 element mediating the perinuclear positioning of the 4q telomere is also responsible for the CTCF and A-type lamin-dependent transcriptional insulator function of the repeat. The CTCF binding and insulation activity are lost upon multimerization of the repeats. As such, it has been proposed that FSHD patients have a CTCF gain-of-function phenotype that “protects” certain genes from the influence of nearby repressive chromatin, ultimately generating a 4q35 de-repressed state [295].

1.3.3.4 FSHD candidate genes

The 4q35 locus is a relatively gene-poor region [296,297]. However expression of several candidate genes proximal to 4q35 potentially contributing to the FSHD pathogenic pathway were studied. Among them, the FSHD region gene 1 (FRG1), FSHD region gene 2 (FRG2) and adenine nucleotide translocator 1 (ANT1) were the most relevant candidates. Each of the aforementioned genes were initially found to be up-regulated in FSHD muscle [279].

FRG1

Some evidences suggest a role for FRG1 in pre-mRNA splicing [298-300]. FRG1 is highly conserved in both vertebrates and invertebrates and it has been found overexpressed in some FSHD samples [279,301]. Moreover, transgenic mice overexpressing FRG1 develop, selectively in the skeletal muscle, pathologies with physiological, histological, ultrastructural, and molecular features that mimic human FSHD [298].

It has been shown that H3K27me3 and the two Polycomb proteins YY1 and EZH2 are bound to D4Z4 and FRG1 promoter in myoblasts [279,301] and are reduced during myogenic differentiation [301].

Interestingly, DNA association studies, by using 3C technologies [51], revealed that D4Z4 physically interacts with FRG1 promoter and this DNA loop is reduced upon differentiation. These epigenetic signatures dynamics during myogenesis are accompanied by a gradual upregulation of FRG1 [301]. Conversely, in FSHD1 myoblasts the D4Z4-FRG1 promoter interaction is reduced and FRG1 expression is anticipated during differentiation, suggesting an alteration of epigenetic signatures dynamics occurring when the differentiation starts.

However, FRG1 overexpression in FSHD samples is not a uniform finding [291,302] and thus the contribution of the FRG1 gene to the FSHD phenotype needs further validation. Finally

the mechanism of action and the role of FRG1 in FSHD onset and development is largely unknown.

DUX4

The most prominent feature of the D4Z4 unit is the presence of an open reading frame for a retrogene called DUX4 [255,303]. DUX4 is believed to have resulted from retrotransposition of a processed transcript of DUXC, a gene present in many mammals but lost in the primate lineage [304]. The DUX4 retrogene encodes a 52 kDa double homeobox transcription factor, which is normally expressed at high levels in human testes and pluripotent stem cells, but epigenetically silenced in somatic cells [239].

A detailed analysis of the 4q subtelomeric region revealed that the DUX4 mRNA is generated by transcription of the last, most telomeric D4Z4 repeat. In fact D4Z4 contains only exons 1 and 2 of DUX4, while the rest of the gene is located on the non-repetitive region distal to the last D4Z4 repeat. This region includes also a sequence termed pLAM that contains a polyadenylation signal, necessary for DUX4 transcript stabilization [168] (Figure 3). Only DUX4 transcripts originating from the last repeat on the 4qA chromosome can be polyadenylated and are stable (and can be translated into protein) while those expressed from 4qB or 10q chromosomes are rapidly degraded [214,203].

Therefore only when the DUX4 transcript can make use of this somatic DUX4 polyadenlyation signal unique to 4qA, the D4Z4 chromatin relaxation, characteristic of both forms of the disease, leads to a higher probability of DUX4 expression in skeletal muscle [305,193,239,259,276].

The relatively higher presence of DUX4 in myotubes versus proliferating myoblasts may suggest that DUX4 transcription is induced upon differentiation [306].

DUX4 binds to a double-homeobox motif and regulates the expression of genes associated with stem cell and germ-line development. Mis-expression of DUX4 in skeletal muscle in FSHD patients induces a large transcriptional deregulation that includes up-regulation of germline genes, genes required in stem cell biology as well as lncRNAs, genes involved in RNA splicing and processing, but also of retrotransposons, endogenous retrovirus elements, and pericentromeric satellite HSATII sequences and suppression of innate immune response genes [307,308,239].

More deeply eventual expression of these developmentally regulated genes in FSHD might be incompatible with the post-mitotic status of skeletal muscle cells, leading to tissue dysfunction.

DUX4 activation disrupts RNA metabolism (i.e. RNA splicing, surveillance and transport) as well as cell signaling, polarity and migration pathways [309,310].

Moreover DUX4 binds to the long terminal repeat (LTR) sequence of a class of retrotransposons and approximately one-third of the genes regulated by DUX4 initiate transcription in these LTR sequences [202]. Some of these elements may create alternative promoters for genes causing production of non-physiological transcripts, long noncoding RNAs, or antisense transcripts. All these factors are likely to be involved in the downstream cascade of intracellular and extracellular events leading to muscle degeneration [203].

In addition, DUX4 expression leads to the inhibition of nonsense-mediated RNA decay [308,311,312] and the accumulation of a large number of RNA transcripts that would normally be de-graded.

The DUX4 mRNA, because of the presence of the translation stop codon in the first exon and the polyadenylation signal in the third exon, represents a target of nonsense mediated decay (NMD) machinery, which might explain why the DUX4 mRNA is present at such low abundance and has been difficult to detect. Interestingly, DUX4 expression results in proteolytic degradation of the factor UPF1, an essential component of the nonsense-mediated decay pathway, resulting in a profound inhibition of NMD [312].

This creates a self-reinforcing loop since DUX4-mediated inhibition of NMD results in increased perdurance of the DUX4 mRNA and this could represent a possible mechanism of positive autoregulation causing FSHD pathogenesis [312].

Moreover DUX4 expression also resulted in the increased abundance of many coding RNA isoforms created by alternative splicing containing premature translation termination codons upstream of splice junctions [292] that are predicted substrates for degradation by NMD. The following accumulation of stabilized NMD substrates may cause direct or indirect toxic effects in muscle cells due to intrinsic toxicity of abnormal RNAs, or a stress response to the production of abnormal proteins with antigenic potential, contributing to an immune response [313].

We could summarize that DUX4 expression in FSHD skeletal muscle cells could have a proapoptotic function [314] and toxic effect on cell growth [315], resulting at the end in cell death and atrophic myotube formation [316,317,314].

Finally the generation of transgenic mice carrying D4Z4 arrays from an FSHD1 allele and from a control allele recapitulates several genetic and epigenetic features seen in FSHD patients: high DUX4 expression levels in the germline, (incomplete) epigenetic repression in somatic tissue, and FSHD-specific variegated DUX4 expression in sporadic muscle nuclei associated with D4Z4 chromatin relaxation [318].

However extremely low levels of DUX4 were found in FSHD muscles with the expression only in a small subset of nuclei, approximately 1 in 1000 FSHD muscle cell nuclei, probably representing occasional bursts of DUX4 expression [239].

At the end remain some doubts on the role of this gene in FSHD development [8].

1.4 Atrogin1 muscle specific ubiquitin ligase and the atrophic pathway

Skeletal muscle comprises approximately 40% of the total body mass of an average adult human and has energy requirements even at rest [319].

Skeletal muscle is a dynamic, multicellular tissue capable of adaptation throughout life in response to a variety of signals, including neural activation, mechanical loading, hormones/ growth factors, cytokines, and nutritional status [320].

Moreover, the inherent structure of skeletal muscle represents a substantial reservoir of proteins that can be utilized by hepatic tissue during times of stress. During periods of prolonged inactivity or disease, a substantial reduction in muscle mass is commonly observed, reducing total muscular energy requirements and, during disease, providing substrates for hepatic gluconeogenesis and acute phase protein production [321]. Despite these beneficial features, a dramatic loss of skeletal muscle mass can be debilitating, resulting in prolonged times to recovery, increased risk of subsequent injury and a severe burden on health care provisions [319].

Under many clinical conditions and chronic diseases skeletal muscle mass is lost, leading to muscle weakness, inactivity, and increased mortality. These are: starvation, limb immobilization, bed rest, cancer cachexia [320], diabetes [322], muscular dystrophies [320], renal failure [323], uremia

[322,324,325], sepsis [326,327], congestive heart failure [320], hyperthyroidism [323], Rheumatoid cachexia [328], Denervation [329], Spinal cord injury [328], neurodegeneration [320], nerve injury [330,331],mechanical ventilation [320], burn injury [332], HIV/AIDS, Chronic obstructive pulmonary disease [328], dexamethasone administration [333,334], Alcoholism [328], obesity [335] and aging [320].

The loss of muscle mass, i.e., muscle atrophy, is a complex process that occurs as a consequence of a variety of stressors, including neural inactivity, mechanical unloading, inflammation, metabolic stress, and elevated glucocorticoids [320] and result primarily from activation of a common biochemical program that stimulates muscle proteolysis [324,325,336,337].

Muscle atrophy occurs as the result of changes in the balance between anabolic and catabolic processes, with the result being a loss of muscle mass when protein breakdown exceeds protein synthesis [320].

In eukaryotic cells, four main mechanisms are responsible for the majority of cellular protein degradation, mediated via the actions of either cysteine-dependent aspartate specific proteases (caspases) [338,339], cathepsins [340], calcium-dependent calpains [341,342], or the ubiquitin proteasome system (UPS) [343,344].

The primary pathway responsible for the increase in muscle proteolysis under these catabolic conditions and therefore the principal regulator of skeletal muscle atrophy [324], was the

ATP-dependent, ubiquitin-proteasome pathway [345-347].

Proteins destined for degradation by the Ub-proteasome pathway are first covalently linked to a chain of Ub molecules, which marks them for rapid breakdown to short peptides by the 26S proteasome [348].

The ubiquitylation process includes a series of reactions involving three classes of proteins: ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes [E2s; also referred to as ubiquitin carrier proteins (UBC)], and ubiquitinprotein ligases (E3s) [349].

The process begins with the ATP-dependent activation of ubiquitin (Ub) by an E1, which results in a high-energy thioester linkage between the COOH terminus of ubiquitin and the active site cysteine of the E1. The activated ubiquitin is then transferred to an E2, again forming a thioester bond between Ub and a cysteine residue of the E2 enzyme.

The final step is the transfer of the ubiquitin from the E2 to a substrate via an E3 ligase.

Ubiquitin is usually conjugated to the target protein via an isopeptide bond between the ε-amino group of a lysine residue in the target protein and the carboxy-terminal glycine residue 76 in Ub [350,351].

Ubiquitin, a 76-amino acid protein with a molecular mass of 8.5 kDa, can be added to a protein as a single entity (monoubiquitin) or as a chain of variable length (polyubiquitin). Ubiquitin contains seven lysines (K6, K11, K27, K29, K33, K48,

and K63), and thus ubiquitin molecules can be linked through any one of these seven lysines. Ubiquitin chains can be comprised of a single type of linkage (homotypic) or a mixture of linkages (heterotypic). Furthermore, chains can be unbranched or branched (forked), the latter being the result of two or more ubiquitin molecules being attached to a single ubiquitin [352].

However, the process is repeated until a minimum of four Ub monomers are covalently attached via lysine residue 48 of Ub to the target protein, the classical formation that is recognized by the 26S proteasome as a signal to degrade the target protein [353].

Individual E3s ubiquitinate specific classes of proteins; hence, the identity of the proteins degraded by the proteasome is largely determined by the complement of E3s active in individual cells [323].

Following successful ubiquitination, and only once the criterion for recognition by the 26S proteasome has been met, proteins are unfolded and fed into the proteasome in an ATP-dependent process [344]. The 26S proteasome consists of a 20S catalytic core and 19S regulatory caps. Structurally, the 20S proteasome consists of four heptameric rings, formed from alpha subunits providing structural support [354], and the beta subunits responsible for the chymotrypsinlike, trypsin-like and caspase-like activities. The proteasome cleaves tagged proteins into short oligonucleotides, after which the activity of tripeptidyl-peptidase II and exopeptidases result in almost complete

degradation of the original protein [355].

Indeed, even under normal physiological conditions, the UPS is constantly degrading damaged or malformed proteins [356], so as to maintain normal cell function.

In Bodine et al. [329] two E3 ligases, MuRF1 (Trim63) and MAFbx (FBX032), were identified as genes that are similarly altered under disparate atrophy conditions. One gene was a RING finger protein previously identified as MuRF1 in cardiac tissue but not previously associated with skeletal muscle atrophy [357]. The other gene was novel and contained an F-box domain, which was characteristic of a family of E3 ligases that function as one component of a SCF (Skp1-Cullin1-F-box protein) ubiquitin ligase complex [358,359,329,323]. This novel gene was shown to bind to both Skip and Cullin and was subsequently named MAFbx (muscle atrophy F-box). The E3 ubiquitin ligase FBX032 is referred to in the literature as both MAFbx and Atrogin-1.

The discovery of MuRF1 and MAFbx yielded two genes having characteristics of key regulators of skeletal muscle atrophy. In fact both genes are selectively expressed in striated muscle, both genes are expressed at relatively low levels in resting skeletal muscle, and the expression of both genes increases rapidly upon the onset of a variety of stressors and prior to the onset of muscle loss.

Since their initial discovery, MuRF1 and MAFbx mRNA expression has been reported to be elevated in a wide range of

atrophy-inducing conditions, and thus the two genes have become recognized as key markers of muscle atrophy [320].

Moreover since they are rapidly activated it is suggested that they may contribute to the initiation of an atrophy program [328].

Atrogin1-1 presents an F-box at amino acid 228–267. This highly degenerate motif is found in a family of proteins, most of which function as one component of the SCF family of Ub-protein ligases (E3s) [360].

The F-box protein functions as an adaptor that binds proteins to be ubiquitinated and, through the F-box, associates with the Skp1 protein (or a homolog) and thus with a cullin protein (CUL1) that acts as a scaffold, and a RING finger protein (RBX1 or RBX2) that associates with CUL1 and recruits the ubiquitin-charged E2s [361,362].

In mammals, there are three classes of F-box proteins based on the structure of the substrate- binding domain.

MAFbx is part of the FBX class because it lacks WD40 repeats and leucine-rich regions, which are common structural motifs found in the other two classes [320].

However, at its extreme carboxy terminus (amino acids 346–355), Atrogin-1 does contain a motif known to interact with proteins containing class II PDZ domains. PDZ domain-containing proteins bind to specific sequences at the extreme C-termini of target proteins [363,364] and may represent a new type of substrate-binding region in F-box proteins.

Additional characterization of MAFbx has revealed that it

contains a leucine zipper domain and a leucine-charged residue- rich domain near its amino terminus that increases the number of substrates that MAFbx can potentially recognize [365].

Other features also present in the molecule include a potential bipartite nuclear localization sequence (amino acid 267–284) and a cytochrome c family heme-binding site signature (amino acid 317–322).

The presence in Atrogin-1 of a putative nuclear localization signal is of appreciable interest. Many other F-box proteins also contain nuclear localization sequences [366]. The presence of such a sequence in Atrogin-1, along with its muscle-specific expression, suggests that it may function in ubiquitinating muscle-specific transcription factors or other nuclear proteins involved in muscle growth. However, Atrogin-1 may also function to ubiquitinate cytosolic proteins such as components of the myofibril, which are rapidly degraded during muscle atrophy [323].

Thus, increased expression of MuRF1 and MAFbx following an atrophy-inducing stressor is thought to be responsible for the shift in protein balance from net synthesis to net degradation, thus inducing a loss of muscle mass [320].

Recent evidence has suggested that the transcriptional regulation of MAFbx/Atrogin-1 and MuRF1 Ub-ligases are intrinsically linked to both cellular metabolic status and inflammatory state via a coordinated pathway of signaling

events.

Li and colleagues discovered that both H2O2 and the catabolic cytokine tumor necrosis factor-alpha (TNFα), were competent inducers of MAFbx/Atrogin-1 mRNA expression in skeletal muscle [367].

Another example of activation of MAFbx is through myostatin, a member of the TGF superfamily and a negative regulator of muscle growth [368-375], which when elevated in vitro results in an increase in the expression of MAFbx [376,377]. Indeed, elevated myostatin has been shown to induce atrophy and fibrosis and impair regeneration in skeletal muscle [378-381].

At the end of the cascades different transcription factors explain their work.

The first transcription factors shown to regulate transcription of MAFBX E3 ligase were the class O-type forkhead transcription factors (FOXO), which include FOXO1, FOXO3a, and FOXO4 [320].

All of the FOXO transcription factors are expressed in skeletal muscle, and expression of FOXO1 and FOXO3a increases during certain forms of atrophy [382-384,322,385-387]. In 2004, two reports demonstrating the ability of the FOXO transcription factors to activate MAFbx were published [388,386]. In Stitt et al. [388], it was revealed that activated FOXO1 was necessary but not sufficient to increase MAFbx gene expression in cultured myotubes. FoxO1 does not directly increase MAFbx levels but instead blocks the IGF-1 inhibition of their

upregulation.

In contrast, Sandri et al. [386] concentrated their investigation on FOXO3a and demonstrated that FoxO3a binds directly to the MAFbx promoter, increasing its transcription and induce muscle atrophy.

The MAFbx promoter have also putative binding sites for other transcription factors such as the NF-κB transcription factors [327], CCAAT/enhancer-binding protein-β (C/EBPβ) transcription factor [389] and Small Mothers Against Decapentaplegic (Smad3) [335].

The two most widely acknowledged targets of MAFbx in skeletal muscle are MyoD, a myogenic regulatory factor, and eukaryotic translation initiation factor 3 subunit f (eIF3-f) [390,365]. eIF3-f appears to be a key effector of MAFbx as targeted increases and decreases in eIF3-f levels cause skeletal muscle hypertrophy and atrophy, respectively [390]. The MAFbx–MyoD interaction occurs in the nucleus. Overexpression of MAFbx results in the polyubiquitination of MyoD and an inhibition of MyoD-induced myotube differentiation and formation [391]. Similar to MyoD, MAFbx binds to myogenin resulting in its polyubiquitination and subsequent degradation during dexamethasone-induced myotube atrophy [392]. Together, these studies demonstrate that these myogenic transcription factors are targets of MAFbx and suggest a role for MAFbx in the regulation of myoblast differentiation [328].

More recently, MyHC and other sarcomeric proteins, such as the intermediate filament proteins vimentin and desmin, have been identified as potential MAFbx substrates [376,377].

To date, both MuRF1 and MAFbx have been reported to have the ability to bind MyHC and other myofibrillar proteins and to mediate the ubiquitination of these proteins in vitro. It is possible that the two E3 ligases have overlapping sets of substrates; however, it is more likely the case that they have distinct sets of substrates [320].

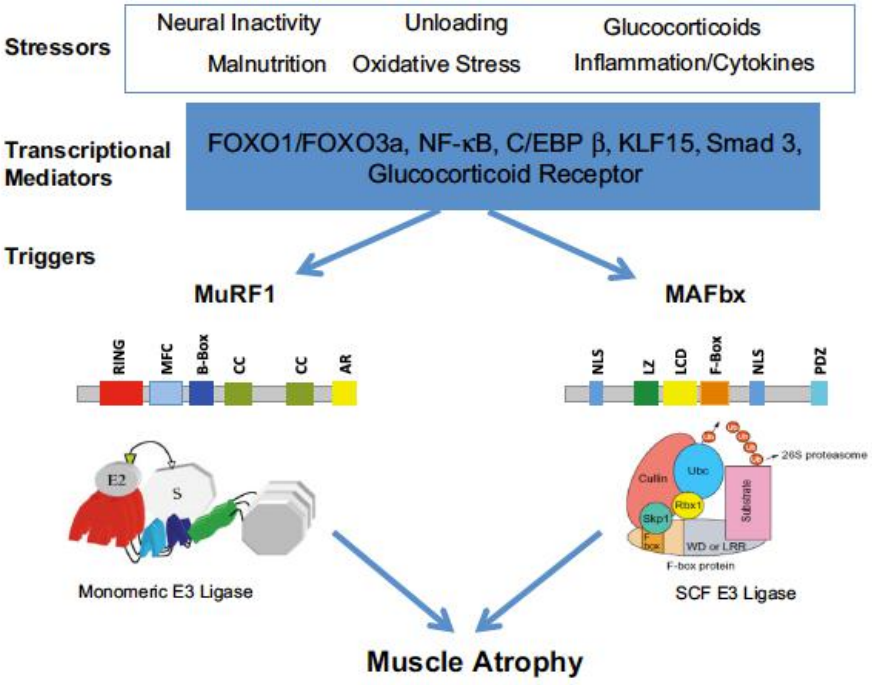


Figure 5. Regulation of muscle RING finger 1 (MuRF1) and muscle atrophy F-box (MAFbx) expression in skeletal muscle. Skeletal muscle atrophy is induced by a number of stressors, as illustrated here. These stressors can lead to the increase in the expression of a number of transcription factors. These transcriptional mediators can bind to the promoter regions of either the MuRF1 or MAFbx genes, leading to an increase in their expression levels within the muscle. A schematic of the putative domain structure of each protein is shown. MAFbx contains 2 nuclear localization signals (NLS), a leucine-zipper domain (LZ), a leucinecharged residue-rich domain (LCD), an F-box domain (F-box), and a PDZ domain (PDZ). SCF, Skp1-Cullin1-F-box protein [320].

Scope of the thesis

Facioscapulohumeral (FSHD) dystrophy is the third most common myopathy, with an incidence of 1:14.000 in the general population; it is caused by deletion of D4Z4 units under the threshold of 11 or their epigenetic deregulation, on chromosome 4q35.2 (FSHD locus), where it is organized as a tandem array of 11 to 100-150 copies in the general population [190,274]. D4Z4 belongs to a family of repeats with high sequence identity present also in 10q26, as a polymorphic tandem array, and 1p12 and the p-arm of acrocentric chromosomes as interspersed units; also the region upstream the array is heavily duplicated in the human genome [191,393,188,394]. Thus, it is almost unfeasible to distinguish 4q-derived sequences in next generation sequencing (NGS) studies, and this is the reason why omics approaches are still lacking for FSHD.

The model of inheritance of FSHD is autosomal dominant, with complete penetrance by age 20. No biochemical, histological, or instrumental markers are available to independently confirm a specific FSHD diagnosis that remains mainly clinical [395].

We are interested in the dissection of the molecular mechanism at the basis of FSHD manifestation. In this regard, unfortunately, almost 20 years after the definition of the genetic defect underlying this disease, the quest for the FSHD gene(s) is still open. The rate of progression of the disease is often unpredictable, as periods of remission can be followed by

dramatic exacerbation with no clear correlation to specific events. Thus, clinical prognosis is difficult. Furthermore, the degree of severity is highly variable even within the same family, suggesting that factors such as modifier genes, gender and environment may also affect disease onset and severity. The finding of asymptomatic or minimally affected carriers of contracted 4q alleles [272,234,396,397], as well as the recognition of the various genetic and epigenetic contributors to the FSHD phenotype [202], highlight the complexity on the molecular mechanism underlying FSHD, claiming for the development of novel approaches to achieve a global comprehension of the genetic determinants at the basis of the disease.

A recognized and unifying model for FSHD is that D4Z4 derepression results in the pathogenic activation of DUX4, a retrogene present within the last D4Z4 repeat of the array, encoding a homeobox containing transcription factor [259]; in healthy subjects, DUX4 is expressed only in the germ line, while it is epigenetically silenced in somatic tissues. In FSHD, DUX4 is aberrantly expressed in skeletal muscle [168,239]. DUX4 is a transcriptional regulator that activates transcription of genes and retroelements specifically in FSHD muscle cells [307,308]. A consensus is not reached yet if the DUX4-dependent transcriptome altered in FSHD can globally recapitulate the transcriptional signature for FSHD cells, as different models for DUX4 expression in muscle cells are

employed and few datasets (RNA-seq) in general are available [398,309,399]. Moreover, analyzing the sole RNA seq profile, it is not possible to distinguish between DUX4 targets, indirect effects or genes altered for different, still unknown, mechanisms. We have hypothesized that 4q D4Z4 array may regulate *in trans* the epigenetic state of a subset of coding and/or non-coding transcripts, interacting 3D in the genome of human muscle cells [301,278,400,401]**.** To demonstrate our hypothesis, we have used a combination of omics approaches including 4q-specific D4Z4 4C-seq, ChIP-seq (H3K27me3, H3K4me3, H3K4me1, H3K27ac, H3K36me3), FSHD transcriptome and published DUX4 dependent transcriptome [399]. Among the D4Z4 interactors epigenetically unpaired in FSHD, a subset is composed by DUX4 targets, and shows a chromatin switch versus a transcriptionally active state in myoblast cells, where DUX4 is not expressed yet; a subset is not targeted by DUX4, and among them we have deeply investigated Atrogin1-D4Z4 interaction with independent experimental strategies (including 3D-multicolor-FISH).

References

1. Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., Allis, C.D.: Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Molecular and cellular biology **26**(7), 2560-2569 (2006). doi:10.1128/MCB.26.7.2560-2569.2006

2. Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I., Fakan, S.: Chromosome territories--a functional nuclear landscape. Current opinion in cell biology **18**(3), 307-316 (2006). doi:10.1016/j.ceb.2006.04.007

3. Lu, L., Sun, K., Chen, X., Zhao, Y., Wang, L., Zhou, L., Sun, H., Wang, H.: Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. The EMBO journal **32**(19), 2575-2588 (2013). doi:10.1038/emboj.2013.182

4. Mao, Y.S., Zhang, B., Spector, D.L.: Biogenesis and function of nuclear bodies. Trends in genetics : TIG **27**(8), 295-306 (2011). doi:10.1016/j.tig.2011.05.006

5. Pombo, A., Dillon, N.: Three-dimensional genome architecture: players and mechanisms. Nature reviews. Molecular cell biology **16**(4), 245-257 (2015). doi:10.1038/nrm3965

6. Baylin, S.B., Jones, P.A.: A decade of exploring the cancer epigenome - biological and translational implications. Nature reviews. Cancer **11**(10), 726-734 (2011). doi:10.1038/nrc3130

7. Bhaumik, S.R., Smith, E., Shilatifard, A.: Covalent modifications of histones during development and disease pathogenesis. Nature structural & molecular biology **14**(11), 1008-1016 (2007). doi:10.1038/nsmb1337

8. Lanzuolo, C.: Epigenetic alterations in muscular disorders. Comparative and functional genomics **2012**, 256892 (2012). doi:10.1155/2012/256892

9. Barrera, L.O., Ren, B.: The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding. Current opinion in cell biology **18**(3), 291-298 (2006). doi:10.1016/j.ceb.2006.04.002

10. Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., van Steensel, B.: Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell **143**(2), 212-224 (2010). doi:10.1016/j.cell.2010.09.009

11. Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M.Y., Luquette, L.J., Xi, R., Jung, Y.L., Park, R.W., Bishop, E.P., Canfield, T.K., Sandstrom, R., Thurman, R.E., MacAlpine, D.M., Stamatoyannopoulos, J.A., Kellis, M., Elgin, S.C., Kuroda, M.I., Pirrotta, V., Karpen, G.H., Park, P.J.: Comprehensive analysis of the chromatin landscape

in Drosophila melanogaster. Nature **471**(7339), 480-485 (2011). doi:10.1038/nature09725

12. Schwartz, Y.B., Kahn, T.G., Stenberg, P., Ohno, K., Bourgon, R., Pirrotta, V.: Alternative epigenetic chromatin states of polycomb target genes. PLoS genetics **6**(1), e1000805 (2010). doi:10.1371/journal.pgen.1000805

13. Consortium, E.P.: An integrated encyclopedia of DNA elements in the human genome. Nature **489**(7414), 57-74 (2012). doi:10.1038/nature11247

14. Bohmdorfer, G., Wierzbicki, A.T.: Control of Chromatin Structure by Long Noncoding RNA. Trends in cell biology **25**(10), 623-632 (2015). doi:10.1016/j.tcb.2015.07.002

15. De Lucia, F., Dean, C.: Long non-coding RNAs and chromatin regulation. Current opinion in plant biology **14**(2), 168-173 (2011). doi:10.1016/j.pbi.2010.11.006

16. Malecova, B., Morris, K.V.: Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. Current opinion in molecular therapeutics **12**(2), 214-222 (2010).

17. Umlauf, D., Fraser, P., Nagano, T.: The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. Biological chemistry **389**(4), 323-331 (2008). doi:10.1515/BC.2008.047

18. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J.,

Heard, E.: Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature **485**(7398), 381-385 (2012). doi:10.1038/nature11049

19. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B.: Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485**(7398), 376-380 (2012). doi:10.1038/nature11082

20. Hou, C., Li, L., Qin, Z.S., Corces, V.G.: Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Molecular cell **48**(3), 471-484 (2012). doi:10.1016/j.molcel.2012.08.031

21. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., Cavalli, G.: Three-dimensional folding and functional organization principles of the Drosophila genome. Cell **148**(3), 458-472 (2012). doi:10.1016/j.cell.2012.01.010

22. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J.: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326**(5950), 289-293 (2009). doi:10.1126/science.1181369

23. Cavalli, G., Misteli, T.: Functional implications of genome topology. Nature structural & molecular biology **20**(3), 290-299 (2013). doi:10.1038/nsmb.2474

24. Li, G., Zhu, P.: Structure and organization of chromatin fiber in the nucleus. FEBS letters **589**(20 Pt A), 2893-2904 (2015). doi:10.1016/j.febslet.2015.04.023

25. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J.: Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389**(6648), 251-260 (1997). doi:10.1038/38444

26. Kornberg, R.D.: Chromatin structure: a repeating unit of histones and DNA. Science **184**(4139), 868-871 (1974).

27. Robinson, P.J., Rhodes, D.: Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Current opinion in structural biology **16**(3), 336-343 (2006). doi:10.1016/j.sbi.2006.05.007

28. Li, G., Reinberg, D.: Chromatin higher-order structures and gene regulation. Current opinion in genetics & development **21**(2), 175-186 (2011). doi:10.1016/j.gde.2011.01.022

29. Woodcock, C.L., Dimitrov, S.: Higher-order structure of chromatin and chromosomes. Current opinion in genetics & development **11**(2), 130-135 (2001).

30. Hou, C., Corces, V.G.: Throwing transcription for a loop: expression of the genome in the 3D nucleus. Chromosoma **121**(2), 107-116 (2012). doi:10.1007/s00412-011-0352-7

31. Yang, J., Corces, V.G.: Insulators, long-range interactions, and genome function. Current opinion in genetics & development **22**(2), 86-92 (2012). doi:10.1016/j.gde.2011.12.007

32. Deng, W., Blobel, G.A.: Do chromatin loops provide epigenetic gene expression states? Current opinion in genetics & development **20**(5), 548-554 (2010). doi:10.1016/j.gde.2010.06.007

33. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., Bernstein, B.E.: Mapping and analysis of chromatin state dynamics in nine human cell types. Nature **473**(7345), 43-49 (2011). doi:10.1038/nature09906

34. Liu, T., Rechtsteiner, A., Egelhofer, T.A., Vielle, A., Latorre, I., Cheung, M.S., Ercan, S., Ikegami, K., Jensen, M., Kolasinska-Zwierz, P., Rosenbaum, H., Shin, H., Taing, S., Takasaki, T., Iniguez, A.L., Desai, A., Dernburg, A.F., Kimura, H., Lieb, J.D., Ahringer, J., Strome, S., Liu, X.S.: Broad chromosomal domains of histone modification patterns in C. elegans. Genome research **21**(2), 227-236 (2011). doi:10.1101/gr.115519.110

35. Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., Heisterkamp, S., van Kampen, A., Versteeg, R.: The human transcriptome map: clustering of

highly expressed genes in chromosomal domains. Science **291**(5507), 1289-1292 (2001). doi:10.1126/science.1056794

36. Deng, W., Blobel, G.A.: Manipulating nuclear architecture. Current opinion in genetics & development **25**, 1-7 (2014). doi:10.1016/j.gde.2013.10.014

37. Fraser, P., Bickmore, W.: Nuclear organization of the genome and the potential for gene regulation. Nature **447**(7143), 413-417 (2007). doi:10.1038/nature05916

38. Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G., Cremer, T.: Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nature reviews. Genetics **8**(2), 104-115 (2007). doi:10.1038/nrg2041

39. Soutoglou, E., Misteli, T.: Mobility and immobility of chromatin in transcription and genome stability. Current opinion in genetics & development **17**(5), 435-442 (2007). doi:10.1016/j.gde.2007.08.004

40. Schoenfelder, S., Clay, I., Fraser, P.: The transcriptional interactome: gene expression in 3D. Current opinion in genetics & development **20**(2), 127-133 (2010). doi:10.1016/j.gde.2010.02.002

41. Cremer, T., Cremer, M.: Chromosome territories. Cold Spring Harbor perspectives in biology **2**(3), a003889 (2010). doi:10.1101/cshperspect.a003889

42. Williams, A., Spilianakis, C.G., Flavell, R.A.: Interchromosomal association and gene regulation in trans.

Trends in genetics : TIG **26**(4), 188-197 (2010). doi:10.1016/j.tig.2010.01.007

43. Cortesi, A., Bodega, B.: Chromosome Conformation Capture in Primary Human Cells. Methods in molecular biology **1480**, 213-221 (2016). doi:10.1007/978-1-4939-6380-5_19

44. Bickmore, W.A.: The spatial organization of the human genome. Annual review of genomics and human genetics **14**, 67-84 (2013). doi:10.1146/annurev-genom-091212-153515

45. Volpi, E.V., Bridger, J.M.: FISH glossary: an overview of the fluorescence in situ hybridization technique. BioTechniques **45**(4), 385-386, 388, 390 passim (2008). doi:10.2144/000112811

46. de Wit, E., de Laat, W.: A decade of 3C technologies: insights into nuclear organization. Genes & development **26**(1), 11-24 (2012). doi:10.1101/gad.179804.111

47. Giorgetti, L., Heard, E.: Closing the loop: 3C versus DNA FISH. Genome biology **17**(1), 215 (2016). doi:10.1186/s13059-016-1081-2

48. Boyle, S., Rodesch, M.J., Halvensleben, H.A., Jeddeloh, J.A., Bickmore, W.A.: Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology **19**(7), 901-909 (2011). doi:10.1007/s10577-011-9245-0

49. Cremer, T., Cremer, C.: Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nature reviews. Genetics **2**(4), 292-301 (2001). doi:10.1038/35066075
50. Fraser, J., Williamson, I., Bickmore, W.A., Dostie, J.: An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiology and molecular biology reviews : MMBR **79**(3), 347-372 (2015). doi:10.1128/MMBR.00006-15
51. Dekker, J., Rippe, K., Dekker, M., Kleckner, N.: Capturing chromosome conformation. Science **295**(5558), 1306-1311 (2002). doi:10.1126/science.1067799
52. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., de Laat, W.: Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell **10**(6), 1453-1465 (2002).
53. Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., de Laat, W.: The beta-globin nuclear compartment in development and erythroid differentiation. Nature genetics **35**(2), 190-194 (2003). doi:10.1038/ng1244
54. Dekker, J.: The three 'C' s of chromosome conformation capture: controls, controls, controls. Nature methods **3**(1), 17-21 (2006). doi:10.1038/nmeth823
55. Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., Axel, R.: Interchromosomal interactions and olfactory receptor choice. Cell **126**(2), 403-413 (2006). doi:10.1016/j.cell.2006.06.035

56. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., Flavell, R.A.: Interchromosomal associations between alternatively expressed loci. Nature **435**(7042), 637-645 (2005). doi:10.1038/nature03574

57. Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., Ohlsson, R.: CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proceedings of the National Academy of Sciences of the United States of America **103**(28), 10684-10689 (2006). doi:10.1073/pnas.0600326103

58. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., Orlando, V.: Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nature cell biology **9**(10), 1167-1174 (2007). doi:10.1038/ncb1637

59. Xu, N., Tsai, C.L., Lee, J.T.: Transient homologous chromosome pairing marks the onset of X inactivation. Science **311**(5764), 1149-1152 (2006). doi:10.1126/science.1122984

60. Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., Heard, E.: Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nature cell biology **8**(3), 293-299 (2006). doi:10.1038/ncb1365

61. Taft, R.J., Pheasant, M., Mattick, J.S.: The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays : news and reviews in molecular, cellular and

developmental biology **29**(3), 288-299 (2007). doi:10.1002/bies.20544

62. Boeke, J.D., Devine, S.E.: Yeast retrotransposons: finding a nice quiet neighborhood. Cell **93**(7), 1087-1089 (1998).

63. Neguembor, M.V., Gabellini, D.: In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy. Epigenomics **2**(2), 271-287 (2010). doi:10.2217/epi.10.8

64. de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., Pollock, D.D.: Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics **7**(12), e1002384 (2011). doi:10.1371/journal.pgen.1002384

65. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A.,

Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown,

D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing, C.: Initial sequencing and analysis of the human genome. Nature **409**(6822), 860-921 (2001). doi:10.1038/35057062

66. Casa, V., Gabellini, D.: A repetitive elements perspective in Polycomb epigenetics. Frontiers in genetics **3**, 199 (2012). doi:10.3389/fgene.2012.00199

67. Kidwell, M.G., Lisch, D.R.: Transposable elements and host genome evolution. Trends in ecology & evolution **15**(3), 95-99 (2000).

68. Feschotte, C.: Transposable elements and the evolution of regulatory networks. Nature reviews. Genetics **9**(5), 397-405 (2008). doi:10.1038/nrg2337

69. Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., Rivera, M.N., Bardeesy, N., Maheswaran, S., Haber, D.A.: Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science **331**(6017), 593-596 (2011). doi:10.1126/science.1200801

70. Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., Verma, I.M.: BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature **477**(7363), 179-184 (2011). doi:10.1038/nature10371

71. Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., Haussler, D.: Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proceedings of the National Academy of Sciences of the United States of America **104**(47), 18613-18618 (2007). doi:10.1073/pnas.0703637104

72. Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., Liu, E.T.: Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome research **18**(11), 1752-1762 (2008). doi:10.1101/gr.080663.108

73. Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., Waki, K., Hornig, N., Arakawa, T., Takahashi, H.,

Kawai, J., Forrest, A.R., Suzuki, H., Hayashizaki, Y., Hume, D.A., Orlando, V., Grimmond, S.M., Carninci, P.: The regulated retrotransposon transcriptome of mammalian cells. Nature genetics **41**(5), 563-571 (2009). doi:10.1038/ng.368

74. Tyekucheva, S., Yolken, R.H., McCombie, W.R., Parla, J., Kramer, M., Wheelan, S.J., Sabunciyan, S.: Establishing the baseline level of repetitive element expression in the human cortex. BMC genomics **12**, 495 (2011). doi:10.1186/1471-2164-12-495

75. Chadwick, B.P.: DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. Genome research **18**(8), 1259-1269 (2008). doi:10.1101/gr.075713.107

76. Simeonova, I., Lejour, V., Bardot, B., Bouarich-Bourimi, R., Morin, A., Fang, M., Charbonnier, L., Toledo, F.: Fuzzy tandem repeats containing p53 response elements may define species-specific p53 target genes. PLoS genetics **8**(6), e1002731 (2012). doi:10.1371/journal.pgen.1002731

77. Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., Antonarakis, S.E.: Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature **332**(6160), 164-166 (1988). doi:10.1038/332164a0

78. Batzer, M.A., Deininger, P.L.: Alu repeats and human genomic diversity. Nature reviews. Genetics **3**(5), 370-379 (2002). doi:10.1038/nrg798

79. Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V.: Repetitive sequences in complex genomes: structure and evolution. Annual review of genomics and human genetics **8**, 241-259 (2007). doi:10.1146/annurev.genom.8.080706.092416

80. Kim, P.M., Lam, H.Y., Urban, A.E., Korbel, J.O., Affourtit, J., Grubert, F., Chen, X., Weissman, S., Snyder, M., Gerstein, M.B.: Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. Genome research **18**(12), 1865-1874 (2008). doi:10.1101/gr.081422.108

81. Britten, R.J.: Transposable element insertions have strongly affected human evolution. Proceedings of the National Academy of Sciences of the United States of America **107**(46), 19945-19948 (2010). doi:10.1073/pnas.1014330107

82. Hua-Van, A., Le Rouzic, A., Boutin, T.S., Filee, J., Capy, P.: The struggle for life of the genome's selfish architects. Biology direct **6**, 19 (2011). doi:10.1186/1745-6150-6-19

83. Faulkner, G.J.: Retrotransposons: mobile and mutagenic from conception to death. FEBS letters **585**(11), 1589-1594 (2011). doi:10.1016/j.febslet.2011.03.061

84. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian, H.H., Jr.: Hot L1s account for the bulk of retrotransposition in the human population.

Proceedings of the National Academy of Sciences of the United States of America **100**(9), 5280-5285 (2003). doi:10.1073/pnas.0831042100

85. Bennett, E.A., Coleman, L.E., Tsui, C., Pittard, W.S., Devine, S.E.: Natural genetic variation caused by transposable elements in humans. Genetics **168**(2), 933-951 (2004). doi:10.1534/genetics.104.031757

86. Mills, R.E., Bennett, E.A., Iskow, R.C., Devine, S.E.: Which transposable elements are active in the human genome? Trends in genetics : TIG **23**(4), 183-191 (2007). doi:10.1016/j.tig.2007.02.006

87. Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., Devine, S.E.: Natural mutagenesis of human genomes by endogenous retrotransposons. Cell **141**(7), 1253-1261 (2010). doi:10.1016/j.cell.2010.05.020

88. Ekram, M.B., Kang, K., Kim, H., Kim, J.: Retrotransposons as a major source of epigenetic variations in the mammalian genome. Epigenetics **7**(4), 370-382 (2012). doi:10.4161/epi.19462

89. Solyom, S., Kazazian, H.H., Jr.: Mobile elements in the human genome: implications for disease. Genome medicine **4**(2), 12 (2012). doi:10.1186/gm311

90. Mitra, R., Li, X., Kapusta, A., Mayhew, D., Mitra, R.D., Feschotte, C., Craig, N.L.: Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active

mammalian DNA transposon. Proceedings of the National Academy of Sciences of the United States of America **110**(1), 234-239 (2013). doi:10.1073/pnas.1217548110

91. Ray, D.A., Pagan, H.J., Thompson, M.L., Stevens, R.D.: Bats with hATs: evidence for recent DNA transposon activity in genus Myotis. Molecular biology and evolution **24**(3), 632-639 (2007). doi:10.1093/molbev/msl192

92. Hancks, D.C., Kazazian, H.H., Jr.: Roles for retrotransposon insertions in human disease. Mobile DNA **7**, 9 (2016). doi:10.1186/s13100-016-0065-9

93. Richardson, S.R., Doucet, A.J., Kopera, H.C., Moldovan, J.B., Garcia-Perez, J.L., Moran, J.V.: The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiology spectrum **3**(2), MDNA3-0061-2014 (2015). doi:10.1128/microbiolspec.MDNA3-0061-2014

94. Munoz-Lopez, M., Garcia-Perez, J.L.: DNA transposons: nature and applications in genomics. Current genomics **11**(2), 115-128 (2010). doi:10.2174/138920210790886871

95. Bodega, B., Orlando, V.: Repetitive elements dynamics in cell identity programming, maintenance and disease. Current opinion in cell biology **31**, 67-73 (2014). doi:10.1016/j.ceb.2014.09.002

96. Mager, D.L., Stoye, J.P.: Mammalian Endogenous Retroviruses. Microbiology spectrum **3**(1), MDNA3-0009-2014 (2015). doi:10.1128/microbiolspec.MDNA3-0009-2014

97. Lee, Y.N., Bieniasz, P.D.: Reconstitution of an infectious human endogenous retrovirus. PLoS pathogens **3**(1), e10 (2007). doi:10.1371/journal.ppat.0030010

98. Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., Tristem, M.: Rate of recombinational deletion among human endogenous retroviruses. Journal of virology **81**(17), 9437-9442 (2007). doi:10.1128/JVI.02216-06

99. Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., Mager, D.L.: Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS genetics **2**(1), e2 (2006). doi:10.1371/journal.pgen.0020002

100. Garcia-Perez, J.L., Widmann, T.J., Adams, I.R.: The impact of transposable elements on mammalian development. Development **143**(22), 4101-4114 (2016). doi:10.1242/dev.132639

101. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian, H.H., Jr.: High frequency retrotransposition in cultured mammalian cells. Cell **87**(5), 917-927 (1996).

102. Hohjoh, H., Singer, M.F.: Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. The EMBO journal **16**(19), 6034-6043 (1997). doi:10.1093/emboj/16.19.6034

103. Martin, S.L., Bushman, F.D.: Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1

retrotransposon. Molecular and cellular biology **21**(2), 467-475 (2001). doi:10.1128/MCB.21.2.467-475.2001

104. Feng, Q., Moran, J.V., Kazazian, H.H., Jr., Boeke, J.D.: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87**(5), 905-916 (1996).

105. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., Gabriel, A.: Reverse transcriptase encoded by a human transposable element. Science **254**(5039), 1808-1810 (1991).

106. Munoz-Lopez, M., Vilar-Astasio, R., Tristan-Ramos, P., Lopez-Ruiz, C., Garcia-Perez, J.L.: Study of Transposable Elements and Their Genomic Impact. Methods in molecular biology **1400**, 1-19 (2016). doi:10.1007/978-1-4939-3372-3_1

107. Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., Moran, J.V.: LINE-1 retrotransposition activity in human genomes. Cell **141**(7), 1159-1170 (2010). doi:10.1016/j.cell.2010.05.021

108. Dewannieux, M., Esnault, C., Heidmann, T.: LINE-mediated retrotransposition of marked Alu sequences. Nature genetics **35**(1), 41-48 (2003). doi:10.1038/ng1223

109. Dewannieux, M., Heidmann, T.: L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. Journal of molecular biology **349**(2), 241-247 (2005). doi:10.1016/j.jmb.2005.03.068

110. Hancks, D.C., Goodier, J.L., Mandal, P.K., Cheung, L.E., Kazazian, H.H., Jr.: Retrotransposition of marked SVA elements

by human L1s in cultured cells. Human molecular genetics **20**(17), 3386-3400 (2011). doi:10.1093/hmg/ddr245

111. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Lower, J., Stratling, W.H., Lower, R., Schumann, G.G.: The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic acids research **40**(4), 1666-1683 (2012). doi:10.1093/nar/gkr863

112. Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., Gage, F.H.: Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature **435**(7044), 903-910 (2005). doi:10.1038/nature03663

113. Brack, C., Hirama, M., Lenhard-Schuller, R., Tonegawa, S.: A complete immunoglobulin gene is created by somatic recombination. Cell **15**(1), 1-14 (1978).

114. Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., Gage, F.H.: L1 retrotransposition in neurons is modulated by MeCP2. Nature **468**(7322), 443-446 (2010). doi:10.1038/nature09544

115. Heras, S.R., Macias, S., Caceres, J.F., Garcia-Perez, J.L.: Control of mammalian retrotransposons by cellular RNA processing activities. Mobile genetic elements **4**, e28439 (2014). doi:10.4161/mge.28439

116. Pizarro, J.G., Cristofari, G.: Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic

Cells. Frontiers in cell and developmental biology **4**, 14 (2016). doi:10.3389/fcell.2016.00014

117. Cohen, C.J., Lock, W.M., Mager, D.L.: Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene **448**(2), 105-114 (2009). doi:10.1016/j.gene.2009.06.020

118. Perepelitsa-Belancio, V., Deininger, P.: RNA truncation by premature polyadenylation attenuates human mobile element activity. Nature genetics **35**(4), 363-366 (2003). doi:10.1038/ng1269

119. Han, J.S., Szak, S.T., Boeke, J.D.: Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature **429**(6989), 268-274 (2004). doi:10.1038/nature02536

120. Belancio, V.P., Hedges, D.J., Deininger, P.: LINE-1 RNA splicing and influences on mammalian gene expression. Nucleic acids research **34**(5), 1512-1521 (2006). doi:10.1093/nar/gkl027

121. Piriyapongsa, J., Polavarapu, N., Borodovsky, M., McDonald, J.: Exonization of the LTR transposable elements in human genome. BMC genomics **8**, 291 (2007). doi:10.1186/1471-2164-8-291

122. Heras, S.R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyras, E., Garcia-Perez, J.L., Caceres, J.F.: The Microprocessor controls the activity of mammalian

retrotransposons. Nature structural & molecular biology **20**(10), 1173-1181 (2013). doi:10.1038/nsmb.2658

123. Piriyapongsa, J., Rutledge, M.T., Patel, S., Borodovsky, M., Jordan, I.K.: Evaluating the protein coding potential of exonized transposable element sequences. Biology direct **2**, 31 (2007). doi:10.1186/1745-6150-2-31

124. Zarnack, K., Konig, J., Tajnik, M., Martincorena, I., Eustermann, S., Stevant, I., Reyes, A., Anders, S., Luscombe, N.M., Ule, J.: Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell **152**(3), 453-466 (2013). doi:10.1016/j.cell.2012.12.023

125. Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., Feschotte, C.: Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS genetics **9**(4), e1003470 (2013). doi:10.1371/journal.pgen.1003470

126. Macia, A., Munoz-Lopez, M., Cortes, J.L., Hastings, R.K., Morell, S., Lucena-Aguilar, G., Marchal, J.A., Badge, R.M., Garcia-Perez, J.L.: Epigenetic control of retrotransposon expression in human embryonic stem cells. Molecular and cellular biology **31**(2), 300-316 (2011). doi:10.1128/MCB.00561-10

127. Macia, A., Blanco-Jimenez, E., Garcia-Perez, J.L.: Retrotransposons in pluripotent cells: Impact and new roles in

cellular plasticity. Biochimica et biophysica acta **1849**(4), 417-426 (2015). doi:10.1016/j.bbagrm.2014.07.007

128. Mercer, T.R., Dinger, M.E., Mattick, J.S.: Long non-coding RNAs: insights into functions. Nature reviews. Genetics **10**(3), 155-159 (2009). doi:10.1038/nrg2521

129. Moran, V.A., Perera, R.J., Khalil, A.M.: Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic acids research **40**(14), 6391-6400 (2012). doi:10.1093/nar/gks296

130. Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., Noro, Y., Wong, C.H., de Hoon, M., Andersson, R., Sandelin, A., Suzuki, H., Wei, C.L., Koseki, H., Consortium, F., Hasegawa, Y., Forrest, A.R., Carninci, P.: Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics **46**(6), 558-566 (2014). doi:10.1038/ng.2965

131. Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J.L., Root, D.E., Lander, E.S.: lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature **477**(7364), 295-300 (2011). doi:10.1038/nature10398

132. Lu, X., Sachs, F., Ramsay, L., Jacques, P.E., Goke, J., Bourque, G., Ng, H.H.: The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell

identity. Nature structural & molecular biology **21**(4), 423-425 (2014). doi:10.1038/nsmb.2799

133. Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., Schumann, G.G., Chen, W., Lorincz, M.C., Ivics, Z., Hurst, L.D., Izsvak, Z.: Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature **516**(7531), 405-409 (2014). doi:10.1038/nature13804

134. Esnault, C., Maestre, J., Heidmann, T.: Human LINE retrotransposons generate processed pseudogenes. Nature genetics **24**(4), 363-367 (2000). doi:10.1038/74184

135. Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., Moran, J.V.: Human L1 retrotransposition: cis preference versus trans complementation. Molecular and cellular biology **21**(4), 1429-1439 (2001). doi:10.1128/MCB.21.4.1429-1439.2001

136. Moran, J.V., DeBerardinis, R.J., Kazazian, H.H., Jr.: Exon shuffling by L1 retrotransposition. Science **283**(5407), 1530-1534 (1999).

137. Grimaldi, G., Skowronski, J., Singer, M.F.: Defining the beginning and end of KpnI family segments. The EMBO journal **3**(8), 1753-1759 (1984).

138. McClintock, B.: Controlling elements and the gene. Cold Spring Harbor symposia on quantitative biology **21**, 197-216 (1956).

139. Britten, R.J., Davidson, E.H.: Gene regulation for higher cells: a theory. Science **165**(3891), 349-357 (1969).

140. Warburton, P.E., Hasson, D., Guillem, F., Lescale, C., Jin, X., Abrusan, G.: Analysis of the largest tandemly repeated DNA families in the human genome. BMC genomics **9**, 533 (2008). doi:10.1186/1471-2164-9-533

141. Gelfand, Y., Rodriguez, A., Benson, G.: TRDB--the Tandem Repeats Database. Nucleic acids research **35**(Database issue), D80-87 (2007). doi:10.1093/nar/gkl1013

142. Ames, D., Murphy, N., Helentjaris, T., Sun, N., Chandler, V.: Comparative analyses of human single- and multilocus tandem repeats. Genetics **179**(3), 1693-1704 (2008). doi:10.1534/genetics.108.087882

143. Rudd, M.K., Wray, G.A., Willard, H.F.: The evolutionary dynamics of alpha-satellite. Genome research **16**(1), 88-96 (2006). doi:10.1101/gr.3810906

144. Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., Willard, H.F.: Genomic and genetic definition of a functional human centromere. Science **294**(5540), 109-115 (2001). doi:10.1126/science.1065042

145. Morris, C.A., Moazed, D.: Centromere assembly and propagation. Cell **128**(4), 647-650 (2007). doi:10.1016/j.cell.2007.02.002

146. Blackburn, E.H.: The molecular structure of centromeres and telomeres. Annual review of biochemistry **53**, 163-194 (1984). doi:10.1146/annurev.bi.53.070184.001115

147. Rudd, M.K., Willard, H.F.: Analysis of the centromeric regions of the human genome assembly. Trends in genetics : TIG **20**(11), 529-533 (2004). doi:10.1016/j.tig.2004.08.008

148. Vissel, B., Nagy, A., Choo, K.H.: A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA. Cytogenetics and cell genetics **61**(2), 81-86 (1992).

149. Kim, J.H., Ebersole, T., Kouprina, N., Noskov, V.N., Ohzeki, J., Masumoto, H., Mravinac, B., Sullivan, B.A., Pavlicek, A., Dovat, S., Pack, S.D., Kwon, Y.W., Flanagan, P.T., Loukinov, D., Lobanenkov, V., Larionov, V.: Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. Genome research **19**(4), 533-544 (2009). doi:10.1101/gr.086496.108

150. Prosser, J., Frommer, M., Paul, C., Vincent, P.C.: Sequence relationships of three human satellite DNAs. Journal of molecular biology **187**(2), 145-155 (1986).

151. Waye, J.S., Willard, H.F.: Human beta satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA. Proceedings of the National Academy of Sciences of the United States of America **86**(16), 6250-6254 (1989).

152. Lin, C.C., Sasi, R., Lee, C., Fan, Y.S., Court, D.: Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. Chromosoma **102**(5), 333-339 (1993).

153. Lee, C., Critcher, R., Zhang, J.G., Mills, W., Farr, C.J.: Distribution of gamma satellite DNA on the human X and Y chromosomes suggests that it is not required for mitotic centromere function. Chromosoma **109**(6), 381-389 (2000).
154. Schueler, M.G., Dunn, J.M., Bird, C.P., Ross, M.T., Viggiano, L., Program, N.C.S., Rocchi, M., Willard, H.F., Green, E.D.: Progressive proximal expansion of the primate X chromosome centromere. Proceedings of the National Academy of Sciences of the United States of America **102**(30), 10563-10568 (2005). doi:10.1073/pnas.0503346102
155. Giacalone, J., Friedes, J., Francke, U.: A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. Nature genetics **1**(2), 137-143 (1992). doi:10.1038/ng0592-137
156. Tremblay, D.C., Alexander, G., Jr., Moseley, S., Chadwick, B.P.: Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. BMC genomics **11**, 632 (2010). doi:10.1186/1471-2164-11-632
157. Deidda, G., Cacurri, S., Grisanti, P., Vigneti, E., Piazzo, N., Felicetti, L.: Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. European journal of human genetics : EJHG **3**(3), 155-167 (1995).
158. Winokur, S.T., Bengtsson, U., Vargas, J.C., Wasmuth, J.J., Altherr, M.R., Weiffenbach, B., Jacobsen, S.J.: The

evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. Human molecular genetics **5**(10), 1567-1575 (1996).

159. Gondo, Y., Okada, T., Matsuyama, N., Saitoh, Y., Yanagisawa, Y., Ikeda, J.E.: Human megasatellite DNA RS447: copy-number polymorphisms and interspecies conservation. Genomics **54**(1), 39-49 (1998). doi:10.1006/geno.1998.5545

160. Warburton, P.E., Haaf, T., Gosden, J., Lawson, D., Willard, H.F.: Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. Genomics **33**(2), 220-228 (1996). doi:10.1006/geno.1996.0187

161. McLaughlin, C.R., Chadwick, B.P.: Characterization of DXZ4 conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. Genome biology **12**(4), R37 (2011). doi:10.1186/gb-2011-12-4-r37

162. Lopez Castel, A., Cleary, J.D., Pearson, C.E.: Repeat instability as the basis for human diseases and as a potential target for therapy. Nature reviews. Molecular cell biology **11**(3), 165-170 (2010). doi:10.1038/nrm2854

163. Eisenberg, D.T.: An evolutionary review of human telomere biology: the thrifty telomere hypothesis and notes on potential adaptive paternal effects. American journal of human

biology : the official journal of the Human Biology Council **23**(2), 149-167 (2011). doi:10.1002/ajhb.21127

164. Weinert, T.A., Hartwell, L.H.: The RAD9 gene controls the cell cycle response to DNA damage in Saccharomyces cerevisiae. Science **241**(4863), 317-322 (1988).

165. Sandell, L.L., Zakian, V.A.: Loss of a yeast telomere: arrest, recovery, and chromosome loss. Cell **75**(4), 729-739 (1993).

166. Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., de Lange, T.: Mammalian telomeres end in a large duplex loop. Cell **97**(4), 503-514 (1999).

167. van der Maarel, S.M., Frants, R.R., Padberg, G.W.: Facioscapulohumeral muscular dystrophy. Biochimica et biophysica acta **1772**(2), 186-194 (2007). doi:10.1016/j.bbadis.2006.05.009

168. Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Matteotti, C., van Acker, A.M., Leo, O., Figlewicz, D., Barro, M., Laoudj-Chenivesse, D., Belayew, A., Coppee, F., Chen, Y.W.: DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proceedings of the National Academy of Sciences of the United States of America **104**(46), 18157-18162 (2007). doi:10.1073/pnas.0708659104

169. Ehrlich, M., Jackson, K., Tsumagari, K., Camano, P., Lemmers, R.J.: Hybridization analysis of D4Z4 repeat arrays

linked to FSHD. Chromosoma **116**(2), 107-116 (2007). doi:10.1007/s00412-006-0080-6

170. Weber, J.L., Wong, C.: Mutation of human short tandem repeats. Human molecular genetics **2**(8), 1123-1128 (1993).

171. Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., Rolf, B.: Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. American journal of human genetics **62**(6), 1408-1415 (1998). doi:10.1086/301869

172. Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E.: Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular ecology **11**(12), 2453-2465 (2002).

173. Legendre, M., Pochet, N., Pak, T., Verstrepen, K.J.: Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome research **17**(12), 1787-1796 (2007). doi:10.1101/gr.6554007

174. Vinces, M.D., Legendre, M., Caldara, M., Hagihara, M., Verstrepen, K.J.: Unstable tandem repeats in promoters confer transcriptional evolvability. Science **324**(5931), 1213-1216 (2009). doi:10.1126/science.1170097

175. Gemayel, R., Vinces, M.D., Legendre, M., Verstrepen, K.J.: Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual review of genetics **44**, 445-477 (2010). doi:10.1146/annurev-genet-072610-155046

176. Rockman, M.V., Wray, G.A.: Abundant raw material for cis-regulatory evolution in humans. Molecular biology and evolution **19**(11), 1991-2004 (2002).
177. Kashi, Y., King, D.G.: Simple sequence repeats as advantageous mutators in evolution. Trends in genetics : TIG **22**(5), 253-259 (2006). doi:10.1016/j.tig.2006.03.005
178. King, M.C., Wilson, A.C.: Evolution at two levels in humans and chimpanzees. Science **188**(4184), 107-116 (1975).
179. Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., Romano, L.A.: The evolution of transcriptional regulation in eukaryotes. Molecular biology and evolution **20**(9), 1377-1419 (2003). doi:10.1093/molbev/msg140
180. Tirosh, I., Weinberger, A., Carmi, M., Barkai, N.: A genetic signature of interspecies variations in gene expression. Nature genetics **38**(7), 830-834 (2006). doi:10.1038/ng1819
181. Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J., Hartl, D.L.: Genetic properties influencing the evolvability of gene expression. Science **317**(5834), 118-121 (2007). doi:10.1126/science.1140247
182. Choi, J.K., Kim, Y.J.: Epigenetic regulation and the variability of gene expression. Nature genetics **40**(2), 141-147 (2008). doi:10.1038/ng.2007.58
183. Payseur, B.A., Jing, P., Haasl, R.J.: A genomic portrait of human microsatellite variation. Molecular biology and evolution **28**(1), 303-312 (2011). doi:10.1093/molbev/msq198

184. Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M.A., Gemmell, N.: Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PloS one **8**(2), e54710 (2013). doi:10.1371/journal.pone.0054710

185. Bilgin Sonay, T., Carvalho, T., Robinson, M.D., Greminger, M.P., Krutzen, M., Comas, D., Highnam, G., Mittelman, D., Sharp, A., Marques-Bonet, T., Wagner, A.: Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. Genome research **25**(11), 1591-1599 (2015). doi:10.1101/gr.190868.115

186. Meneveri, R., Agresti, A., Marozzi, A., Saccone, S., Rocchi, M., Archidiacono, N., Corneo, G., Della Valle, G., Ginelli, E.: Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. Gene **123**(2), 227-234 (1993).

187. Zhang, X.Y., Loflin, P.T., Gehrke, C.W., Andrews, P.A., Ehrlich, M.: Hypermethylation of human DNA sequences in embryonal carcinoma cells and somatic tissues but not in sperm. Nucleic acids research **15**(22), 9429-9449 (1987).

188. Lyle, R., Wright, T.J., Clark, L.N., Hewitt, J.E.: 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**(3), 389-397 (1995). doi:10.1006/geno.1995.1166

189. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al.: Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature genetics **2**(1), 26-30 (1992). doi:10.1038/ng0992-26

190. van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., Frants, R.R.: FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Human molecular genetics **2**(12), 2037-2042 (1993).

191. Bodega, B., Cardone, M.F., Muller, S., Neusser, M., Orzan, F., Rossi, E., Battaglioli, E., Marozzi, A., Riva, P., Rocchi, M., Meneveri, R., Ginelli, E.: Evolutionary genomic remodelling of the human 4q subtelomere (4q35.2). BMC evolutionary biology **7**, 39 (2007). doi:10.1186/1471-2148-7-39

192. Zeng, W., Chen, Y.Y., Newkirk, D.A., Wu, B., Balog, J., Kong, X., Ball, A.R., Jr., Zanotti, S., Tawil, R., Hashimoto, N., Mortazavi, A., van der Maarel, S.M., Yokomori, K.: Genetic and epigenetic characteristics of FSHD-associated 4q and 10q D4Z4 that are distinct from non-4q/10q D4Z4 homologs. Human mutation **35**(8), 998-1010 (2014). doi:10.1002/humu.22593

193. Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., Kimonis, V.E., Balog, J., Frants, R.R., Ball,

A.R., Jr., Lock, L.F., Donovan, P.J., van der Maarel, S.M., Yokomori, K.: Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS genetics **5**(7), e1000559 (2009). doi:10.1371/journal.pgen.1000559

194. Ballarati, L., Piccini, I., Carbone, L., Archidiacono, N., Rollier, A., Marozzi, A., Meneveri, R., Ginelli, E.: Human genome dispersal and evolution of 4q35 duplications and interspersed LSau repeats. Gene **296**(1-2), 21-27 (2002).

195. Clark, L.N., Koehler, U., Ward, D.C., Wienberg, J., Hewitt, J.E.: Analysis of the organisation and localisation of the FSHD-associated tandem array in primates: implications for the origin and evolution of the 3.3 kb repeat family. Chromosoma **105**(3), 180-189 (1996).

196. Huebert, D.J., Bernstein, B.E.: Genomic views of chromatin. Current opinion in genetics & development **15**(5), 476-481 (2005). doi:10.1016/j.gde.2005.08.001

197. 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. The Journal of cell biology **167**(2), 269-279 (2004). doi:10.1083/jcb.200403128

198. Masny, P.S., Bengtsson, U., Chung, S.A., Martin, J.H., van Engelen, B., van der Maarel, S.M., Winokur, S.T.: Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope

disease? Human molecular genetics **13**(17), 1857-1871 (2004). doi:10.1093/hmg/ddh205

199. Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M.R., Cremer, T.: Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS biology **3**(5), e157 (2005). doi:10.1371/journal.pbio.0030157

200. Tawil, R., Van Der Maarel, S.M.: Facioscapulohumeral muscular dystrophy. Muscle & nerve **34**(1), 1-15 (2006). doi:10.1002/mus.20522

201. Padberg, G.W., Lunt, P.W., Koch, M., Fardeau, M.: Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD **1**(4), 231-234 (1991).

202. Daxinger, L., Tapscott, S.J., van der Maarel, S.M.: Genetic and epigenetic contributors to FSHD. Current opinion in genetics & development **33**, 56-61 (2015). doi:10.1016/j.gde.2015.08.007

203. Sacconi, S., Salviati, L., Desnuelle, C.: Facioscapulohumeral muscular dystrophy. Biochimica et biophysica acta **1852**(4), 607-614 (2015). doi:10.1016/j.bbadis.2014.05.021

204. Tawil, R., Storvick, D., Feasby, T.E., Weiffenbach, B., Griggs, R.C.: Extreme variability of expression in monozygotic twins with FSH muscular dystrophy. Neurology **43**(2), 345-348 (1993).

205. Griggs, R.C., Tawil, R., McDermott, M., Forrester, J., Figlewicz, D., Weiffenbach, B.: Monozygotic twins with facioscapulohumeral dystrophy (FSHD): implications for genotype/phenotype correlation. FSH-DY Group. Muscle & nerve. Supplement **2**, S50-55 (1995).

206. Wijmenga, C., Padberg, G.W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O.F., Milner, E.C., Weber, J.L., van Ommen, G.B., Sandkuyl, L.A., et al.: Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics **9**(4), 570-575 (1991).

207. Mathews, K.D., Mills, K.A., Bosch, E.P., Ionasescu, V.V., Wiles, K.R., Buetow, K.H., Murray, J.C.: Linkage localization of facioscapulohumeral muscular dystrophy (FSHD) in 4q35. American journal of human genetics **51**(2), 428-431 (1992).

208. Upadhyaya, M., Lunt, P., Sarfarazi, M., Broadhead, W., Farnham, J., Harper, P.S.: The mapping of chromosome 4q markers in relation to facioscapulohumeral muscular dystrophy (FSHD). American journal of human genetics **51**(2), 404-410 (1992).

209. Statland, J.M., Tawil, R.: Facioscapulohumeral Muscular Dystrophy. Continuum **22**(6, Muscle and Neuromuscular Junction Disorders), 1916-1931 (2016). doi:10.1212/CON.0000000000000399

210. van Deutekom, J.C., Bakker, E., Lemmers, R.J., van der Wielen, M.J., Bik, E., Hofker, M.H., Padberg, G.W., Frants,

R.R.: Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. Human molecular genetics **5**(12), 1997-2003 (1996).

211. Zatz, M., Marie, S.K., Passos-Bueno, M.R., Vainzof, M., Campiotto, S., Cerqueira, A., Wijmenga, C., Padberg, G., Frants, R.: High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. American journal of human genetics **56**(1), 99-105 (1995).

212. Bakker, E., Wijmenga, C., Vossen, R.H., Padberg, G.W., Hewitt, J., van der Wielen, M., Rasmussen, K., Frants, R.R.: The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle & nerve. Supplement **2**, S39-44 (1995).

213. Lemmers, R.J., Van Overveld, P.G., Sandkuijl, L.A., Vrieling, H., Padberg, G.W., Frants, R.R., van der Maarel, S.M.: Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. American journal of human genetics **75**(1), 44-53 (2004). doi:10.1086/422175

214. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., Krom, Y.D., Klooster, R., Sun, Y., den Dunnen, J.T., Helmer, Q., Donlin-Smith, C.M., Padberg, G.W., van Engelen, B.G., de Greef, J.C., Aartsma-Rus, A.M., Frants,

R.R., de Visser, M., Desnuelle, C., Sacconi, S., Filippova, G.N., Bakker, B., Bamshad, M.J., Tapscott, S.J., Miller, D.G., van der Maarel, S.M.: Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics **44**(12), 1370-1374 (2012). doi:10.1038/ng.2454

215. de Greef, J.C., Lemmers, R.J., van Engelen, B.G., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R., van der Maarel, S.M.: Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Human mutation **30**(10), 1449-1459 (2009). doi:10.1002/humu.21091

216. Pandya, S., King, W.M., Tawil, R.: Facioscapulohumeral dystrophy. Physical therapy **88**(1), 105-113 (2008). doi:10.2522/ptj.20070104

217. Tawil, R., Kissel, J.T., Heatwole, C., Pandya, S., Gronseth, G., Benatar, M., Guideline Development, D., Implementation Subcommittee of the American Academy of, N., Practice Issues Review Panel of the American Association of, N., Electrodiagnostic, M.: Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the Practice Issues Review Panel of the American Association of Neuromuscular &

Electrodiagnostic Medicine. Neurology **85**(4), 357-364 (2015). doi:10.1212/WNL.0000000000001783

218. Laforet, P., de Toma, C., Eymard, B., Becane, H.M., Jeanpierre, M., Fardeau, M., Duboc, D.: Cardiac involvement in genetically confirmed facioscapulohumeral muscular dystrophy. Neurology **51**(5), 1454-1456 (1998).

219. Trevisan, C.P., Pastorello, E., Armani, M., Angelini, C., Nante, G., Tomelleri, G., Tonin, P., Mongini, T., Palmucci, L., Galluzzi, G., Tupler, R.G., Barchitta, A.: Facioscapulohumeral muscular dystrophy and occurrence of heart arrhythmia. European neurology **56**(1), 1-5 (2006). doi:10.1159/000094248

220. Statland, J.M., Sacconi, S., Farmakidis, C., Donlin-Smith, C.M., Chung, M., Tawil, R.: Coats syndrome in facioscapulohumeral dystrophy type 1: frequency and D4Z4 contraction size. Neurology **80**(13), 1247-1250 (2013). doi:10.1212/WNL.0b013e3182897116

221. Fitzsimons, R.B., Gurwin, E.B., Bird, A.C.: Retinal vascular abnormalities in facioscapulohumeral muscular dystrophy. A general association with genetic and therapeutic implications. Brain : a journal of neurology **110 (Pt 3)**, 631-648 (1987).

222. Padberg, G.W., Brouwer, O.F., de Keizer, R.J., Dijkman, G., Wijmenga, C., Grote, J.J., Frants, R.R.: On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. Muscle & nerve. Supplement(2), S73-80 (1995).

223. Trevisan, C.P., Pastorello, E., Tomelleri, G., Vercelli, L., Bruno, C., Scapolan, S., Siciliano, G., Comacchio, F.: Facioscapulohumeral muscular dystrophy: hearing loss and other atypical features of patients with large 4q35 deletions. European journal of neurology **15**(12), 1353-1358 (2008). doi:10.1111/j.1468-1331.2008.02314.x

224. Lutz, K.L., Holte, L., Kliethermes, S.A., Stephan, C., Mathews, K.D.: Clinical and genetic features of hearing loss in facioscapulohumeral muscular dystrophy. Neurology **81**(16), 1374-1377 (2013). doi:10.1212/WNL.0b013e3182a84140

225. Trevisan, C.P., Pastorello, E., Ermani, M., Angelini, C., Tomelleri, G., Tonin, P., Mongini, T., Palmucci, L., Galluzzi, G., Tupler, R.G., Marioni, G., Rimini, A.: Facioscapulohumeral muscular dystrophy: a multicenter study on hearing function. Audiology & neuro-otology **13**(1), 1-6 (2008). doi:10.1159/000107431

226. Funakoshi, M., Goto, K., Arahata, K.: Epilepsy and mental retardation in a subset of early onset 4q35-facioscapulohumeral muscular dystrophy. Neurology **50**(6), 1791-1794 (1998).

227. Saito, Y., Miyashita, S., Yokoyama, A., Komaki, H., Seki, A., Maegaki, Y., Ohno, K.: Facioscapulohumeral muscular dystrophy with severe mental retardation and epilepsy. Brain & development **29**(4), 231-233 (2007). doi:10.1016/j.braindev.2006.08.012

228. Miura, K., Kumagai, T., Matsumoto, A., Iriyama, E., Watanabe, K., Goto, K., Arahata, K.: Two cases of

chromosome 4q35-linked early onset facioscapulohumeral muscular dystrophy with mental retardation and epilepsy. Neuropediatrics **29**(5), 239-241 (1998). doi:10.1055/s-2007-973568

229. Chen, T.H., Lai, Y.H., Lee, P.L., Hsu, J.H., Goto, K., Hayashi, Y.K., Nishino, I., Lin, C.W., Shih, H.H., Huang, C.C., Liang, W.C., Wang, W.F., Jong, Y.J.: Infantile facioscapulohumeral muscular dystrophy revisited: Expansion of clinical phenotypes in patients with a very short EcoRI fragment. Neuromuscular disorders : NMD **23**(4), 298-305 (2013). doi:10.1016/j.nmd.2013.01.005

230. Cabianca, D.S., Gabellini, D.: The cell biology of disease: FSHD: copy number variations on the theme of muscular dystrophy. The Journal of cell biology **191**(6), 1049-1060 (2010). doi:10.1083/jcb.201007028

231. Tawil, R., McDermott, M.P., Mendell, J.R., Kissel, J., Griggs, R.C.: Facioscapulohumeral muscular dystrophy (FSHD): design of natural history study and results of baseline testing. FSH-DY Group. Neurology **44**(3 Pt 1), 442-446 (1994).

232. Statland, J.M., Tawil, R.: Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. Current opinion in neurology **24**(5), 423-428 (2011). doi:10.1097/WCO.0b013e32834959af

233. Zatz, M., Marie, S.K., Cerqueira, A., Vainzof, M., Pavanello, R.C., Passos-Bueno, M.R.: The facioscapulohumeral muscular dystrophy (FSHD1) gene affects

males more severely and more frequently than females. American journal of medical genetics **77**(2), 155-161 (1998).
234. Tonini, M.M., Passos-Bueno, M.R., Cerqueira, A., Matioli, S.R., Pavanello, R., Zatz, M.: Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromuscular disorders : NMD **14**(1), 33-38 (2004).
235. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., Desnuelle, C., Spuler, S., Tarnopolsky, M., Venance, S.L., Frants, R.R., van der Maarel, S.M., Tawil, R.: Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology **75**(17), 1548-1554 (2010). doi:10.1212/WNL.0b013e3181f96175
236. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., Casarin, A., Pegoraro, E., Tawil, R., Bakker, E., Tapscott, S.J., Desnuelle, C., van der Maarel, S.M.: The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics **93**(4), 744-751 (2013). doi:10.1016/j.ajhg.2013.08.004
237. Hewitt, J.E.: Loss of epigenetic silencing of the DUX4 transcription factor gene in facioscapulohumeral muscular dystrophy. Human molecular genetics **24**(R1), R17-23 (2015). doi:10.1093/hmg/ddv237

238. van der Maarel, S.M., Tawil, R., Tapscott, S.J.: Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends in molecular medicine **17**(5), 252-258 (2011). doi:10.1016/j.molmed.2011.01.001

239. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., Miller, D.G.: Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics **6**(10), e1001181 (2010). doi:10.1371/journal.pgen.1001181

240. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., Miller, D.G., Upadhyaya, M., Verschuuren, J.J., Lopez de Munain Arregui, A., van Engelen, B.G., Padberg, G.W., Sacconi, S., Tawil, R., Tapscott, S.J., Bakker, B., van der Maarel, S.M.: Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics **24**(3), 659-669 (2015). doi:10.1093/hmg/ddu486

241. Mitsuhashi, S., Boyden, S.E., Estrella, E.A., Jones, T.I., Rahimov, F., Yu, T.W., Darras, B.T., Amato, A.A., Folkerth, R.D., Jones, P.L., Kunkel, L.M., Kang, P.B.: Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscular disorders : NMD **23**(12), 975-980 (2013). doi:10.1016/j.nmd.2013.08.009

242. Winston, J., Duerden, L., Mort, M., Frayling, I.M., Rogers, M.T., Upadhyaya, M.: Identification of two novel SMCHD1 sequence variants in families with FSHD-like muscular dystrophy. European journal of human genetics : EJHG **23**(1), 67-71 (2015). doi:10.1038/ejhg.2014.58

243. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., Muller, C.R.: Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG **23**(6), 808-816 (2015). doi:10.1038/ejhg.2014.191

244. van den Boogaard, M.L., Lemmers, R.J., Camano, P., van der Vliet, P.J., Voermans, N., van Engelen, B.G., Lopez de Munain, A., Tapscott, S.J., van der Stoep, N., Tawil, R., van der Maarel, S.M.: Double SMCHD1 variants in FSHD2: the synergistic effect of two SMCHD1 variants on D4Z4 hypomethylation and disease penetrance in FSHD2. European journal of human genetics : EJHG **24**(1), 78-85 (2016). doi:10.1038/ejhg.2015.55

245. Lemmers, R.J., van den Boogaard, M.L., van der Vliet, P.J., Donlin-Smith, C.M., Nations, S.P., Ruivenkamp, C.A., Heard, P., Bakker, B., Tapscott, S., Cody, J.D., Tawil, R., van der Maarel, S.M.: Hemizygosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. Human mutation **36**(7), 679-683 (2015). doi:10.1002/humu.22792

246. Tawil, R., van der Maarel, S.M., Tapscott, S.J.: Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. Skeletal muscle **4**, 12 (2014). doi:10.1186/2044-5040-4-12

247. Chuang, P.T., Albertson, D.G., Meyer, B.J.: DPY-27:a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X chromosome. Cell **79**(3), 459-474 (1994).

248. Dej, K.J., Ahn, C., Orr-Weaver, T.L.: Mutations in the Drosophila condensin subunit dCAP-G: defining the role of condensin for chromosome condensation in mitosis and gene expression in interphase. Genetics **168**(2), 895-906 (2004). doi:10.1534/genetics.104.030908

249. Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M., Matzke, A.J.: A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nature genetics **40**(5), 670-675 (2008). doi:10.1038/ng.119

250. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., Kay, G.F., Whitelaw, E.: SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics **40**(5), 663-669 (2008). doi:10.1038/ng.142

251. Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., Giannoulatou, E., Heard, E., Stancheva, I., Brockdorff, N.: Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental cell **23**(2), 265-279 (2012). doi:10.1016/j.devcel.2012.06.011

252. Mould, A.W., Pang, Z., Pakusch, M., Tonks, I.D., Stark, M., Carrie, D., Mukhopadhyay, P., Seidel, A., Ellis, J.J., Deakin, J., Wakefield, M.J., Krause, L., Blewitt, M.E., Kay, G.F.: Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics & chromatin **6**(1), 19 (2013). doi:10.1186/1756-8935-6-19

253. Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M., Heard, E., Brockdorff, N.: Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Molecular and cellular biology **33**(16), 3150-3165 (2013). doi:10.1128/MCB.00145-13

254. Lemmers, R.J., de Kievit, P., Sandkuijl, L., Padberg, G.W., van Ommen, G.J., Frants, R.R., van der Maarel, S.M.: Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nature genetics **32**(2), 235-236 (2002). doi:10.1038/ng999

255. Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M., et al.: Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Human molecular genetics **3**(8), 1287-1295 (1994).

256. Winokur, S.T., Bengtsson, U., Feddersen, J., Mathews, K.D., Weiffenbach, B., Bailey, H., Markovich, R.P., Murray, J.C., Wasmuth, J.J., Altherr, M.R., et al.: The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology **2**(3), 225-234 (1994).

257. Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M., de Knijff, P., Padberg, G.W., Frants, R.R., van der Maarel, S.M.: Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. American journal of human genetics **81**(5), 884-894 (2007). doi:10.1086/521986

258. Lemmers, R.J., Wohlgemuth, M., Frants, R.R., Padberg, G.W., Morava, E., van der Maarel, S.M.: Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. American journal of human genetics **75**(6), 1124-1130 (2004). doi:10.1086/426035

259. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., Miller, D.G., Tapscott, S.J., Tawil, R., Frants, R.R., van der Maarel, S.M.: A unifying genetic model for facioscapulohumeral muscular dystrophy. Science **329**(5999), 1650-1653 (2010). doi:10.1126/science.1189044

260. Lemmers, R.J., van der Vliet, P.J., van der Gaag, K.J., Zuniga, S., Frants, R.R., de Knijff, P., van der Maarel, S.M.: Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. American journal of human genetics **86**(3), 364-377 (2010). doi:10.1016/j.ajhg.2010.01.035

261. Spurlock, G., Jim, H.P., Upadhyaya, M.: Confirmation that the specific SSLP microsatellite allele 4qA161 segregates with fascioscapulohumeral muscular dystrophy (FSHD) in a cohort of multiplex and simplex FSHD families. Muscle & nerve **42**(5), 820-821 (2010). doi:10.1002/mus.21766

262. Tsumagari, K., Chang, S.C., Lacey, M., Baribault, C., Chittur, S.V., Sowden, J., Tawil, R., Crawford, G.E., Ehrlich, M.: Gene expression during normal and FSHD myogenesis. BMC medical genomics **4**, 67 (2011). doi:10.1186/1755-8794-4-67

263. Thomas, N.S., Wiseman, K., Spurlock, G., MacDonald, M., Ustek, D., Upadhyaya, M.: A large patient study confirming that facioscapulohumeral muscular dystrophy (FSHD) disease expression is almost exclusively associated with an FSHD locus

located on a 4qA-defined 4qter subtelomere. Journal of medical genetics **44**(3), 215-218 (2007). doi:10.1136/jmg.2006.042804

264. Bakker, E., Wijmenga, C., Vossen, R.H., Padberg, G.W., Hewitt, J., van der Wielen, M., Rasmussen, K., Frants, R.R.: The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle & nerve. Supplement(2), S39-44 (1995).

265. van Geel, M., Dickson, M.C., Beck, A.F., Bolland, D.J., Frants, R.R., van der Maarel, S.M., de Jong, P.J., Hewitt, J.E.: Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics **79**(2), 210-217 (2002). doi:10.1006/geno.2002.6690

266. Lemmers, R.J.L., de Kievit, P., van Geel, M., van der Wielen, M.J., Bakker, E., Padberg, G.W., Frants, R.R., van der Maarel, S.M.: Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. Annals of neurology **50**(6), 816-819 (2001).

267. Lunt, P.W., Jardine, P.E., Koch, M.C., Maynard, J., Osborn, M., Williams, M., Harper, P.S., Upadhyaya, M.: Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). Human molecular genetics **4**(5), 951-958 (1995).

268. Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V.,

Silvestri, G., Mirabella, M., Mangiola, F., Tonali, P., Felicetti, L.: Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype. Annals of neurology **45**(6), 751-757 (1999).

269. Tawil, R., Forrester, J., Griggs, R.C., Mendell, J., Kissel, J., McDermott, M., King, W., Weiffenbach, B., Figlewicz, D.: Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. Annals of neurology **39**(6), 744-748 (1996). doi:10.1002/ana.410390610

270. Lunt, P.W., Jardine, P.E., Koch, M., Maynard, J., Osborn, M., Williams, M., Harper, P.S., Upadhyaya, M.: Phenotypic-genotypic correlation will assist genetic counseling in 4q35-facioscapulohumeral muscular dystrophy. Muscle & nerve. Supplement(2), S103-109 (1995).

271. Goto, K., Lee, J.H., Matsuda, C., Hirabayashi, K., Kojo, T., Nakamura, A., Mitsunaga, Y., Furukawa, T., Sahashi, K., Arahata, K.: DNA rearrangements in Japanese facioscapulohumeral muscular dystrophy patients: clinical correlations. Neuromuscular disorders : NMD **5**(3), 201-208 (1995).

272. Scionti, I., Greco, F., Ricci, G., Govi, M., Arashiro, P., Vercelli, L., Berardinelli, A., Angelini, C., Antonini, G., Cao, M., Di Muzio, A., Moggio, M., Morandi, L., Ricci, E., Rodolico, C., Ruggiero, L., Santoro, L., Siciliano, G., Tomelleri, G., Trevisan,

C.P., Galluzzi, G., Wright, W., Zatz, M., Tupler, R.: Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. American journal of human genetics **90**(4), 628-635 (2012). doi:10.1016/j.ajhg.2012.02.019

273. Tupler, R., Berardinelli, A., Barbierato, L., Frants, R., Hewitt, J.E., Lanzi, G., Maraschio, P., Tiepolo, L.: Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. Journal of medical genetics **33**(5), 366-370 (1996).

274. Wijmenga, C., Frants, R.R., Hewitt, J.E., van Deutekom, J.C., van Geel, M., Wright, T.J., Padberg, G.W., Hofker, M.H., van Ommen, G.J.: Molecular genetics of facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD **3**(5-6), 487-491 (1993).

275. Cabianca, D.S., Casa, V., Gabellini, D.: A novel molecular mechanism in human genetic disease: a DNA repeat-derived lncRNA. RNA biology **9**(10), 1211-1217 (2012). doi:10.4161/rna.21922

276. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., van der Maarel, S.M.: Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics **7**(6), 579-584 (2012). doi:10.4161/epi.20001

277. van Overveld, P.G., Lemmers, R.J., Deidda, G., Sandkuijl, L., Padberg, G.W., Frants, R.R., van der Maarel, S.M.: Interchromosomal repeat array interactions between

chromosomes 4 and 10: a model for subtelomeric plasticity. Human molecular genetics **9**(19), 2879-2884 (2000).

278. Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., Gabellini, D.: A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell **149**(4), 819-831 (2012). doi:10.1016/j.cell.2012.03.035

279. 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**(3), 339-348 (2002).

280. Mihaly, J., Mishra, R.K., Karch, F.: A conserved sequence motif in Polycomb-response elements. Molecular cell **1**(7), 1065-1066 (1998).

281. Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., Sakonju, S.: The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. Development **128**(11), 2163-2173 (2001).

282. Mishra, R.K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S.E., Schedl, P.: The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. Molecular and cellular biology **21**(4), 1311-1318 (2001). doi:10.1128/MCB.21.4.1311-1318.2001

283. Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., Wysocka, J.: Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell **139**(7), 1290-1302 (2009). doi:10.1016/j.cell.2009.12.002

284. Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., Orkin, S.H.: Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell **139**(7), 1303-1314 (2009). doi:10.1016/j.cell.2009.12.003

285. Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., Brookes, E., Pombo, A., Fisher, C., Skarnes, W.C., Snoek, T., Bezstarosti, K., Demmers, J., Klose, R.J., Casanova, M., Tavares, L., Brockdorff, N., Merkenschlager, M., Fisher, A.G.: Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nature cell biology **12**(6), 618-624 (2010). doi:10.1038/ncb2065

286. Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., Reinberg, D.: Jarid2 and PRC2, partners in regulating gene expression. Genes & development **24**(4), 368-380 (2010). doi:10.1101/gad.1886410

287. Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., Helin, K.: JARID2 regulates binding of the

Polycomb repressive complex 2 to target genes in ES cells. Nature **464**(7286), 306-310 (2010). doi:10.1038/nature08788

288. Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., Cavalli, G.: Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature **434**(7032), 533-538 (2005). doi:10.1038/nature03386

289. Matharu, N.K., Hussain, T., Sankaranarayanan, R., Mishra, R.K.: Vertebrate homologue of Drosophila GAGA factor. Journal of molecular biology **400**(3), 434-447 (2010). doi:10.1016/j.jmb.2010.05.010

290. Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H., Di Croce, L.: The histone variant macroH2A is an epigenetic regulator of key developmental genes. Nature structural & molecular biology **16**(10), 1074-1079 (2009). doi:10.1038/nsmb.1665

291. Jiang, G., Yang, F., van Overveld, P.G., Vedanarayanan, V., van der Maarel, S., Ehrlich, M.: Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Human molecular genetics **12**(22), 2909-2921 (2003). doi:10.1093/hmg/ddg323

292. Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R., van der Maarel, S.M., Filippova, G.N., Tapscott, S.J.: RNA transcripts, miRNA-sized fragments and proteins

produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Human molecular genetics **18**(13), 2414-2430 (2009). doi:10.1093/hmg/ddp180

293. Ottaviani, A., Schluth-Bolard, C., Rival-Gervier, S., Boussouar, A., Rondier, D., Foerster, A.M., Morere, J., Bauwens, S., Gazzo, S., Callet-Bauchu, E., Gilson, E., Magdinier, F.: Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. The EMBO journal **28**(16), 2428-2436 (2009). doi:10.1038/emboj.2009.201

294. Filippova, G.N.: Genetics and epigenetics of the multifunctional protein CTCF. Current topics in developmental biology **80**, 337-360 (2008). doi:10.1016/S0070-2153(07)80009-3

295. Ottaviani, A., Rival-Gervier, S., Boussouar, A., Foerster, A.M., Rondier, D., Sacconi, S., Desnuelle, C., Gilson, E., Magdinier, F.: The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. PLoS genetics **5**(2), e1000394 (2009). doi:10.1371/journal.pgen.1000394

296. van Geel, M., Heather, L.J., Lyle, R., Hewitt, J.E., Frants, R.R., de Jong, P.J.: The FSHD region on human chromosome 4q35 contains potential coding regions among pseudogenes and a high density of repeat elements. Genomics **61**(1), 55-65 (1999). doi:10.1006/geno.1999.5942

297. Blair, I.P., Adams, L.J., Badenhop, R.F., Moses, M.J., Scimone, A., Morris, J.A., Ma, L., Austin, C.P., Donald, J.A., Mitchell, P.B., Schofield, P.R.: A transcript map encompassing a susceptibility locus for bipolar affective disorder on chromosome 4q35. Molecular psychiatry **7**(8), 867-873 (2002). doi:10.1038/sj.mp.4001113

298. Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R., Green, M.R., Tupler, R.: Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature **439**(7079), 973-977 (2006). doi:10.1038/nature04422

299. van Koningsbruggen, S., Dirks, R.W., Mommaas, A.M., Onderwater, J.J., Deidda, G., Padberg, G.W., Frants, R.R., van der Maarel, S.M.: FRG1P is localised in the nucleolus, Cajal bodies, and speckles. Journal of medical genetics **41**(4), e46 (2004).

300. van Koningsbruggen, S., Straasheijm, K.R., Sterrenburg, E., de Graaf, N., Dauwerse, H.G., Frants, R.R., van der Maarel, S.M.: FRG1P-mediated aggregation of proteins involved in pre-mRNA processing. Chromosoma **116**(1), 53-64 (2007). doi:10.1007/s00412-006-0083-3

301. Bodega, B., Ramirez, G.D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli, E., Ginelli, E.: 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 biology **7**, 41 (2009). doi:10.1186/1741-7007-7-41

302. Osborne, R.J., Welle, S., Venance, S.L., Thornton, C.A., Tawil, R.: Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. Neurology **68**(8), 569-577 (2007). doi:10.1212/01.wnl.0000251269.31442.d9

303. Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., Belayew, A.: Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene **236**(1), 25-32 (1999).

304. Leidenroth, A., Hewitt, J.E.: A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. BMC evolutionary biology **10**, 364 (2010). doi:10.1186/1471-2148-10-364

305. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., van der Maarel, S.M.: Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nature genetics **35**(4), 315-317 (2003). doi:10.1038/ng1262

306. Tassin, A., Laoudj-Chenivesse, D., Vanderplanck, C., Barro, M., Charron, S., Ansseau, E., Chen, Y.W., Mercier, J., Coppee, F., Belayew, A.: DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? Journal

of cellular and molecular medicine **17**(1), 76-89 (2013). doi:10.1111/j.1582-4934.2012.01647.x

307. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., Tapscott, S.J.: DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell **22**(1), 38-51 (2012). doi:10.1016/j.devcel.2011.11.013

308. Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., Tapscott, S.J.: DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS genetics **9**(11), e1003947 (2013). doi:10.1371/journal.pgen.1003947

309. Rickard, A.M., Petek, L.M., Miller, D.G.: Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Human molecular genetics **24**(20), 5901-5914 (2015). doi:10.1093/hmg/ddv315

310. Gatica, L.V., Rosa, A.L.: A complex interplay of genetic and epigenetic events leads to abnormal expression of the DUX4 gene in facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD **26**(12), 844-852 (2016). doi:10.1016/j.nmd.2016.09.015

311. Sharma, V., Harafuji, N., Belayew, A., Chen, Y.W.: DUX4 differentially regulates transcriptomes of human

rhabdomyosarcoma and mouse C2C12 cells. PloS one **8**(5), e64691 (2013). doi:10.1371/journal.pone.0064691

312. Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S.M., Tapscott, S.J., Bradley, R.K.: A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. eLife **4** (2015). doi:10.7554/eLife.04996

313. Pastor, F., Kolonias, D., Giangrande, P.H., Gilboa, E.: Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature **465**(7295), 227-230 (2010). doi:10.1038/nature08999

314. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., Wattiez, R., Figlewicz, D., Laoudj-Chenivesse, D., Belayew, A., Coppee, F., Rosa, A.L.: The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular disorders : NMD **17**(8), 611-623 (2007). doi:10.1016/j.nmd.2007.04.002

315. Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R., Kyba, M.: DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Experimental neurology **214**(1), 87-96 (2008). doi:10.1016/j.expneurol.2008.07.022

316. Vanderplanck, C., Ansseau, E., Charron, S., Stricwant, N., Tassin, A., Laoudj-Chenivesse, D., Wilton, S.D., Coppee, F., Belayew, A.: The FSHD atrophic myotube phenotype is caused

by DUX4 expression. PloS one **6**(10), e26820 (2011). doi:10.1371/journal.pone.0026820

317. Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., Harper, S.Q.: DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Annals of neurology **69**(3), 540-552 (2011). doi:10.1002/ana.22275

318. Krom, Y.D., Thijssen, P.E., Young, J.M., den Hamer, B., Balog, J., Yao, Z., Maves, L., Snider, L., Knopp, P., Zammit, P.S., Rijkers, T., van Engelen, B.G., Padberg, G.W., Frants, R.R., Tawil, R., Tapscott, S.J., van der Maarel, S.M.: Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. PLoS genetics **9**(4), e1003415 (2013). doi:10.1371/journal.pgen.1003415

319. Murton, A.J., Constantin, D., Greenhaff, P.L.: The involvement of the ubiquitin proteasome system in human skeletal muscle remodelling and atrophy. Biochimica et biophysica acta **1782**(12), 730-743 (2008). doi:10.1016/j.bbadis.2008.10.011

320. Bodine, S.C., Baehr, L.M.: Skeletal muscle atrophy and the E3 ubiquitin ligases MuRF1 and MAFbx/atrogin-1. American journal of physiology. Endocrinology and metabolism **307**(6), E469-484 (2014). doi:10.1152/ajpendo.00204.2014

321. Hasselgren, P.O., Pedersen, P., Sax, H.C., Warner, B.W., Fischer, J.E.: Current concepts of protein turnover and amino

acid transport in liver and skeletal muscle during sepsis. Archives of surgery **123**(8), 992-999 (1988).

322. Lecker, S.H., Jagoe, R.T., Gilbert, A., Gomes, M., Baracos, V., Bailey, J., Price, S.R., Mitch, W.E., Goldberg, A.L.: Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **18**(1), 39-51 (2004). doi:10.1096/fj.03-0610com

323. Gomes, M.D., Lecker, S.H., Jagoe, R.T., Navon, A., Goldberg, A.L.: Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy. Proceedings of the National Academy of Sciences of the United States of America **98**(25), 14440-14445 (2001). doi:10.1073/pnas.251541198

324. Lecker, S.H., Solomon, V., Mitch, W.E., Goldberg, A.L.: Muscle protein breakdown and the critical role of the ubiquitin-proteasome pathway in normal and disease states. The Journal of nutrition **129**(1S Suppl), 227S-237S (1999).

325. Mitch, W.E., Goldberg, A.L.: Mechanisms of muscle wasting. The role of the ubiquitin-proteasome pathway. The New England journal of medicine **335**(25), 1897-1905 (1996). doi:10.1056/NEJM199612193352507

326. Li, Y.P., Chen, Y., John, J., Moylan, J., Jin, B., Mann, D.L., Reid, M.B.: TNF-alpha acts via p38 MAPK to stimulate expression of the ubiquitin ligase atrogin1/MAFbx in skeletal muscle. FASEB journal : official publication of the Federation of

American Societies for Experimental Biology **19**(3), 362-370 (2005). doi:10.1096/fj.04-2364com

327. Cai, D., Frantz, J.D., Tawa, N.E., Jr., Melendez, P.A., Oh, B.C., Lidov, H.G., Hasselgren, P.O., Frontera, W.R., Lee, J., Glass, D.J., Shoelson, S.E.: IKKbeta/NF-kappaB activation causes severe muscle wasting in mice. Cell **119**(2), 285-298 (2004). doi:10.1016/j.cell.2004.09.027

328. Foletta, V.C., White, L.J., Larsen, A.E., Leger, B., Russell, A.P.: The role and regulation of MAFbx/atrogin-1 and MuRF1 in skeletal muscle atrophy. Pflugers Archiv : European journal of physiology **461**(3), 325-335 (2011). doi:10.1007/s00424-010-0919-9

329. Bodine, S.C., Latres, E., Baumhueter, S., Lai, V.K., Nunez, L., Clarke, B.A., Poueymirou, W.T., Panaro, F.J., Na, E., Dharmarajan, K., Pan, Z.Q., Valenzuela, D.M., DeChiara, T.M., Stitt, T.N., Yancopoulos, G.D., Glass, D.J.: Identification of ubiquitin ligases required for skeletal muscle atrophy. Science **294**(5547), 1704-1708 (2001). doi:10.1126/science.1065874

330. Jaspers, S.R., Tischler, M.E.: Atrophy and growth failure of rat hindlimb muscles in tail-cast suspension. Journal of applied physiology: respiratory, environmental and exercise physiology **57**(5), 1472-1479 (1984).

331. Tischler, M.E., Rosenberg, S., Satarug, S., Henriksen, E.J., Kirby, C.R., Tome, M., Chase, P.: Different mechanisms of increased proteolysis in atrophy induced by denervation or

unweighting of rat soleus muscle. Metabolism: clinical and experimental **39**(7), 756-763 (1990).

332. Lang, C.H., Huber, D., Frost, R.A.: Burn-induced increase in atrogin-1 and MuRF-1 in skeletal muscle is glucocorticoid independent but downregulated by IGF-I. American journal of physiology. Regulatory, integrative and comparative physiology **292**(1), R328-336 (2007). doi:10.1152/ajpregu.00561.2006

333. Clarke, B.A., Drujan, D., Willis, M.S., Murphy, L.O., Corpina, R.A., Burova, E., Rakhilin, S.V., Stitt, T.N., Patterson, C., Latres, E., Glass, D.J.: The E3 Ligase MuRF1 degrades myosin heavy chain protein in dexamethasone-treated skeletal muscle. Cell metabolism **6**(5), 376-385 (2007). doi:10.1016/j.cmet.2007.09.009

334. Frost, R.A., Nystrom, G.J., Jefferson, L.S., Lang, C.H.: Hormone, cytokine, and nutritional regulation of sepsis-induced increases in atrogin-1 and MuRF1 in skeletal muscle. American journal of physiology. Endocrinology and metabolism **292**(2), E501-512 (2007). doi:10.1152/ajpendo.00359.2006

335. Goodman, C.A., McNally, R.M., Hoffmann, F.M., Hornberger, T.A.: Smad3 induces atrogin-1, inhibits mTOR and protein synthesis, and promotes muscle atrophy in vivo. Molecular endocrinology **27**(11), 1946-1957 (2013). doi:10.1210/me.2013-1194

336. Attaix, D., Combaret, L., Pouch, M.N., Taillandier, D.: Regulation of proteolysis. Current opinion in clinical nutrition and metabolic care **4**(1), 45-49 (2001).

337. Hasselgren, P.O.: Glucocorticoids and muscle catabolism. Current opinion in clinical nutrition and metabolic care **2**(3), 201-205 (1999).

338. Du, J., Wang, X., Miereles, C., Bailey, J.L., Debigare, R., Zheng, B., Price, S.R., Mitch, W.E.: Activation of caspase-3 is an initial step triggering accelerated muscle proteolysis in catabolic conditions. The Journal of clinical investigation **113**(1), 115-123 (2004). doi:10.1172/JCI18330

339. Nunez, G., Benedict, M.A., Hu, Y., Inohara, N.: Caspases: the proteases of the apoptotic pathway. Oncogene **17**(25), 3237-3245 (1998). doi:10.1038/sj.onc.1202581

340. Bechet, D., Tassa, A., Taillandier, D., Combaret, L., Attaix, D.: Lysosomal proteolysis in skeletal muscle. The international journal of biochemistry & cell biology **37**(10), 2098-2114 (2005). doi:10.1016/j.biocel.2005.02.029

341. Bartoli, M., Richard, I.: Calpains in muscle wasting. The international journal of biochemistry & cell biology **37**(10), 2115-2133 (2005). doi:10.1016/j.biocel.2004.12.012

342. Costelli, P., Reffo, P., Penna, F., Autelli, R., Bonelli, G., Baccino, F.M.: Ca(2+)-dependent proteolysis in muscle wasting. The international journal of biochemistry & cell biology **37**(10), 2134-2146 (2005). doi:10.1016/j.biocel.2005.03.010

343. Jackman, R.W., Kandarian, S.C.: The molecular basis of skeletal muscle atrophy. American journal of physiology. Cell physiology **287**(4), C834-843 (2004). doi:10.1152/ajpcell.00579.2003

344. Attaix, D., Ventadour, S., Codran, A., Bechet, D., Taillandier, D., Combaret, L.: The ubiquitin-proteasome system and skeletal muscle wasting. Essays in biochemistry **41**, 173-186 (2005). doi:10.1042/EB0410173

345. Mitch, W.E., Medina, R., Grieber, S., May, R.C., England, B.K., Price, S.R., Bailey, J.L., Goldberg, A.L.: Metabolic acidosis stimulates muscle protein degradation by activating the adenosine triphosphate-dependent pathway involving ubiquitin and proteasomes. The Journal of clinical investigation **93**(5), 2127-2133 (1994). doi:10.1172/JCI117208

346. Price, S.R., Bailey, J.L., Wang, X., Jurkovitz, C., England, B.K., Ding, X., Phillips, L.S., Mitch, W.E.: Muscle wasting in insulinopenic rats results from activation of the ATP-dependent, ubiquitin-proteasome proteolytic pathway by a mechanism including gene transcription. The Journal of clinical investigation **98**(8), 1703-1708 (1996). doi:10.1172/JCI118968

347. Wing, S.S., Goldberg, A.L.: Glucocorticoids activate the ATP-ubiquitin-dependent proteolytic system in skeletal muscle during fasting. The American journal of physiology **264**(4 Pt 1), E668-676 (1993).

348. Hershko, A., Ciechanover, A.: The ubiquitin system. Annual review of biochemistry **67**, 425-479 (1998). doi:10.1146/annurev.biochem.67.1.425

349. Metzger, M.B., Hristova, V.A., Weissman, A.M.: HECT and RING finger families of E3 ubiquitin ligases at a glance. Journal

of cell science **125**(Pt 3), 531-537 (2012). doi:10.1242/jcs.091777

350. Navon, A., Ciechanover, A.: The 26 S proteasome: from basic mechanisms to drug targeting. The Journal of biological chemistry **284**(49), 33713-33718 (2009). doi:10.1074/jbc.R109.018481

351. Passmore, L.A., Barford, D.: Getting into position: the catalytic mechanisms of protein ubiquitylation. The Biochemical journal **379**(Pt 3), 513-525 (2004). doi:10.1042/BJ20040198

352. Kravtsova-Ivantsiv, Y., Ciechanover, A.: Non-canonical ubiquitin-based signals for proteasomal degradation. Journal of cell science **125**(Pt 3), 539-548 (2012). doi:10.1242/jcs.093567

353. Thrower, J.S., Hoffman, L., Rechsteiner, M., Pickart, C.M.: Recognition of the polyubiquitin proteolytic signal. The EMBO journal **19**(1), 94-102 (2000). doi:10.1093/emboj/19.1.94

354. Rivett, A.J., Mason, G.G., Murray, R.Z., Reidlinger, J.: Regulation of proteasome structure and function. Molecular biology reports **24**(1-2), 99-102 (1997).

355. Tomkinson, B., Lindas, A.C.: Tripeptidyl-peptidase II: a multi-purpose peptidase. The international journal of biochemistry & cell biology **37**(10), 1933-1937 (2005). doi:10.1016/j.biocel.2005.02.009

356. Etlinger, J.D., Goldberg, A.L.: A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. Proceedings of the National Academy

of Sciences of the United States of America **74**(1), 54-58 (1977).

357. Centner, T., Yano, J., Kimura, E., McElhinny, A.S., Pelin, K., Witt, C.C., Bang, M.L., Trombitas, K., Granzier, H., Gregorio, C.C., Sorimachi, H., Labeit, S.: Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain. Journal of molecular biology **306**(4), 717-726 (2001). doi:10.1006/jmbi.2001.4448

358. Skaar, J.R., Pagan, J.K., Pagano, M.: Mechanisms and function of substrate recruitment by F-box proteins. Nature reviews. Molecular cell biology **14**(6), 369-381 (2013). doi:10.1038/nrm3582

359. Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., Harper, J.W.: F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. Cell **91**(2), 209-219 (1997).

360. Kipreos, E.T., Pagano, M.: The F-box protein family. Genome biology **1**(5), REVIEWS3002 (2000). doi:10.1186/gb-2000-1-5-reviews3002

361. Cardozo, T., Pagano, M.: The SCF ubiquitin ligase: insights into a molecular machine. Nature reviews. Molecular cell biology **5**(9), 739-751 (2004). doi:10.1038/nrm1471

362. Petroski, M.D., Deshaies, R.J.: Function and regulation of cullin-RING ubiquitin ligases. Nature reviews. Molecular cell biology **6**(1), 9-20 (2005). doi:10.1038/nrm1547

363. Ponting, C.P., Phillips, C., Davies, K.E., Blake, D.J.: PDZ domains: targeting signalling molecules to sub-membranous sites. BioEssays : news and reviews in molecular, cellular and developmental biology **19**(6), 469-479 (1997). doi:10.1002/bies.950190606

364. Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., MacKinnon, R.: Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell **85**(7), 1067-1076 (1996).

365. Tintignac, L.A., Lagirand, J., Batonnet, S., Sirri, V., Leibovitch, M.P., Leibovitch, S.A.: Degradation of MyoD mediated by the SCF (MAFbx) ubiquitin ligase. The Journal of biological chemistry **280**(4), 2847-2856 (2005). doi:10.1074/jbc.M411346200

366. Blondel, M., Galan, J.M., Chi, Y., Lafourcade, C., Longaretti, C., Deshaies, R.J., Peter, M.: Nuclear-specific degradation of Far1 is controlled by the localization of the F-box protein Cdc4. The EMBO journal **19**(22), 6085-6097 (2000). doi:10.1093/emboj/19.22.6085

367. Li, Y.P., Chen, Y., Li, A.S., Reid, M.B.: Hydrogen peroxide stimulates ubiquitin-conjugating activity and expression of genes for specific E2 and E3 proteins in skeletal muscle myotubes. American journal of physiology. Cell physiology **285**(4), C806-812 (2003). doi:10.1152/ajpcell.00129.2003

368. Aversa, Z., Bonetto, A., Penna, F., Costelli, P., Di Rienzo, G., Lacitignola, A., Baccino, F.M., Ziparo, V., Mercantini, P.,

Rossi Fanelli, F., Muscaritoli, M.: Changes in myostatin signaling in non-weight-losing cancer patients. Annals of surgical oncology **19**(4), 1350-1356 (2012). doi:10.1245/s10434-011-1720-5

369. Leger, B., Derave, W., De Bock, K., Hespel, P., Russell, A.P.: Human sarcopenia reveals an increase in SOCS-3 and myostatin and a reduced efficiency of Akt phosphorylation. Rejuvenation research **11**(1), 163-175B (2008). doi:10.1089/rej.2007.0588

370. Bish, L.T., George, I., Maybaum, S., Yang, J., Chen, J.M., Sweeney, H.L.: Myostatin is elevated in congenital heart disease and after mechanical unloading. PloS one **6**(9), e23818 (2011). doi:10.1371/journal.pone.0023818

371. Hittel, D.S., Berggren, J.R., Shearer, J., Boyle, K., Houmard, J.A.: Increased secretion and expression of myostatin in skeletal muscle from extremely obese women. Diabetes **58**(1), 30-38 (2009). doi:10.2337/db08-0943

372. Gustafsson, T., Osterlund, T., Flanagan, J.N., von Walden, F., Trappe, T.A., Linnehan, R.M., Tesch, P.A.: Effects of 3 days unloading on molecular regulators of muscle size in humans. Journal of applied physiology **109**(3), 721-727 (2010). doi:10.1152/japplphysiol.00110.2009

373. Ju, C.R., Chen, R.C.: Serum myostatin levels and skeletal muscle wasting in chronic obstructive pulmonary disease. Respiratory medicine **106**(1), 102-108 (2012). doi:10.1016/j.rmed.2011.07.016

374. Verzola, D., Procopio, V., Sofia, A., Villaggio, B., Tarroni, A., Bonanni, A., Mannucci, I., De Cian, F., Gianetta, E., Saffioti, S., Garibotto, G.: Apoptosis and myostatin mRNA are upregulated in the skeletal muscle of patients with chronic kidney disease. Kidney international **79**(7), 773-782 (2011). doi:10.1038/ki.2010.494

375. Gonzalez-Cadavid, N.F., Taylor, W.E., Yarasheski, K., Sinha-Hikim, I., Ma, K., Ezzat, S., Shen, R., Lalani, R., Asa, S., Mamita, M., Nair, G., Arver, S., Bhasin, S.: Organization of the human myostatin gene and expression in healthy men and HIV-infected men with muscle wasting. Proceedings of the National Academy of Sciences of the United States of America **95**(25), 14938-14943 (1998).

376. Lokireddy, S., McFarlane, C., Ge, X., Zhang, H., Sze, S.K., Sharma, M., Kambadur, R.: Myostatin induces degradation of sarcomeric proteins through a Smad3 signaling mechanism during skeletal muscle wasting. Molecular endocrinology **25**(11), 1936-1949 (2011). doi:10.1210/me.2011-1124

377. Lokireddy, S., Wijesoma, I.W., Sze, S.K., McFarlane, C., Kambadur, R., Sharma, M.: Identification of atrogin-1-targeted proteins during the myostatin-induced skeletal muscle wasting. American journal of physiology. Cell physiology **303**(5), C512-529 (2012). doi:10.1152/ajpcell.00402.2011

378. Durieux, A.C., Amirouche, A., Banzet, S., Koulmann, N., Bonnefoy, R., Pasdeloup, M., Mouret, C., Bigard, X., Peinnequin, A., Freyssenet, D.: Ectopic expression of myostatin

induces atrophy of adult skeletal muscle by decreasing muscle gene expression. Endocrinology **148**(7), 3140-3147 (2007). doi:10.1210/en.2006-1500

379. Li, Z.B., Kollias, H.D., Wagner, K.R.: Myostatin directly regulates skeletal muscle fibrosis. The Journal of biological chemistry **283**(28), 19371-19378 (2008). doi:10.1074/jbc.M802585200

380. Zhu, J., Li, Y., Shen, W., Qiao, C., Ambrosio, F., Lavasani, M., Nozaki, M., Branca, M.F., Huard, J.: Relationships between transforming growth factor-beta1, myostatin, and decorin: implications for skeletal muscle fibrosis. The Journal of biological chemistry **282**(35), 25852-25863 (2007). doi:10.1074/jbc.M704146200

381. Elkasrawy, M., Immel, D., Wen, X., Liu, X., Liang, L.F., Hamrick, M.W.: Immunolocalization of myostatin (GDF-8) following musculoskeletal injury and the effects of exogenous myostatin on muscle and bone healing. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society **60**(1), 22-30 (2012). doi:10.1369/0022155411425389

382. Baehr, L.M., Furlow, J.D., Bodine, S.C.: Muscle sparing in muscle RING finger 1 null mice: response to synthetic glucocorticoids. The Journal of physiology **589**(Pt 19), 4759-4776 (2011). doi:10.1113/jphysiol.2011.212845

383. Cho, J.E., Fournier, M., Da, X., Lewis, M.I.: Time course expression of Foxo transcription factors in skeletal muscle

following corticosteroid administration. Journal of applied physiology **108**(1), 137-145 (2010). doi:10.1152/japplphysiol.00704.2009

384. Furuyama, T., Kitayama, K., Yamashita, H., Mori, N.: Forkhead transcription factor FOXO1 (FKHR)-dependent induction of PDK4 gene expression in skeletal muscle during energy deprivation. The Biochemical journal **375**(Pt 2), 365-371 (2003). doi:10.1042/BJ20030022

385. Sacheck, J.M., Hyatt, J.P., Raffaello, A., Jagoe, R.T., Roy, R.R., Edgerton, V.R., Lecker, S.H., Goldberg, A.L.: Rapid disuse and denervation atrophy involve transcriptional changes similar to those of muscle wasting during systemic diseases. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **21**(1), 140-155 (2007). doi:10.1096/fj.06-6604com

386. Sandri, M., Sandri, C., Gilbert, A., Skurk, C., Calabria, E., Picard, A., Walsh, K., Schiaffino, S., Lecker, S.H., Goldberg, A.L.: Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. Cell **117**(3), 399-412 (2004).

387. Senf, S.M., Dodd, S.L., Judge, A.R.: FOXO signaling is required for disuse muscle atrophy and is directly regulated by Hsp70. American journal of physiology. Cell physiology **298**(1), C38-45 (2010). doi:10.1152/ajpcell.00315.2009

388. Stitt, T.N., Drujan, D., Clarke, B.A., Panaro, F., Timofeyva, Y., Kline, W.O., Gonzalez, M., Yancopoulos, G.D., Glass, D.J.:

The IGF-1/PI3K/Akt pathway prevents expression of muscle atrophy-induced ubiquitin ligases by inhibiting FOXO transcription factors. Molecular cell **14**(3), 395-403 (2004).

389. Zhang, G., Jin, B., Li, Y.P.: C/EBPbeta mediates tumour-induced ubiquitin ligase atrogin1/MAFbx upregulation and muscle wasting. The EMBO journal **30**(20), 4323-4335 (2011). doi:10.1038/emboj.2011.292

390. Lagirand-Cantaloube, J., Offner, N., Csibi, A., Leibovitch, M.P., Batonnet-Pichon, S., Tintignac, L.A., Segura, C.T., Leibovitch, S.A.: The initiation factor eIF3-f is a major target for atrogin1/MAFbx function in skeletal muscle atrophy. The EMBO journal **27**(8), 1266-1276 (2008). doi:10.1038/emboj.2008.52

391. Lagirand-Cantaloube, J., Cornille, K., Csibi, A., Batonnet-Pichon, S., Leibovitch, M.P., Leibovitch, S.A.: Inhibition of atrogin-1/MAFbx mediated MyoD proteolysis prevents skeletal muscle atrophy in vivo. PloS one **4**(3), e4973 (2009). doi:10.1371/journal.pone.0004973

392. Jogo, M., Shiraishi, S., Tamura, T.A.: Identification of MAFbx as a myogenin-engaged F-box protein in SCF ubiquitin ligase. FEBS letters **583**(17), 2715-2719 (2009). doi:10.1016/j.febslet.2009.07.033

393. Giussani, M., Cardone, M.F., Bodega, B., Ginelli, E., Meneveri, R.: Evolutionary history of linked D4Z4 and Beta satellite clusters at the FSHD locus (4q35). Genomics **100**(5), 289-296 (2012). doi:10.1016/j.ygeno.2012.07.011

394. Bodega, B., Cardone, M.F., Rocchi, M., Meneveri, R., Marozzi, A., Ginelli, E.: The boundary of macaque rDNA is constituted by low-copy sequences conserved during evolution. Genomics **88**(5), 564-571 (2006). doi:10.1016/j.ygeno.2006.05.001

395. Tawil, R., van der Maarel, S., Padberg, G.W., van Engelen, B.G.: 171st ENMC international workshop: Standards of care and management of facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD **20**(7), 471-475 (2010). doi:10.1016/j.nmd.2010.04.007

396. Ricci, G., Zatz, M., Tupler, R.: Facioscapulohumeral muscular dystrophy: more complex than it appears. Current molecular medicine (2014).

397. Scionti, I., Fabbri, G., Fiorillo, C., Ricci, G., Greco, F., D'Amico, R., Termanini, A., Vercelli, L., Tomelleri, G., Cao, M., Santoro, L., Percesepe, A., Tupler, R.: Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. Journal of medical genetics **49**(3), 171-178 (2012). doi:10.1136/jmedgenet-2011-100454

398. Ferreboeuf, M., Mariot, V., Bessieres, B., Vasiljevic, A., Attie-Bitach, T., Collardeau, S., Morere, J., Roche, S., Magdinier, F., Robin-Ducellier, J., Rameau, P., Whalen, S., Desnuelle, C., Sacconi, S., Mouly, V., Butler-Browne, G., Dumonceaux, J.: DUX4 and DUX4 downstream target genes

are expressed in fetal FSHD muscles. Human molecular genetics **23**(1), 171-181 (2014). doi:10.1093/hmg/ddt409

399. Jagannathan, S., Shadle, S.C., Resnick, R., Snider, L., Tawil, R.N., van der Maarel, S.M., Bradley, R.K., Tapscott, S.J.: Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Human molecular genetics (2016). doi:10.1093/hmg/ddw271

400. Pirozhkova, I., Petrov, A., Dmitriev, P., Laoudj, D., Lipinski, M., Vassetzky, Y.: 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**(10), e3389 (2008). doi:10.1371/journal.pone.0003389

401. Robin, J.D., Ludlow, A.T., Batten, K., Gaillard, M.C., Stadler, G., Magdinier, F., Wright, W.E., Shay, J.W.: SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy. Genome research **25**(12), 1781-1790 (2015). doi:10.1101/gr.190660.115

Chapter 2

Chromosome Conformation Capture in primary human cells

A. Cortesi[1], B. Bodega[1]

[1]Genome Biology Unit, Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi" (INGM), Via Francesco Sforza 35, 20122 Milan, Italy

[#] to whom correspondence should be addressed: bodega@ingm.org

Methods Mol Biol, 2016, 1480:213-21

Summary

3D organization of the genome, its structural and regulatory function of cell identity, is acquiring prominent features in epigenetics studies; more efforts have been done to develop techniques that allow studying nuclear structure. Chromosome Conformation Capture (3C) has been set up in 2002 from Dekker and from that moment great investments were in develop genomics variants of 3C technology (4C, 5C, Hi-C) providing new tools to investigate the shape of the genome in a more systematic and unbiased manner. 3C method allow scientists to fix dynamic and variable 3D interactions in nuclear space; consequently, to study which sequences interact, how a gene is regulated by different and distant enhancer or how a set of enhancer could regulate transcriptional units; to follow is the conformation that mediate regulation change in development; to evaluate if this fine epigenetic mechanism is impaired in disease condition.

Key words: Chromosome Conformation Capture (3C), 3D interactions, nuclear structure, primary human cells

1. Introduction

Structure of the cell nucleus has sparked interest from years. The development of chromosome conformation capture (3C) technology has allowed to deeply investigate genome topology [1-3].

3C approach is a biochemical method that enables to study the three-dimensional organization of the genome in cells. Using formaldehyde, physical interactions are fixed; a subsequent digestion leads the formation from interacting sequences of free ends that are in close proximity. Then a ligation reaction in diluted conditions disadvantage inter-molecular ligation, promoting intra-molecular fusions. As a result, is generated a collection of products that are composed of DNA fragments that were originally physically near together in the nuclear space. Using a bait primer on the region of interest and the other one on the supposing region of interaction, in a simple PCR it could be possible to visualize the band, if the interaction occurs [2].

First study, using 3C to explore the three-dimensional organization of chromosomes at high resolution, describes intrachromosomal interactions between telomeres as well as interchromosomal interactions between centromeres and between homologous chromosomes in yeast [2]. The idea that the 3D genome topology is in some way functional to the regulation of nuclear activities takes place and a lot of work is done in this direction. 3C method was applied to analyze

physical connections between genes and *cis* enhancers (mouse and human β-*globin* locus model [4-6]). Then, highly specific associations between loci located on separate chromosomes are described. These *trans*-interactions can be between a distant enhancer with different target genes (olfactory receptor genes [7]). In other cases, *trans*-interactions appear to play a role in a higher level of gene control to coordinately regulate multiple loci with a set of both intrachromosomal and interchromosomal interactions (T helper 2 cytokine locus [8]), or providing additional levels of gene regulation by allowing combinatorial association of genes and sets of regulatory elements (imprinted loci [9]). It is also reported that specific DNA binding sites could mediate the formation of this topologically complex structure (Polycomb Response Elements in Drosophila [10]). Finally, it is demonstrated that 3D structure could also have a role in developmental processes (mammalian X-chromosome inactivation [11,12]).

Here, we describe the 3C technique adapted to study primary human cells, allowing to capture 3D chromatin interactions.

2. Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

Cell Lysis Buffer: 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 5 mM

$MgCl_2$, 100 μM EGTA. Before the use complement with protease inhibitors, PIC (Protease Cocktail Inhibitor) and PMSF (phenylmethanesulfonylfluoride).

Crosslinking Buffer: 10 mM Tris-HCl pH 8, 50 mM NaCl, 10 mM $MgCl_2$. Before the use complement with protease inhibitors, PIC and PMSF, and 1 mM DTT (Dithiothreitol).

Formaldehyde Solution 36,5-38% in H_2O.

Glycine 2,5 M.

PBS 1X (Phosphate buffered saline pH 7.4).

Sodium dodecyl sulfate (SDS) 10%.

Digestion Buffer: buffer of the specific restriction enzyme diluted at 1,5X. Before the use complement with 0.3% SDS.

Triton X-100 100%.

Ligation Buffer (1,15X): 57 mM Tris-HCl pH 7.5, 11 mM $MgCl_2$, 11 mM DTT, 1,1 mM ATP, 1% Triton X-100. Prepare at the use.

T4 DNA Ligase

Phenol-chloroform-isoamylalcohol 25:24:1.

Ethanol (100% and 70%).

3 M Sodium Acetate pH 5.2.

Glycogen.

Proteinase K.

RNAse cocktail (500 U/ml of RNase A and 20,000 U/ml of RNase T1).

Resuspension Buffer: 10 mM Tris-HCl pH 7.5.

Qubit dsDNA Broad Range Assay Kit.

3. Methods

3.1 Nuclei isolation

1. Grow cells in appropriate culture conditions.
2. Count cells using an automated cell counter. Consider to use $3,5*10^6$ cells for each experiment.
3. Collect cells by centrifugation at 1.000 rpm for 5 minutes at room temperature.
4. Resuspend cells, pipetting up and down, in 10 ml of cold Cell Lysis Buffer (See Note 1).
5. Incubate on ice for 10 minutes.
6. Check the lysis at microscope (See Note 2).

3.2 Crosslinking

1. Collect nuclei by centrifugation at 1.350 rpm for 5 minutes at 4°C.
2. Resuspend nuclei in 5 ml of Crosslinking Buffer.
3. Crosslink by the addition of 270 μl of 36,5-38% Formaldehyde Solution (2%) (See Note 3).
4. Incubate for 10 minutes at room temperature on a shaker.
5. Quench the reaction by the addition of 250 μl of 2,5 M Glycine (125 mM).
6. Incubate for 5 minutes at room temperature on a shaker.
7. Collect cross-linked nuclei by centrifugation at 1.350 rpm for 10 minutes at room temperature.

8. Wash twice with ice-cold PBS 1X.
9. Collect cross-linked nuclei by centrifugation at 1.350 rpm for 5 minutes at 4°C.

3.3 Digestion

1. Resuspend cross-linked nuclei in 500 μl of Digestion Buffer.
2. Incubate for 1 hour at 37°C with bland agitation.
3. Add 110 μl of 100% Triton X-100 (1.8%) to sequester SDS.
4. Incubate for 1 hour at 37°C with bland agitation.
5. Collect 60 μl aliquot of the sample and label as undigested genomic DNA control (UND).
6. Digest with 500 U of restriction enzyme at appropriate conditions overnight with bland agitation (See Note 4).
7. Collect 60 μl aliquot of the sample and label as digested genomic DNA control (DIG).
8. Inactivate the restriction enzyme by the addition of 95 μl of 10% SDS (1.6%).
9. Incubate for 25 minutes at 65°C with bland agitation.

3.4 Ligation

1. Transfer the reaction in a 50 ml tube.
2. Dilute DNA to a final concentration of 2,5 ng/μl in 10 ml of Ligation Buffer (See Note 5).
3. Incubate for 1 hour at 37°C with bland agitation.

4. Add 2.000 U of T4 DNA Ligase.
5. Incubate for 8 hours at 16°C with bland agitation.
6. Leave sample at room temperature for 30 minutes.
7. Treat with 1 mg of Proteinase K.
8. Incubate O.N. at 65°C with constant agitation to reverse the formaldehyde crosslinks.

3.5 Purification

1. Add 25 μl of RNAse cocktail.
2. Incubate for 40 minutes at 37°C with bland agitation to remove RNA.
3. Purify DNA, adding one volume of phenol-chloroform-isoamylalcohol, vortexing for 2 minutes and centrifuging at 4.000 rpm for 15 minutes at RT.
4. Transfer the supernatant to a new tube.
5. Dilute 1:2 in bi-distillated water.
6. Precipitate with three volumes of 100% Ethanol, 1/10 volume of 3 M Sodium Acetate pH 5.2 and 10 μl of Glycogen, leaving at -80°C for at least 1 hour.
7. Centrifuge at 3.000 rpm for 45 minutes at 4°C.
8. Discard the supernatant.
9. Wash the pellet with 10 ml of 70% Ethanol.
10. Centrifuge at 3.000 rpm for 15 minutes at 4°C.
11. Air dry the pellet.
12. Dissolve in 100 μl of Resuspension Buffer.
13. Quantify DNA by Qubit dsDNA Broad Range Assay Kit.

3.6 Purification of Controls aliquots (UND and DIG)

1. Process aliquots of UND and DIG, bringing volume to 100 µl with Resuspension Buffer.
2. Treat with 100 µg of Proteinase K.
3. Incubate for 1 hour at 65°C with constant agitation to reverse the formaldehyde crosslinks.
4. Add 1 µl of RNAse cocktail.
5. Incubate for 30 minutes at 37°C with bland agitation to remove RNA.
6. Purify DNA, adding one volume of phenol-chloroform-isoamylalcohol, vortexing for 2 minutes and centrifuging at 10.000 rpm for 4 minutes at RT.
7. Transfer the supernatant to a new tube.
8. Precipitate with two volumes of 100% Ethanol, 1/10 volume of 3 M Sodium Acetate pH 5.2 and 1 µl of Glycogen, leaving at -80°C for at least 1 hour.
9. Centrifuge at 14.000 rpm for 30 minutes at 4°C.
10. Discard the supernatant.
11. Wash the pellet with 100 µl of 70% Ethanol.
12. Centrifuge at 14.000 rpm for 5 minutes at 4°C.
13. Air dry the pellet.
14. Dissolve in 20 µl of Resuspension Buffer.
15. Quantify DNA by Qubit dsDNA Broad Range Assay Kit.
16. Perform PCR amplification using a Taq DNA Polymerase with provided reagents in 50 µl of reaction, using 0,3 µM of primers and 10 ng of template (See Note 6).

3.6 PCR

1. Design primers so that they will be in close proximity (proximal to 100 – 50 bp) to the restriction site for the regions to be amplified (See Note 7).
2. Select BACs covering the regions that are supposed to interact, quantify them by Qubit dsDNA Broad Range Assay Kit. Digest equimolar amounts of different BACs with the restriction enzyme of interest, than mix and ligate in 50 µl; after a step of purification by phenol extraction and ethanol precipitation, resuspend the BAC mix in 50 µl of Resuspension Buffer, obtaining the BAC control. It represents the PCR template that contains all possible ligation products that are relevant for the genomic regions of interest, and as such is considered a positive control.
3. Perform PCR amplification using a Taq DNA Polymerase with provided reagents in 50 µl of reaction, using 0,3 µM of primers and 25 ng of template (See Note 8). Remember to perform the PCR on 3C sample and BAC control in parallel, using for the BAC control 1/10 of the quantity used for 3C sample.
4. Run PCR products on 2% agarose gels, stained with ethidium bromide.
5. Quantify with the Image J (See Note 9).

4. Notes

1. 3C method could be performed using directly intact cells or nuclei. It depends on the cellular type in study and also on the percentage of crosslinking that you use. Indeed if you use cells with a great cytoscheleton or you need to crosslink with high percentage of formaldehyde to reveal less frequent interactions, it could be better to performed 3C experiment on isolated nuclei to promote next steps of the procedure and consequently the detection of higher signals of interaction.

2. If the lysis was incomplete you could use a Dounce homogenizer, performing approximately fifteen strokes.

3. Formaldehyde is used to cross-link protein-protein and protein-DNA interactions by means of their amino and imino groups. Advantages of this cross-linking agent are that it works over a relatively short distance (2 Å) and that cross-links can be reversed at higher temperatures [13-15]. The percentage of formaldehyde used depends on the frequency and stability of the interactions analyzed and has to be set for every new 3C experiment.

4. The restriction enzyme used depends on the locus to be analyzed. If small loci (<10–20 kb) are studied it is needed

to use frequently cutting restriction enzymes, while larger loci are investigated, six- base cutters (six-cutters) can also be used. It has to be considered that a high percentage of cross-linking could decrease digestion efficiency [16]. Indeed it is recommend to be sure that at least 60–70% of the DNA, but preferably 80% or more, be digested before continuing with the ligation step [17].

5. Ligation reaction has to be performed in a condition of diluting DNA (2,5 ng/μl), in a way that intramolecular ligation events between cross-linked DNA fragments are favoured. Moreover, independently of the restriction site analyzed, two types of junction are always over-represented. The first most abundant junction is with the neighboring DNA sequences, resulting from incomplete digestion and can constitute up to 20–30% of all the junctions. The second most abundant junction is with the other end of the same restriction fragment, after circularization, and can account for up to 5–10% of all the junctions. The formation of other junctions from fragments that are close together in the nuclear space represents only 0.2–0.5% of the junctions and with increasing genomic site separation decrease to <0.1%. It is therefore clear that to accurately quantify such rare events, it is necessary to include many genome equivalents in a PCR reaction [17].

6. Biological parameters, such as the heterogeneity of the cells, and technical parameters, as different efficiency in digestion and ligation step, also have to be considered. Therefore it is necessary to quantify the efficiency of digestion and ligation for each sample. It could be quantified by PCR or real-time PCR, amplifying fragments containing a restriction site of the enzyme used in UND, DIG and 3C samples and a region without restriction site as control to normalize the results of the PCR. The Ct values (cycle thresholds) were used to calculate the digestion efficiency according to the formula reported in Hagege et al. 2007 [18]: % restriction = 100 – 100/2^(Ct_{DIG}- Ct_{UND}).

7. To assess if the interaction detected in 3C reflects a functional 3D contact, the frequency of random collision has to be determined. This is important especially for *in cis* interactions. The reason is that the flexibility of the chromatin fiber could cause an engagement of DNA segments from the same fiber in random collisions, with a frequency inversely proportional to the genomic distance between them. For this porpoise, it is necessary to design primers that scan the entire region that is supposed to interact and to evaluate if two regions interact more frequently with each other than with neighbouring

sequences. Only in this case it is possible to affirm that you have a specific interaction. Instead *in trans* interactions random collisions are not expected, indeed any interaction that is detected indicates a specific contact.

8. The standard 3C PCR protocol uses a standard number of PCR cycles and a standard amount of DNA template for the analysis of all different ligation products. This approach is only semiquantitative and prone to inaccuracies. In fact PCR products that come to saturation, going outside from the linear range of the amplification reaction, are not suitable for differential analysis. To overcome this limitation, a real-time PCR approach using TaqMan probes, called 3C-qPCR, was developed [19,20].

9. Interaction frequencies are determined by dividing the intensity of PCR products bands obtained from 3C sample with that from the BAC control, for each pair of primers.

Figure legends

Figure 1: Chromosome Conformation Capture (3C) procedure.

Schematic representation of 3C steps: isolation of nuclei, crosslinking of 3D interactions, digestion with a specific restriction enzyme, creating free ends, ligation of them, obtaining circles of interacting sequences, purification and PCR amplification to evaluate the occurrence of a specific interaction.

Figure 2: 3C controls.

1) Quality check of total control materials: UND sample genomic band, DIG sample smear and 3C sample recovering of the band. 2) Quality check of control materials on specific restriction site: it should be a band of amplification product in UND sample, that decrease in DIG sample and recover in 3C sample. 3) Output of specific interaction, comparing with BAC control.

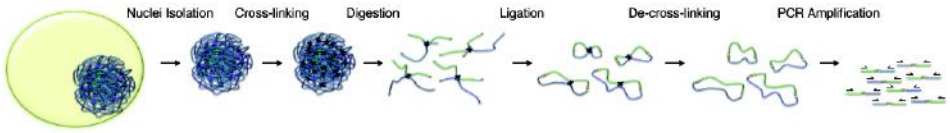


Figure 1: Chromosome Conformation Capture (3C) procedure.

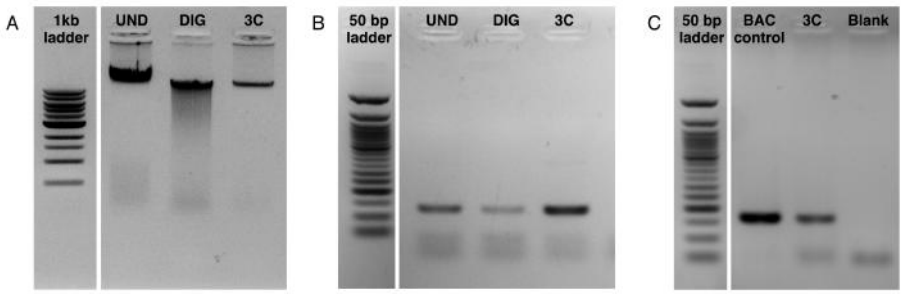


Figure 2: 3C controls.

References

1. de Laat, W., Dekker, J.: 3C-based technologies to study the shape of the genome. Methods **58**(3), 189-191 (2012). doi:10.1016/j.ymeth.2012.11.005

2. Dekker, J., Rippe, K., Dekker, M., Kleckner, N.: Capturing chromosome conformation. Science **295**(5558), 1306-1311 (2002). doi:10.1126/science.1067799

3. Calandra, P., Cascino, I., Lemmers, R.J., Galluzzi, G., Teveroni, E., Monforte, M., Tasca, G., Ricci, E., Moretti, F., van der Maarel, S.M., Deidda, G.: Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. Journal of medical genetics **53**(5), 348-355 (2016). doi:10.1136/jmedgenet-2015-103436

4. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., de Laat, W.: Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell **10**(6), 1453-1465 (2002).

5. Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., de Laat, W.: The beta-globin nuclear compartment in development and erythroid differentiation. Nature genetics **35**(2), 190-194 (2003). doi:10.1038/ng1244

6. Dekker, J.: The three 'C' s of chromosome conformation capture: controls, controls, controls. Nature methods **3**(1), 17-21 (2006). doi:10.1038/nmeth823

7. Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., Axel, R.: Interchromosomal interactions and olfactory receptor choice. Cell **126**(2), 403-413 (2006). doi:10.1016/j.cell.2006.06.035

8. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., Flavell, R.A.: Interchromosomal associations between alternatively expressed loci. Nature **435**(7042), 637-645 (2005). doi:10.1038/nature03574

9. Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., Ohlsson, R.: CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proceedings of the National Academy of Sciences of the United States of America **103**(28), 10684-10689 (2006). doi:10.1073/pnas.0600326103

10. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., Orlando, V.: Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nature cell biology **9**(10), 1167-1174 (2007). doi:10.1038/ncb1637

11. Xu, N., Tsai, C.L., Lee, J.T.: Transient homologous chromosome pairing marks the onset of X inactivation. Science **311**(5764), 1149-1152 (2006). doi:10.1126/science.1122984

12. Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., Heard, E.: Transient colocalization of X-

inactivation centres accompanies the initiation of X inactivation. Nature cell biology **8**(3), 293-299 (2006). doi:10.1038/ncb1365

13. Solomon, M.J., Varshavsky, A.: Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proceedings of the National Academy of Sciences of the United States of America **82**(19), 6470-6474 (1985).

14. Orlando, V., Strutt, H., Paro, R.: Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods **11**(2), 205-214 (1997). doi:10.1006/meth.1996.0407

15. Jackson, V.: Formaldehyde cross-linking for studying nucleosomal dynamics. Methods **17**(2), 125-139 (1999). doi:10.1006/meth.1998.0724

16. Splinter, E., Grosveld, F., de Laat, W.: 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods in enzymology **375**, 493-507 (2004).

17. Simonis, M., Kooren, J., de Laat, W.: An evaluation of 3C-based methods to capture DNA interactions. Nature methods **4**(11), 895-901 (2007). doi:10.1038/nmeth1114

18. Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., Forne, T.: Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nature protocols **2**(7), 1722-1733 (2007). doi:10.1038/nprot.2007.243

19. Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., de Laat, W.: CTCF mediates long-range chromatin looping and local histone modification in the

beta-globin locus. Genes & development **20**(17), 2349-2354 (2006). doi:10.1101/gad.399506

20. Wurtele, H., Chartrand, P.: Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology **14**(5), 477-495 (2006). doi:10.1007/s10577-006-1075-0

Chapter 3

Chromatin landscape of D4Z4 repeat interactome reveals a muscle atrophy signature in Facioscapulohumeral Dystrophy

A. Cortesi*[1-2], M. Pesant*[1], S Shina[1], L. Antonelli[3], F. Gregoretti[3], G. Oliva[3], G. Soldà[4], B. Bodega[1]

[1]Genome Biology Unit, Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi" (INGM), Via Francesco Sforza 35, 20122 Milan, Italy
[2]DIMET, Università degli Studi di Milano Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy
[3]CNR Institute for High Performance Computing and Networking (ICAR), Via P. Castellino 111, 80131 Naples, Italy
[4]Applied Biology Department of Biomedical Sciences, Humanitas University, Via Manzoni 113, 20089 Rozzano (Milan), Italy

*Equal Contribution

[#] to whom correspondence should be addressed: bodega@ingm.org

Submitted

Abstract

The autosomal dominant genetic disease Facioscapulohumeral Dystrophy (FSHD) represents the most common myopathy found in adults. The genetic defect of FSHD does not reside in any protein-coding gene, but instead is linked to the contraction below 11 copies of the 3.3kb D4Z4 tandem repeat macrosatellite located on chromosome 4q35 (FSHD locus). Several mechanisms underlying FSHD have been described, however they don't explain completely the variability of this complex epigenetic disease.

With our work we aimed to elucidate the contribution of D4Z4 repeat in mediating the 3D genomic architecture during normal and FSHD-associated myogenesis. We have used a combination of genome-wide approaches including circular chromosome conformation capture (4C)-seq specific for the 4q chromosome, ChIP-seq with chromatin state dynamics analysis as well as RNA-seq datasets to investigate the dynamics of chromatin epigenetic landscapes and folding in FSHD.

From these analysis we retrived that genome topology of FSHD cells was altered as well as the epigenetic landscape. Moreover in FSHD we retrived number of genes not regulated by DUX4 that instead present a deregulated interaction with 4q (lost and gained interactions) and showing active enhancers that are involved in the atrophic pathway.

Among others, we further investigated a *trans* interaction with

the muscle atrophy marker Atrogin1 (8q24) and showed that contact frequency between D4Z4 and Atrogin1, genomic locus topology, revealed by 4C-seq, chromosome conformation capture (3C) and 3D-FISH assays as well as Atrogin1 transcription, as well as epigenetic marks on the promoter, were deregulated during FSHD muscle cell differentiation.

We have revealed through 4C assay that Atrogin1 3D-interactome comprise also important players in atrophic pathway, differentially regulated in FSHD cells, and therefore contributing to the FSHD patients phenotype.

Finally we suggest that is the impairment in the epigentic environment of the FSHD locus, that occurs in patients, responsible for the deregulation of 3D contacts, at least for Atrogin1. Indeed when we restore another 4q D4Z4 array we are able to recover Atrogin1 expression.

Altogether our data highlight the contribution of the macrosatellite D4Z4 in regulating the 3D genome architecture during normal myogenic differentiation. We propose that the profound coordinated alterations of chromatin folding and functional states observed in FSHD may be a key trait for the disease.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is the most common myopathy found in adults, with an overall incidence of more than 1:14.000. FSHD is a dominant autosomal myopathy characterized by progressive, often asymmetric weakness and wasting of facial (facio), shoulder, and upper arm (scapulohumeral) muscles [1,2], but progressing to affect almost all skeletal muscles [3,4]. It is classified among progressive muscular dystrophies, characterized by muscular fiber necrosis and degeneration giving rise to progressive muscular atrophy.

FSHD is a genetically heterogeneous disorder and its genetic bases are unique and involve both genetic and epigenetic alterations [5]. There is a general consensus in the field in supporting the view that epigenetic mechanisms are important players in FSHD. Several FSHD clinical features, such as the variability in severity and rate of progression, the gender bias in penetrance, the asymmetric muscle wasting and the discordance of the disease in monozygotic twins, suggest a strong epigenetic contribution to the pathology [2,6-10].

The vast majority of FSHD are transmitted as an autosomal dominant trait with the disease locus mapping to the subtelomeric region of chromosome 4q, more specifically at 4q35-qter [11-13], on a specific FSHD-permissive haplotype of chromosome 4q [14]. This form has been termed FSHD1 and

accounts for approximately 95% of the cases [15]. This chromosomal region lacks classical genes, but contains an array of repeated units of a 3.3 kb macrosatellite repeat ordered head to tail [16], called D4Z4. In the general population, this repeat array is extremely polymorphic ranging from 11 to 150 copies [17,18], while FSHD1 patients present a reduction of the D4Z4 array, which leaves 1–10 units on the affected allele [19].

D4Z4 repeat arrays are not restricted to chromosome 4q, but homologous sequences have been identified on many chromosomes [20]. Due to a duplication event, a highly similar and equally polymorphic repeat array localizes to the subtelomere of chromosome 10q [21]. In particular, the subtelomere of chromosome 10q is almost identical to the region in 4q containing D4Z4 repeats, containing highly homologous and equally polymorphic repeat arrays [22,21] and sequence homology extending to 45 kb proximal of D4Z4 array and 15–25 kb distal [23].

However some families, with an undistinguishable clinical phenotype display a more complex pattern of inheritance and a distinct genetic defect. This second form is termed FSHD2 [5] and represents about 5% of patients meeting clinical criteria for FSHD [15,24]. This type of FSHD is caused by mutations in SMCHD1, a member of the condensin/cohesin family of chromatin compaction complexes, that binds to the D4Z4 repeat array. Interestingly, FSHD2 patients, who phenotypically show FSHD though lacking D4Z4 contractions, display general

D4Z4 hypomethylation [25], similar to the contracted D4Z4 repeat arrays in FSHD1 [24,4], indicating an important epigenetic condition necessary to develop or generate the disease.

In healthy subjects, the D4Z4 array present extensive PcG binding, DNA methylation, histone de-acetylation and chromatin compaction leading to a repressive chromatin organization. In FSHD1 patients chromatin conformation of FSHD locus is altered at multiple levels, from DNA methylation, through histone modifications up to higher-order chromosome structures, leading to a structural and epigenetic remodeling of the FSHD locus, toward a more active chromatin state, which is responsible for the de-repression of 4q35 genes [26-28].

A model explaining the pathology implicates the DUX4 retrogene encoded by D4Z4 and transcribed from the last, most telomeric D4Z4 repeat [29,30], including a non-repetitive region termed pLAM that contains a polyadenylation signal, necessary for DUX4 transcript stabilization [31].

The DUX4 retrogene encodes a double homeobox transcription factor, which is normally expressed at high levels in human testes and pluripotent stem cells, but epigenetically silenced in somatic cells [32]. However it has been found ectopically overexpressed in bursts in patients' myotubes, leading to the activation of genes associated with stem cell and germ-line development and resulting in the induction of toxicity [33] and apoptosis of muscle cells [34,35]. However the role of DUX4

remains unclear due to the complexity of its expression pattern [31,32,36], low mRNA and protein abundance [32,37]. Moreover, DUX4 might not be sufficient to explain the wide variability of this complex epigenetic disease, suggesting that other factors are likely involved [28,38-41].

Recent studies pointed out other mechanisms underlying FSHD manifestation. Namely, the 4q-D4Z4 proximal region was shown to engage multiple short- and long-range chromatin loops within the FSHD locus thereby modifying the expression of proximal genes such as FRG1 and more distal ones such as SORBS2. In a recent study aimed at defining the chromatin architecture of the proximal 4q-D4Z4 array region conducted with targeted Hi-C, SORBS2 was shown to loose a loop established with FRG1 in FSHD cells presenting short telomeres (by a mechanism linked to so-called Telomere Position Effect-Over Long Distances) thereby leading to overexpression of the gene and candidating it as a potential player in FSHD deregulated myogenic differentiation [42]. However, to which extent 4q-D4Z4 array may engage genome-wide contacts as well as an integrated view of FSHD-associated epigenome and the functional outcomes of these mechanisms for the disease are lacking at the moment.

In this paper we demonstrated that 4q-D4Z4 array is not only able to regulate *cis*-genes via chromatin looping *in cis* [43], but moreover is able to mediate long range functional interactions with genes *in trans*. Therefore in patients the contraction of the

array is responsible not only for the FSHD locus decompaction [28], but more globally acts in concert with broad alterations of the chromatin landscape to affect 3D 4q-D4Z4 interactome. In this context we show the functional deregulation of a 4q-D4Z4-mediated *trans*-interaction with Atrogin1, a gene codifying for a muscle specific ubiquitin ligase, which is over-expressed in muscle atrophy that is a primary symptom of FSHD.

Results and Discussion

To address how the interplay between molecular and epigenetic mechanisms and three-dimensional chromatin organization impact normal and FSHD-associated skeletal muscle cell differentiation, we used an integrated multiomics approach. We thus specifically performed chromosome conformation capture-seq (4C-seq) combined with chromatin segmentation characterization obtained from H3K4me3, H3K4me1, H3K27ac,H3K36me3 and H3K27me3 ChIP-seq datsets and integrated these datasets with published RNA-seq data (Figure 1 A).

Specifically, we set up a 4q-D4Z4 specific 4C-seq whose procedure is illustrated in Figure 1 B. The 40kb proximal region of the polymorphic chr4q D4Z4 array displays more than 98% of sequence identity with chr10q26. We thus took advantage of a Single Sequence Length Polymorphism (SSLP) upstream the D4Z4 array to design a paired-end allele-specific 4C. Mate reads give us the possibility to precisely assign the chromosome of origin (chr4 and chr10) as well as the contribution of chr4 alleles (4qA161, 4qB163 and 4qB168) and chr10 alleles (10qA166) to the interactions we retrieved. Pearson correlation analysis of our 4C-seq data for both viewpoints 4qsslp and 10qsslp showed high reproducibility between biological samples as well as 50-75% of reads mapping *in cis* respect to the viewpoints as described by [44,45]

(Figure S1 A-B). Genome browser snapshots of normalized 4qsslp 4C (4q-4C) coverage at chr4q35 showed high enrichment of 4C reads around the 4qsslp viewpoint (VP) but also revealed long-range interactions up to around 4Mb from the VP both in control (grey tracks) and FSHD myoblasts (blue tracks). On the other hand, coverage tracks of 10qsslp 4C (10q-4C) revealed genomic contacts around the 4q D4Z4 array suggesting a spatial proximity of 4q35 and 10q26 but displayed low coverage and absence of 4C peaks in the whole 4q35 region (Figure 1 C and Figure S2 A-B). Genome browser snapshot indeed clearly show altered distribution of 4q-4C coverage between CN and FSHD, arguing for 4q D4Z4-mediated alterations of contact loops in 4q35 (Figure 1 C). In particular 4C coverage showed notable differences between CN and FSHD cells at FRG1 candidate gene of the 4q35 (as already reported in [43]) (Figure S2 C). Moreover, CN and FSHD primary myoblasts cells clearly show genome-wide alterations, as evidenced by the circus plots presented in Figure 1D.

In order to characterize the epigenome of FSHD cells and to unveil possible epigenetic characteristics of 4q-sslp contacted regions, we performed ChIP-seq for H3K4me3, H3K4me1, H3K27ac, H3K36me3 and H3K27me3. Genome wide distribution of ChIP-seq signal between the two replicates of CN and FSHD myoblasts and myotubes showed high correlation between biological replicates for specific histone marks. Good

correlation were also observed comparing the datasets we generated with datasets of corresponding histone marks produced by ENCODE for the same cell types (Figure S3 A-B). Slight differences between CN and FSHD were seen at the genome-wide level in merged replicates as evidenced by pearson correlation and PCA analysis (Figure S3 C-D). We thus sought to use our ChIP-seq datasets to infer chromatin segmentation in CN and FSHD to get a better characterization of epigenetic alterations at the genome-wide level and within the differential interactions we retrieved by 4q-4C-seq. We used ChromHMM to define chromatin states and chose a 15 state-model accordingly named to the Roadmap nomenclature (Figure 1 E). An example of the ChromHMM and ChIP-seq signal tracks is shown around the SORBS2 gene at 4q35, recently described as being positively regulated in FSHD as well as being involved in chromatin loops with the proximal region of 4q D4Z4 array [42] (Figure 1 F).

We next sought to characterize 4q-4C interactions enrichment or depletion for those chromatin states. As shown in Figure S4 A-B, no clear enrichment/depletion for chromatin states could be observed in 4q common or CN-specific and FSHD-specific interactions. Nonetheless, common interactions were statistically enriched for CpG islands, whereas CN-specific and FSHD-specific interactions were enriched for LADs and lincRNAs, respectively (Figure S4 C), potentially suggesting that these features may be linked to the altered FSHD genomic

contact and may play a role in FSHD-associated molecular deregulations.

Finally, we addressed the question whether CN and FSHD show global similarities or differences in their chromatin states. For this, we computed Jaccard values [46] of each chromatin state in pairwise comparison between cell types and this analysis revealed that enhancers states in general were the most variable feature between cells (Figure 1 G).

4q-sslp 4C allowed us to retrieve hundreds of genomic interactions and the genes present within these regions. The venn diagrams presented in Figure 2 A display the common and CN- (lost in FSHD) and FSHD-specific (gained in FSHD) interactions as well as protein-coding genes. We next combined our 4q-4C data with ChromHMM data and recently published RNA-seq datasets to unveil a potential relationship between altered genomic contacts, altered epigenomes as well as transcriptional features of FHSD cells. A recent study by Jagannathan et al. [47] shed light on a robust consensus transcriptional signature mediated by the FSHD hallmark transcription factor DUX4. We thus intersected these RNA-seq datasets with our datasets to retrieve genes from 4C interactions that are lost or gained by FSHD, showing active enhancers in FSHD and regulated or not regulated by DUX4 (Figure 2 B). Genes positively associated with DUX4 and negatively associated to DUX4 were subjected to Gene Ontology analysis. DUX4 dependent genes with a deregulated

interaction with 4q-sslp (lost and gained interactions) and showing active enhancers were enriched for terms associated to development, differentiation and RNA metabolism processes as described in the literature (Figure 2 C-D). DUX4 independent genes that instead present a deregulated interaction with 4q-sslp (lost and gained interactions) and showing active enhancers were instead enriched for terms associated to protein metabolism and modification (Figure 2 E). Interestingly, these processes are involved in protein degradation pathways characteristic or at least associated with muscle atrophy phenotypes. Among the genes enriched for these pathways we focused on FBXO32 (Atrogin1) (Figure 2 F), a muscle specific ubiquitin ligase, marker of atrophy when overexpressed [48-50]. Because muscle atrophy is one of the primary symptom of FSHD, we decided to better investigate this interactor, hypothesizing that 4q-D4Z4-Atrogin1 could be an example of a functional deregulated genomic interaction thus being linked to disease manifestation.

Firstly, we validated by 3C experiment D4Z4 – Atrogin1 interaction; we used as bait primer a sequence on D4Z4 [43] as well as various primers encompassing the Atrogin1 locus as illustrated in Figure S5 A. 3C-PCR analysis revealed multiple interactions between D4Z4 an Atrogin1 in the 5’ and 3’UTR regions as well as within the gene body, thus suggesting a multi-looped structure. These interactions were validated both in CN and FSHD cells.

We next sought to further validate between 4q-D4Z4 proximal regions and Atrogin1 using a more sensitive technique, 3D FISH. Because D4Z4 repeat has different chromosomal localizations in the human genome, we performed 3D multi color FISH using a probe for the repeat itself (green in Figure S5 B) and probes for 4q (red in Figure S5 B) and 10q (pink in Figure S5 B) in unique regions upstream the D4Z4 array. This enabled us to discriminate the D4Z4 spots of 4q and 10q. Moreover we calculated the distance between D4Z4 spots and 4q and 10q spots (approximately 0,5 μm) (Figure S5 C) and decided correct for this distance in the following experiments using 4q and 10q probes.

We performed 3D multicolor FISH with probes for Atrogin1 (red in Figure 3 A-B), 4q (green in Figure 3 A-B) and 10q (pink in Figure 3 A-B), in three controls and three FSHD primary myoblast cell lines. In order to automate image analysis and specifically select Atrogin1, 4q and 10q spots, we developed a software relying on a novel algorithm (exemplified in Figure 3A), that allows us to retrieve distances between all spots and distances of all spots from the nuclear centroid.

First we looked at measurements of distances of Atrogin1 to 4q and 10q. As a *trans* interaction is considered consistent when it is detected in a range of 10% of the cells, we chose as a cutoff the lowest 10 percentile of the measurements.

We found that Atrogin1 gene interacts both with 4q sub tel and 10q sub tel in CN cells, confirming 4C data, whereas 4q-

Atrogin1 interaction is reduced in FSHD cells (Figure 3 C). This strongly suggests that contraction of the 4q-D4Z4 array causative of FSHD might not only be characterized by alterations *in cis*, but might also be involved in deregulated *trans* interactions.

Interestingly 4q and 10q loci interact each other in both CN and FSHD (Figure S5 D-E), validating 4C results shown in Figure 1C and suggesting that these homolog sequences share the same nuclear domains probably because of their reciprocal interaction.

We next further analyzed the nuclear topology of this interaction. Distribution of frequency of localization from the centroid to the nuclear periphery showed that, although occupying a similar nuclear domain, Atrogin1 is statistically less peripheral than 4q and 10q in both controls and FSHD patients (Figure 3 D).

To better characterize the nuclear topology of the interaction between Atrogin1 and 4q-D4Z4 proximal region, we plotted the distribution of distances from the nuclear centroid of the spots that interact from those that do not interact. This revealed that 4q spots interacting with Atrogin1 in both controls and patients are less peripheral than those that do not interact (Figure 3 E). This observation was specific to 4q-D4Z4 proximal region-Atrogin1 interaction as we did not observe any difference when we considered 10q-Atrogin1 interactions (Figure S5 F). This suggests that 4q-D4Z4 array could occupy different positions in

the nucleus, from very peripheral to less peripheral. We can hypothesize that the contact engaged by 4q-D4Z4 with Atrogin1 in CN cells tends to be lost in FSHD cells, in part due or accompanied by a more peripheral re-localization of the contracted 4q-D4Z4 in FSHD cells.

We asked if this structural 3D re-organization impacts Atrogin1 transcriptional regulation. We performed ChIP qPCR experiments and detected an increased binding of RNA PolII at Atrogin1 promoter in FSHD respect to controls (Figure 3 F and S6 A), accompanied by Atrogin1 transcriptional aberrant upregulation during myogenic differentiation (Figure S6 B-C) in patients respect to controls (Figure 3 G and S6 D). This suggest that the decreased interaction between 4q-D4Z4 proximal region and Atrogin1 leads to the transcriptional deregulation of Atrogin1 observed in patients.

Finally, we assessed if the FSHD hallmark transcription factor DUX4 is involved in the transcriptional upregulation of Atrogin1 in FSHD cells. We transfected both control and FSHD myoblasts (in which DUX4 is not expressed and Atrogin1 is minimally expressed and not yet upregulated upon differentiation) with a mock plasmid and a plasmid coding for DUX4 and confirmed that Atrogin1 is not induced by DUX4 in these experimental conditions (Figure S7 A-B).

In order to better investigate the relevance for FSHD manifestation of the alterations linking the 4q-D4Z4 chromatin topololy, FHSD-epigenome and Atrogin1, we decided to further

investigate the genomic interactome of Atrogin1 by performing the reverse 4C-seq analysis, taking the promoter of Atrogin1 as a viewpoint. 4C datasets showed good biological reproducibility and a percentage of mapped reads *in cis* and *trans* within accepted standards (Figure S8 A). Figure 4 A depicts genome browser snapshots of the 4C normalized coverage at the Atrogin1 locus as well as chromatin segmentation tracks. Circos plot representation of genome-wide interactions involving the promoter of Atrogin1 clearly shows alterations in FSHD respect to control (Figure 4 B). Indeed, we detected common but also CN- and FSHD-specific interactions mediated by Atrogin1 as well as associated protein-coding genes (Figure 4 C). In order to assess if genes retrieved from Atrogin1 interactome of CN and Atrogin1 interactome of FSHD display alteration in chromatin states, we performed a chromatin switch analysis using Chromdiff software and retrieved genes that statistically show differences for at least 1 of the 15 chromatin states. Pie charts of Figure 4D show that a small fraction of interacting genes, roughly 10%, display switches in chromatin states when comparing control and FSHD. Genes showing state switches were subjected to Gene Ontology analysis and enriched pathways are represented as clustered heatmaps according to the semantic similarity between ontology terms (Figure 4 E and S8 B). Among ontology terms enriched, both controls (Figure S8 B) and FSHD (Figure 4 E) Atrogin1 interactomes displayed a cluster of terms related to muscle differentiation, muscle

hypertrophy and muscle adaptation, thus suggesting that Atrogin1 engages genomic contacts with genes relevant for muscle hypertrophy/atrophy processes that are further associated with altered chromatin states when comparing control and FHSD cells.

Finally, because 4q-D4Z4 array interaction with Atrogin1 is lost in FSHD condition leading to transcriptional deregulation of this gene and the activation of the atrophic pathway, we addressed if Atrogin1 trancriptional deregulation could be the direct consequence of the D4Z4 derepression. For this purpose we used two different experimental approaches. First we performed 3D multicolor FISH with Atrogin1 and 4q probes in FSHD2 cell lines (Figure 4 F). FSHD2 form of the disease is in fact not related to the contraction of 4q-D4Z4 array but instead shows epigenetic decompaction (DNA hypomethylation) of array [51]. Furthermore, 3D FISH analysis showed that interactions between Atrogin1 and 4q-D4Z4 proximal region is decreased in FSHD2 myoblasts, similar to FSHD1 (Figure 4 G) suggesting that the epigenetic switch of the 4q-D4Z4 array *per se* is responsible for the decreased interaction with Atrogin1. RT-qPCR experiments showed that Atrogin1 is aberrantly upregulated in FSHD2 cells upon differentiation as in FSHD1 cells (Figure S6 D).

To test if the 4q-D4Z4 array could mediate the transcriptional regulation of Atrogin1, we transfected CN and FSHD primary myoblasts with a BAC (CH16-291A23) containing 4q at least 15

D4Z4 repeats (Bodega, unpublished). This BAC is already reported to acquire epigenetic features of the repressed 4q-D4Z4 array when transfected [28]. After checking BAC transfection and the absence of differentiation induction (Figure S8 C, left), we assessed Atrogin1 expression by RT-qPCR and observed that Atrogin1 expression is reduced when 4q D4Z4 array BAC is introduced respect to a control BAC with unrelated sequences in both controls and FSHD cells (Figure 4H). The repression of Atrogin1 in CH16-291A23 transfected myoblasts was not accompanied by a repression of FOXO3, activator of Atrogin1 (Figure S8 C, right), thus suggesting a direct role of 4q-D4Z4 in modulating Atrogin1 expression.

Experimental procedures

Cell cultures

Human primary myoblasts from healthy donors and FSHD patients were obtained from the Telethon BioBank of the C. Besta Neurological Institute, Milano, Italy and the Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA (see Supplementary table 1).

Human primary immortalized myoblasts from healty donors and FSHD 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 (see Supplementary table 1).

The cell lines from the Telethon BioBank were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 20% fetal bovine serum (FBS), 2 mM L-glutamine, 10 μg/ml insulin, 25 ng/ml human fibroblast growth factor (hFGF), 10 ng/ml human epidermal growth factor (hEGF) (proliferating medium), and then induced to differentiate by means of DMEM supplemented with 5% horse serum and 100 μg/ml human insulin (differentiating medium).

The cell lines from the Fields Center for FSHD were cultured in F10 Nutrient Mixture supplemented with 20% FBS, 2 mM L-glutamine, 10 ng/ml hFGF, 0,4 μg/ml dexamethasone (proliferating medium), and then induced to differentiate by

means of DMEM supplemented with 2% horse serum, 10 μg/ml human insulin, 4,5 g/L D-Glucose and 0,11 g/L sodium pyruvate (differentiating medium).

The immortalized cell lines were grown in plates coated with 0,1% gelatin and cultured in Media X (4:1 DMEM/Medium 199 plus 0,88 mg/L sodium pyruvate and 3,4 g/L sodium bicarbonate), supplemented with 15% FBS, 10 ng/ml hFGF, 2,5 ng/ml human hepatocyte growth factor (hHGF), 0,055 μg/ml dexamethasone, 0.03 μg/ml zinc sulfate, 1,4 μg/ml vitamin B12, 0,02 M HEPES (proliferating medium), and then induced to differentiate by means of Media X (4:1 DMEM/Medium 199) supplemented with 2% horse serum, 2 mM L-glutamine, 1mM sodium pyruvate, 20 mM HEPES (differentiating medium).

All of the patients satisfied the accepted clinical criteria for FSHD. They had undergone DNA diagnosis and were identified as carriers of small (<38 kb, <11 repeats) 4q35-located D4Z4 repeat arrays, as determined by p13E-11 hybridization to EcoRI-digested and EcoRI/ BlnI-digested genomic DNA. The details of the cell lines are shown in Additional file.

4C procedure

4C procedure followed the procedure described in [44] with some modifications. Control and FSHD myoblasts were grown in three 150 mm culture plates per sample. Cells were retrieved by trypsinisation, washed in PBS, resuspended in cold lysis buffer (50nM Tris-HCl pH 7.5, 150mM NaCl, 5mM EDTA, 0.5%

NP40, 1% TX-100 and protease inhibitors) and incubated at 4° for 1h. Efficient lysis was assessed by staining with Trypan blue and visualization under the microscope. Obtained nuclei were pelleted and resuspended in 5ml of crosslinking buffer (10 mM Tris pH 8, 10 mM $MgCl_2$, 50 mM NaCl, 1mM DTT and protease inhibitors) and crosslinked with formaldehyde (2% final) for 10 min at room temperature under gentle agitation. Crosslinking was quenched by incubation with glycin (0.125M final) for 5 min at room temperature and nuclei were washed 2 times with cold PBS, and pelleted at 1400 rpm for 5 min at 4°C.

Nuclei were then resuspended in 450 µl of H2O Milli-Q, 60 µl of 10X restriction buffer were added and tube were placed at 37°C. 15 µl 10% SDS were immediately added and samples were incubated in a Thermomixer for 1 hour at 37 °C (900 rpm). 100 µl of 10% TX-100 were added and samples were incubated for 1 hour at 37°C (900 rpm).

The first enzymatic digestion was performed with 3 pulses of 200U HindIII (NEB), for 4 hours, overnight and further 4 hours at 37°C while shaking (900 rpm). The enzymatic digestion was stopped by incubating the samples at 37°C for 20 min. Each sample was then transferred into a 50 ml tube and 5.7 ml of H_2O Milli-Q were added.

Ligation was performed by adding 700 µl 10X Ligase buffer (660mM Tris HCl pH 7.5, 50mM MgCl2, 10mM DTT, 10mM ATP) and 100U T4 DNA Ligase, overnight at 16°C under gentle agitation and further 30 min at room temperature.

Crosslinking was reversed by addition of 30 µl Prot K (10 mg/ml) and incubation overnight at 65°C. Then 30 µl RNase A (10 mg/ml, Roche) were added and samples were incubated 45 min at 37°C. 7 ml of Phenol–Chloroform were added to the samples. Samples were centrifuged 15 min, 3270 g at RT and the aqueous phase was collected in a fresh 50 ml tube. Ligated DNA was precipitated by addition of 7 ml Milli-Q, 1.5 ml 2M NaAC pH 5.6, 7µl glycogen (20 mg/ml) and 35 ml 100% ethanol. Samples were left to precipitate at -80°C for 3 hours and centrifuged 45 min, 4000 rpm at 4°C. DNA pellet was washed with 70% cold ethanol and pellet was resuspended in 150 µl 10mM Tris–HCl pH 7.5.

The second enzymatic digestion digestion was performed by adding 50U DpnII (NEB) followed by incubation overnight at 37°C. Digestion was inactivated by incubation at 65°C for 25 min. Digested 4C products were extracted with 500 µl Phenol–Chloroform-alcohol isoamyl and precipitated at -80°C for 3 hours by addition of 50 µl 3 M NaAc pH 5.2 and 950 µl 100% ethanol. Pelleted 4C products were washed with 70% cold ethanol and resuspended in 200 µl 10mM Tris–HCl pH7.5. Samples were transferred to a fresh 50 ml tube and the second ligation was performed in presence of 12.1 ml Milli-Q, 1.4 ml 10X Ligation buffer and 200 U T4 DNA Ligase overnight at 16°C under gentle agitation. 4C products were then precipitated with 0.7 ml 3M NaAC pH 5.2, 7 µl glycogen (20 mg/ml) and 35 ml 100% ethanol at -80°C for 3 hours. The resulting pellet was

washed with 70% cold ethanol and resuspended in 200 µl 10mM Tris–HCl pH 7.5.

4C templates were purified using 2 PureLink columns per sample and eluted in 15 µl per column. Eluates were combined and quantified using the Qubit dsDNA HS assay Reagent Kit (Thermo Fisher Scientific) and a Qubit Fluorometer (Thermo Fisher Scientific).

4C templates were amplified in 25 µl PCR reactions (12.5ng per reaction, 200ng of 4C template total) using bait specific primers (see Supplementary table 3 and 4) and Phusion Taq Polymerase (Thermo Fisher Scientific). PCR program consisted of 98°C for 3 min, 98°C for 30 sec, 55°C for 1 min, 72°C for 2min and 72°C for 7 min for a total of 29 cycles. PCR products were then purified on PureLink columns (4 columns per sample) and eluted in 10 µl elution buffer per column (40 µl total). Samples were quantified using the Qubit dsDNA BR assay Reagent Kit (Thermo Fisher Scientific) and a Qubit Fluorometer (Thermo Fisher Scientific). 4C sequencing libraries were prepared on 500 ng of 4C template using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, E7370) according to the manufacturer’s protocol without size selection and 5 cycles of PCR. Finally, libraries with cleaned with AMPure XP beads and eluted in 31 µl of TE. Libraries were sequenced on a NextSeq 500 (Illumina).

4C-Seq analysis of 4qSSLP

A 4q/10q specific strategy 4C-sequencing was used for studying 4q D4Z4 array- and 10q D4Z4 array-mediated interactions. Briefly, for 4q/10q specific paired-end 4C sequencing, primers were designed such that one of the selected primers reads in the single sequence length polymorphism (SSLP) sequences located in the proximal 4q and 10q D4Z4 arrays. The second 4C primer reads into the captured sequence ligated to the 'bait' fragment in the 4C procedure. This allows us to identify genomic regions interacting with 4q SSLP and 10q SSLP based on the prior knowledge of SSLP sequences of our samples (see Supplementary table 2).

All 4C-seq reads were demultiplexed based on the 4C reading primer that included the restriction site sequence where one mismatch was allowed for the reading primer and no mismatches were allowed for the restriction site sequence. All reads were then trimmed for read primer excluding restriction site sequence. Reads were then trimmed for low quality using trimmomatic v0.32 [52] and unpaired reads were removed. Read pairs were then categorized for the chromosomes and allele based on the SSLP sequence in accordance with the SSLP identified experimentally for each sample. Sequencing runs with at least 10,000 reads in each category were considered for further analysis (see Supplementary table 3). Reads reading into the captured sequence ligated to bait from

the technical and biological replicates were then pooled and mapped to a reduced genome derived from hg19 consisting of all unique 75nt-long regions surrounding HindIII sites. Reads were mapped with Bowtie2 [53] and parameters -N = 0 and –5 6 was used to trim the read sequences. To assess the reproducibility of 4C signal between the biological replicates, number of reads in each fragend was counted and Pearson's correlation was calculated. To find chromosome-wide interacting domains, 4C-ker [54] was used. 4C-ker uses a 3 state HMM to define regions of high-interaction using an input of window counts corrected for the effect of linear distance from the bait. High interacting domains were defined across the entire bait chromosome and other chromosomes. 4C-ker parameter of k = 10 for cis, k = 4 for nearbait and k = 19 for trans was used. Normalized 4C coverage tracks were generated for display in UCSC Genome Browser and Circos software [55] was used to visualize genome-wide 4C interactions.

4C-Seq Analysis of Atrogin1

All 4C-seq reads were demultiplexed based on the 4C reading primer that included the restriction site sequence where one mismatch was allowed for the reading primer and no mismatches were allowed for the restriction site sequence. All reads were then trimmed for read primer excluding restriction site sequence. Reads were then trimmed for low quality using

trimmomatic v0.32 [52] and unpaired reads were removed. Sequencing runs with at least 500,000 reads were considered for further analysis (see Supplementary table 4). Reads reading into the captured sequence ligated to bait from the technical replicates were then pooled and mapped to a reduced genome derived from hg19 consisting of all unique 75nt-long regions surrounding HindIII sites. Reads were mapped with Bowtie2 [53] and paramteres -N = 0 and –5 6 was used to trim the read sequences. To assess the reproducibility of 4C signal between the biological replicates, number of reads in each fragend was counted and Pearson's correlation was calculated. To find chromosome-wide interacting domains, 4C-ker [54] was used. 4C-ker uses a 3 state HMM to define regions of high-interaction using an input of window counts corrected for the effect of linear distance from the bait. High interacting domains were defined across the entire bait chromosome and other chromosomes. 4C-ker parameter of k = 10 for cis, k = 4 for nearbait and k = 19 for trans was used. Normalized 4C coverage tracks were generated for display in UCSC Genome Browser and Circos software [55] was used to visualize genome-wide 4C interactions.

ChIP assay and sequencing

Human primary myoblasts were seeded in 3-4 100 mm dishes. When cells reached the desired confluence 10 ml of growth medium was leaved and 1 ml of formaldehyde solution was

added (HEPES-KOH 50 mM pH 7.5, NaCl 100 mM, EDTA 1mM, EGTA 0,5 mM and formaldehyde 11%) in a way that cells were fixed in 1% formaldehyde for 12 min at room temperature. After formaldehyde quenching with 125 mM Glycine for 5 min, cells were washed two times in cold PBS with gentle swirl. Cells were harvested by using a scraper and cold PBS. Collected cells were lysed in 8 ml of LB1 solution (50 mM HEPES-KOH pH 7.5, 10 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100 and protease inhibitors) and incubated for 1h at 4°C. Cells were collected and washed in 8 ml of LB2 solution (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors) with gentle swirl 10 min at room temperature. Cells were collected and lysed in 600-800 μl of LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-laurylsarcosine) for 30 min on ice. Lysates were sonicated with Brenson A250 for 45 sec (25% amplitude 0,5 sec on 0,5 sec off) 4-6 times. An aliquot (10 μl) of the sonicated material was collected to determine the quality of the chromatin; control sample was incubated over night at 65°C for cross-link reversal, treated with proteinase K (SIGMA) and 0,5% SDS 1h at 55°C for protein digestion, purified by phenol–chloroform extraction and ethanol precipitation, resuspended in 10 mM Tris-HCl pH 7.5 and finally treated with RNAse (Thermo Fisher Scientific) for 15 min at 37°C. Control sample was loaded on 1% agarose gel for electrophoresis. We considered good a chromatin enriched in

fragments of 500 bp. The samples were quantified with Nanodrop spectrophotometer to determine the concentration of chromatin and clarified.

Chromatin was immunoprecipitated on a rotating wheel overnight at 4°C with anti-H3K4me3 (Millipore, 07-473), anti-H3K4me1 (Millipore, 07-436), anti-H3K27ac (Millipore, 07-360), anti-H3K27me3 (Millipore, 07-449) and anti-H3K36me3 (Abcam, 9050) and RNAPolII (Abcam, 5408) in 300 μl LB3 and 1% Triton X-100. Recovery of the immunocomplexes was obtained by the addition of 40 μl of Dynabeads protein G (Thermo Fisher Scientific) and incubation for 2-3 h at 4°C. Next, beads-antibody-chromatin complexes were washed one time in wash buffer low salt (0,1% SDS, 2 mM EDTA, 1% Triton X-100, 20 mM Tris-HCl pH 8.0 and 150 mM NaCl) for 5 min at 4°C, one time in wash buffer high salt (0,1% SDS, 2 mM EDTA, 1% Triton X-100, 20 mM Tris-HCl pH 8.0 and 500 mM NaCl) for 5 min at 4°C, one time in wash buffer low salt. Followed a brief wash in Tris-EDTA 50 mM NaCl. Immunocomplexes were eluted with elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and cross-linking was reversed by overnight incubation at 65°C. Samples were treated with RNAse (Thermo Fisher Scientific) for 1h at 37°C and with proteinase K (SIGMA) for 1h at 55 °C, purified by phenol–chloroform extraction and ethanol precipitation, and resuspended in 20 μl of 10 mM Tris-HCl pH 7.5.

DNA was quantified using the Qubit dsDNA HS assay Reagent

Kit (Thermo Fisher Scientific) and a Qubit Fluorometer (Thermo Fisher Scientific).

For ChIP-seq experiment we used two biological replicates of primary myoblasts and 4 days-differentiate myotubes for control (CN) and FSHD cells. Sequencing libraries were prepared from or 5 ng of DNA using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, E7370) according to the manufacturer's protocol without size selection and 12 cycles of PCR. Finally, libraries were cleaned with AMPure XP beads and eluted in 31 µl of 10mM Tris-HCl pH 8.0. Libraries were sequenced on a NextSeq 500 (Illumina).

Instead for RNAPolII ChIP we used two biological replicates of only primary myoblasts. Input was diluted 1:100 µl and 0,5 µl of each sample was used in qRT-PCR. Considering the highly representation of D4Z4 repeats present in human samples, for DBE amplification samples were diluted 1:20 and 0,5 µl of each sample was used in qRT-PCR (see Supplementary table 6). All experiments were repeated at least three times.

Quantitative RT-PCR (qRT-PCR) analysis was performed on an Step One Plus real time PCR detection system (Applied Biosystem), using the power SYBR Green q-PCR master mix (Thermo Fisher Scientific). The relative enrichment obtained by using RNA PolII antibody was quantified after normalization for input chromatin.

ChIP-seq data processing and analysis

The following already published ChIP-seq datasets from ENCODE were used: H3K4me1 of human skeletal muscle myoblast (ENCSR000ANI), H3K27ac of human skeletal muscle myoblast (ENCSR000ANF), H3K4me1 of human skeletal muscle myotubes (ENCSR000ANX) and H3K27ac of human skeletal muscle myotubes (ENCSR000ANV). H3K4me3, H3K4me1, H3K27ac, H3K27me3 and H3K36me3 reads were mapped on GRCh37/hg19 using bowtie2 with default parameters [53] and duplicate reads were filtered with samtools. For normalization, we used the deepTools2 package [56]. Aligned read files were corrected for sequencing depth and normalized to the respective input (bamCoverage and bamCompare modules). Reproducibility between replicate samples and already published datasets was assessed using the deepTools2 package. Pearson's coefficient correlation was calculated and represented as heatmaps using the multiBamSummary and plotCorrelation modules. PCA graphs were obtained with the plotPCA module.

Normalized signal tracks of fold-enrichment were generated using MACS2 v2.1.0 [57] with –m FE option. We used respective input for signal normalization of ChIP-seq histone marks coverage.

Chromatin state analysis

Chromatin states of control and FSHD myoblasts and myotubes

were defined with ChromHMM [58] with default parameters. We used the 5 histone marks H3K4me3, H3K4me1, H3K27ac, H3K36me3 and H3K27me3 as well as the respective input files to test different numbers of defined chromatin states. We chose 15 states as the optimal number according to the maximal informative annotated genomic features and minimal redundancy. Chromatin states of this 15-state model were named and color-coded for visualization according to the Roadmap nomenclature. The 15-state model was used to compute state enrichment on different genomic features represented as heatmaps in Figure 1. Calculation of pairwise Jaccard value presented in Figure 1 was performed with Bedtools v2.2.4 [59].

Active enhancers were assigned to their closest protein-coding genes (Gencode v19) using Bedtools v2.2.4 intersect and closest functions.

For chromatin state switch analysis between control cells and FSHD cells, we used the Chromdiff software [60]. Briefly, Chromdiff performs an intersection between chromatin states and any type of annotation file, performs a summarization of chromatin states on given coordinates and statistically tests chromatin state differences between two conditions. We thus used chromatin state bed files of control and FHSD myoblasts as well as Gencode v19 protein-coding genes coordinates as input for Chromdiff. We retrieved protein-coding genes showing statistically significant differences in their major states or any

other chromatin state between control and FSHD.

To determine in 4qSSLP contacted regions obtained by 4qSSLP 4C were associated with a particular chromatin state, we used GAT [61]. Briefly, chromatin states were grouped to define the functional subgroups “active TSS” (chromatin state 1_TssA), “transcription” (chromatin state 5_Tx + 6_TxWk), “genic enhancers” (chromatin state 7_EnhG1 + 8_EnhG2), “active enhancers” (9_EnhA1 + 10_EnhA2), “inactive enhancers” (11_EnhWk), “bivalent TSS” (12_TssBiv) and “Polycomb” (14_ReprPC). GAT was then used to evaluate the observed enrichment of those chromatin states in 4C interacting regions, expressed observed fold change over expected. Enrichment were considered statistically significant for a FDR-corrected p value <0.05.

Gene Ontology analysis

Protein-coding genes retrieved from intersection between 4C datasets, published RNA-seq datasets and chromatin state analysis were used as input in the Cytoscape v3.2.0 [62] plug-in ClueGO v2.1.5 [63]. Satistically enriched Biological Processes (update on 04/18/2016) were then visualized as heatmaps using Euclidian hierarchical clustering on k-score GO terms similarity. Heatmaps were generated in R.

3C Assay

The 3C assay was performed accordingly to Cortesi et al. [64]

with minor adaptations. A total of 3,5 × 10^6 cells were resuspended in 10 ml of cold cell lysis buffer (10 mM Tris, pH 8.0, 10 mM NaCl, 5mM $MgCl_2$, 100µM EGTA and protease inhibitors) and incubated for 10 min at 4°C; nuclei were pelleted and resuspended in 5 ml of crosslinking buffer (10 mM Tris-Cl, pH 8, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM dithiothreitol and protease inhibitors) and crosslinked with 2% formaldehyde for 10 min at room temperature. The reaction was quenched by the addition of 0.125 M glycine for 5 min. The crosslinked nuclei were washed with PBS and resuspended in 500 µl of restriction enzyme buffer (NEB) and, after the addition of 0.3% sodium dodecyl sulfate (SDS), incubated at 37°C for 1 h. SDS was sequestered by the addition of 1,8% Triton X-100 and incubation at 37°C for 1 h. Digestion was performed using 500 U of PvuII (NEB) at 37°C overnight with constant agitation. Enzyme was inactivated by the addition of 1.6% SDS and incubation at 65°C for 25 min. The reaction was diluted to a final concentration of 2.5 ng/µl in a ligation reaction buffer (57 mM Tris-HCl pH 7.5, 11 mM $MgCl_2$, 11 mM DTT, 1,1 mM ATP, 1% Triton X-100) and incubated at 37°C for 1 h to allow SDS sequestration by Triton X-100. Ligation was performed adding 2,000 U of T4 DNA ligase (NEB). The ligations were incubated at 16°C for 8 h. The samples were treated with proteinase K (SIGMA) and incubated overnight at 65°C to reverse the formaldehyde crosslinks. The samples were treated with RNAse cocktail (500 U/ml of RNase A and 20,000 U/ml of

RNase T1)(Thermo Fisher Scientific) and incubated at 37°C for 40 min. DNA fragments were purified by phenol–chloroform extraction and ethanol precipitation and then resuspended in 10 mM Tris-HCl pH 7.5. A reference template was generated by digesting, mixing and ligating bacterial artificial chromosomes spanning the genomic regions of interest: Atrogin1 region and the D4Z4 array. 3C templates and the reference template were used to perform semiquantitative PCR with DreamTaq DNA Polymerase (Thermo Fisher Scientific). Primers flanked the PvuII restriction sites scanned Atrogin1 region, instead the bait primer is located inside D4Z4 repeat (see Supplementary table 5). The PCR products were quantified using the ImageJ software. Data are presented as the ratio of amplification obtained with 3C templates and with the reference template.

Three-dimensional DNA FISH assay

BAC (BACPAC Resources Program, CHORI) DNA (CH16-77M12 for D4Z4 repeat, RP11-279K24 for 4q35.1, RP11-288G11 for 10q26.3 and RP11-174I12 for Atrogin1), 1-3 μg, was labeled with bio-dUTP (Thermo Fisher Scientific), dig-dUTP (ROCHE) and cy3-dUTP (Thermo Fisher Scientific) with nick translation in 50 μl of labeling mix buffer (0,02 mM dNTPs C-G-A, 0,01 mM dTTP, 0,01 mM labeled dUTP, 50 mM Tris-Hcl pH 7.8, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10 ng/μl BSA, 0,05 U/μl DNA PolII (Thermo Fisher Scientific), 0,004 U/μl DNase I (SIGMA)) for 2h at 16°C. Dimensions of probes were

checked on 2% agarose gel till are around 50 bp. Probes were precipitated with ethanol and resuspended in Tris-HCl 10 mM pH 7.5. For each experiment 100 ng of each probe was used, precipitating with 3,5 μg of Cot-1 DNA (Thermo Fisher Scientific) and 20 μg of unlabeled salmon sperm DNA (SIGMA). Probes were resuspended in 6 μl 50% formamide/2xSSC/10% dextran sulfate as follows: the pellet were resolved in the appropriate amount of 100% formamide (shake at 40°C, can take up to a few hours) and then were added the equal volume of 4xSSC/20% dextran sulfate. Cells were plated directly on coverslips and when reached a confluence of approximately 60% they were fixed with 4% paraformaldehyde in PBS 1X+Tween 0,1% (PBS-T) for 10 min at room temperature. During the last minute few drops of 0.5% TritonX-100/PBS were added, followed by three 3-min washes in PBS with 0.01% TritonX-100at room temperature. Cells were first permeabilized in 0.5% TritonX-100/PBS for 10 min at room temperature. A step of RNAse (Thermo Fisher Scientific) treatment was performed for 1h at 37°C. Cells were subjected to others passages of permeabilization in 20% Glycerol in PBS over night at room temperature, followed by freeze and thaw steps interleaved by soak in 20% Glycerol/PBS for 4 times. Three 10-min washes in at room temperature were performed. Cells were then incubated with 0.1M HCl for 5 minutes at room temperature, followed by a rinse in 2xSSC and then incubated in 50% formamide (pH = 7.0)/2xSSC (at least 30 minutes at

room temperature). Slides were equilibrate in 2xSSC for 2 min at room temperature, PBS for 3 min and then treated with 0.01 N HCl/0.0025% pepsin for 3-4 min to eliminate cyotoscheleton. Pepsin was inactivated with two 5-min washes in PBS/MgCl2 50 mM. Nuclei were post-fixed in 1% PFA in PBS for 1 min, washed for 5 min in PBS and two times in 2XSSC, and then back to 50%FA/2xSSC for at least 30 min. Hybridization solution was loaded on a clean microscopic slide, coverslip with nuclei was turned upside down on the drop of hybridization mixture and sealed with rubber cement. Samples were denatured at 75°C for 4 min and leaved to hybridize at 37°C in a metallic box floating in a 37°C water bath over night. Followed three 5-min washes in 2xSSC at 37°C, three 5-min washes in 0.1xSSC at 60°C, a rinse in 4xSSC/0.2%Tween and blocking in 4xSSC/0.2%Tween + 4% BSA for 20 min at 37°C. Samples were then incubated with the appropriate concentration of streptavidin (Thermo Fisher Scientific) (1:1.000) or antidigoxygenin (ROCHE) (1:100) diluted in 4xSSC/0.2%Tween + 4% BSA 35 min in a dark and wet chamber at 37°C. Followed three 3-min washes in 4xSSC/0.2%Tween at 37°C. nuclei were equilibrated in PBS and post-fixed in 2% formaldehyde/PBS for 10 min at room temperature. Followed three 5-min washes in PBS. The 3D-fixed nuclei were counterstained for 10 min with DAPI at room temperature. Followed two 5-min washes in PBS. Coverslips were mounted. An Eclipse Ti-E (Nikon Instruments, Florence, Italy) was used to scan the nuclei, with an axial

distance of between 2-2.5 micron consecutive sections

3D multicolour FISH analysis tool

In order to automatically analyse 3D multicolour FISH in fluorescence cell image z-stacks we developed a tool in MATLAB. The tool is capable to detect and localize fluorescent 3D spots by performing the 2D segmentation of cell nuclei and the detection of spots for each slice of the stack followed by the 3D reconstruction and identification of nuclei and spots. It then measures the relative positioning of spots in the nucleus and inter-spot distances which greatly enrich our understanding of the 3D spatial organization of the spots within cell nucleus. The tool implements the following algorithm:

for each slice n of the stack

$nuclei_n$ = nuclei_seg($I_{dapi,n}$) %Performs 2D nuclei segmentation

for each fluorescent field f=Atrogin1,4q,10q

$spot_{f,n}$ = detect_spot(I_f) %Performs 2D spot detection

$spot\_vol_f(:,:,n) = spot_{f,n}(:,:)$

endfor

$nuclei\_vol(:,:,n) = nuclei_n(:,:)$

endfor

nuclei_CC = bwconncomp(nuclei_vol)

nuclei_L = labelmatrix(nuclei_CC)

compute volume for each nucleus object in nuclei_CC

exclude nuclei whose volume is less than 10% of mean volumes

$\{NCL\}_M$ <- identified 3D nuclei
for each nucleus m in $\{NCL\}_M$
for each fluorescent field f=Atrogin1,4q,10q
$NCL_m.spot_f$ <- detected spots within the 3D nucleus NCL_m
$spot\_CC_f$ = bwconncomp($NCL_m.spot_f$)
compute volume for each spot object in $spot\_CC_f$
exclude spots whose volume is below the SD of the volumes
$\{SPT_f\}_{nf}$ <- identified 3D spots
endfor
if number of identified spots nf = 2 for each f=Atrogin1,4q,10q
nucleus m is deemed suitable for analysis
compute distances of any spot from any other spot
compute distances of any spot from the nuclear periphery
compute distances of any spot from the nuclear centroid
compute distance from nuclear periphery to nuclear centroid
endif
endfor

The function *nuclei_seg* performs a partition of cell image in nuclei regions and background implementing a region based segmentation algorithm. The function *detect_spot* has four major steps. It first filters the image I_f applying the Laplacian of Gaussian (LoG) operator (*fspecial* MATLAB function) of size 13 and standard deviation 7. This enhances the signal in the areas where objects are present. Then the function applies the h-dome transformation [65] that extracts bright structures by

cutting off the intensity of height h from the top, around local intensity maxima. We used h=0.5 with a neighborhood size of 19x19. We decided to not use a global operator after having observed that a spot in one part of the image could be lighter or darker than the background in another part. This is due to the facts that spots have inhomogeneous intensity distribution over the image and that the image may have an uneven background. In the third step, the function performs a thresholding on h-domes image that excludes pixels whose intensity values are below a threshold. The threshold is 1.96 standard deviations above the mean of domes intensity values. We therefore assumed that spot areas have significant intensity disparity with respect to other bright areas present in cell nucleus. Lastly, the function applies a thresholding operation based on the surface areas of the spots, in order to discard too small objects which are probably just noise. It filters out spots smaller than a surface area of 25. *detect_spot* produces an accurate set of spots.

spot_vol$_f$ and *nuclei_vol* are 3D arrays that contain the positions of the detected spots and nuclei from all slices.

3D reconstructions of nuclei are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 26). 3D nuclei are then labeled by applying the *labelmatrix* MATLAB function so they can easily separated each from the others.

The tool compute volume of each 3D reconstruction, discarding objects whose volume is less than 10% of mean volumes which

are just noise.

3D reconstructions of spots are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 26). Then a threshold operation is performed on the 3D spots to obtain the more significant ones: it keeps in all the spots whose volume is above the threshold the standard deviation of the volumes.

For each nucleus, the tool checks the number of identified spots for each fluorescent field. If there are exactly 2 spots for each fluorescent field that nucleus is deemed suitable for the analysis: the tool computes the distances of any spot from any other spot, the distances of any spot from the nuclear periphery and the nuclear centroid and the distance from the nuclear periphery to the nuclear centroid.

Total RNA extraction and cDNA production

Total RNA extraction was performed using the RNeasy Mini Kit (Qiagen), and the purified RNA was treated with RNase-free DNase (Qiagen) to remove any residual DNA, following the manufacturer's instructions. RNA was retrotranscribed using the SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher Scientific) following the suggested conditions. cDNA was finally diluted 1:3 for the following use and 1 ng for each reaction was used in qRT-PCR.

Quantitative RT-PCR (qRT-PCR) analysis was performed on an Step One Plus real time PCR detection system (Applied

Biosystem), using the power SYBR Green q-PCR master mix (Thermo Fisher Scientific). The relative expression of the investigated genes was quantified after normalization against glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and then reported as absolute or normalized on myoblasts.
The primer pairs used for real time amplifications are shown in Supplementary table 6.

Transfection

For DUX4 transfection human primary cells from controls and patients were plated in a 6-well dish containing a glass coverslip at 5×10^5 cells/well in growth medium without antibiotics. Cells were leaved to adhere for 4-5 h. After that 4 µg of plasmid DNA (pCMV-Myc or pCMV-DUX4, a gently gift of Gabellini's lab)) were diluted in room-temperature Opti-MEM® (Thermo Fisher Scientific) to a final volume of 250 µl. 10 µl of Lipofectamine® 2000 Reagent (Thermo Fisher Scientific) were diluted with Opti-MEM to a final volume of 250 µl to achieve a ratio of DNA weight to Lipofectamine volume that is 1:2.5. The Lipofectamine and plasmid DNA preparations in Opti-MEM were gently mixed and incubated for 20 min at room temperature. During this time medium were replaced with 2 ml of fresh medium. Transfection complexes were then added to the cell cultures, and the cultures were incubated at 37°C in 5% CO2. The following day samples were processed for the RNA extraction.

For 4q D4Z4 array transfection immortalized human primary cells from controls and patients were plated in a 12-well dish at 6×10^4 cells/well in growth medium without antibiotics. The following day, 2 µg of BAC DNA (RP11-2A16 for control or CH16-291A23 fo 4q D4Z4 repeat array) were diluted in room-temperature Opti-MEM® (Thermo Fisher Scientific) to a final volume of 50 µl with the addition of 4 µl of P3000™ Reagent (Thermo Fisher Scientific). 4 µl of Lipofectamine® 3000 Reagent (Thermo Fisher Scientific) were diluted with Opti-MEM to a final volume of 50 µl to achieve a ratio of DNA weight to Lipofectamine volume that is 1:2. The Lipofectamine and BAC DNA (plus P3000) preparations in Opti-MEM were gently mixed and incubated for 20 min at room temperature. During this time medium were replaced with 600 µl of fresh medium. Transfection complexes were then added to the cell cultures, and the cultures were incubated at 37°C in 5% CO2. The day after the transfection medium was replaced with growth medium and the cells were maintained under standard conditions until the following day for the next DNA and RNA extraction.

The primer pairs used for PCR or real time amplifications are shown in Supplementary table 6.

Immunofluorescence

For immunofluorescence staining of DUX4 cells transfected with plasmids containing or not DUX4 ORF were fixed in 4% paraformaldehyde in PBS-T for 15 minutes at room temperature

and then washed twice with PBS. Cells were permeabilized with 0,5% Triton X-100 in PBS for 10 minutes at room temperature with gentle rocking and then washed three times with PBS. In order to prevent the nonspecific binding of the antibodies, before using antibodies to detect proteins, all epitopes on cells should be blocked in 4% BSA in PBS-T for 30 minutes at room temperature with gentle rocking.

Cells were incubated for 3 h at room temperature with gentle rocking with the primary rabbit-DUX4 antibody directed against the C-terminal region of DUX4 (E5-5; 1:200, Abcam, ab124699), diluted in 4% BSA in PBS-T.

After washing three times in PBS-T, followed an incubation with diluted Alexa 488 conjugated donkey anti-rabbit

(A21206, 1:1000, Thermo Fisher Scientific) for one hour, gently rocking in the dark. Next, cells were washed three times with PBS-T and counterstained for 10 min with DAPI diluted in PBS-T at room temperature. Followed two washes in PBS-T and two in PBS, coverslips were mounted. A Eclipse Ti-E (Nikon Instruments, Florence, Italy) was used to scan the cells.

DNA extraction

Total DNA extraction from transfected cells was performed in 200 μl of DNA extraction Buffer (Tris-HCl 20 mM, EDTA 5 mM, 150 mM NaCl) with the addition of 0,5 % SDS and 80 μg of proteinase K (SIGMA) at 37°C overnight with bland agitation.

The day after DNA fragments were purified by phenol–

chloroform and phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation and then resuspended in 10 mM Tris-HCl pH 7.5. DNA templates were used to perform qualitative PCR with DreamTaq DNA Polymerase (Thermo Fisher Scientific). Primers used recognized a sequence specific of the BAC vector and common to both BAC used in the experiment. The PCR products confirming the presence of BAC in transfected cells DNA preparations were shown.

The primer pairs used for the real time amplifications are shown Supplementary table 6.

Statistics

For all graphs, mean value and standard error of the mean are presented with the number of replicates indicated in the figure legends. p values and applied statistical tests are indicated in the figure legends.

Figure legends

Figure 1. Definition of 4q specific D4Z4 interactome that is epigentically deregulated in FSHD

(A) Integrated “omics” approach to decipher chromatin structure alterations in FSHD. To probe genome/chromatin remodeling in FSHD we sought to develop a chr4q specific 4C-seq approach. ChIP-seq datasets for classical histone marks were used to define the chromatin states with ChromHMM and RNA-seq published datasets [47] to get insights in the transcriptional deregulations of the structurally deregulated loci.

(B) Scheme of the 4q-allele-specific 4C-seq. The 40kb proximal region of the polymorphic chr4q D4Z4 array displays more than 98% of sequence identity with chr10q26. We took advantage of a Single Sequence Length Polymorphism (SSLP) upstream the D4Z4 array to design a paired-end allele-specific 4C. Mate reads give us the possibility to precisely assign the chromosome of origin (chr4 and chr10) as well as the contribution of chr4 alleles (4qA161, 4qB163 and 4qB168) and chr10 alleles (10qA166) to the interactions we retrieved.

(C) Genome browser snapshot of the FSHD locus (4q35) depicting 4C normalized coverage tracks for 4q and 10q viewpoints in CN and FSHD.

(D) Circus plots depicting all the interactions mediated by the D4Z4 proximal region sslp on chr4q in control (CN) and patients (FSHD). Cis interaction are depicted in red, trans interactions

between sslp and chr8 (see later) in yellow and chr10 in green. Other trans interactions are depicted in grey.

(E) ChromHMM 15-state model of CN and FSHD myoblasts (MB) and myotubes (MT) obtained with ChIP-seq datasets we generated for H3K4me1, H3K4me3, H3K27ac, H3K27me3 and H3K36me3 (left, emission parameters, right, transition parameters). Integrations of our ChIP-seq datasets enabled us to define 15 chromatine states, according to the nomenclature of Roadmap.

(F) Genome browser snapshot of SORBS2 gene region with the ChromHMM and ChIP-seq signal tracks in CN and FSHD myoblasts and myotubes with 4q- and 10q-4C coverage.

(G) Jaccard values of each chromatin state between CN and FSHD (myoblasts and myotubes togheter). Each dot shows one pairwise comparison (lines indicate medians).

Figure 2. FSHD deregulated interactions show genes enriched for active enhancers that are also DUX4 indipendent

(A) Venn diagrams of chr4q sslp interacting regions and associated protein-coding genes in CN and FSHD.

(B) Venn diagrams depicting the intersection of genes that we found in interactions that are lost (up) or gained (bottom) by FSHD, genes that showed active enhancers in FSHD as well as genes transcriptionally regulated or not regulated by DUX4 [47].

(C-E) Gene ontology analysis of genes found in deregulated

interactions (lost or gained) with active enhancer expressed and DUX4 dependent (n=433, 220+213) or indipendent (n=1637, 859+778). Significant enriched terms were clustered according to semantic similarity and represented as heatmaps.
(D-F) Snapshots of an enlarged view of gene DUX4 dependent or independent present in enriched terms of gene ontology shown in C-E respectively, showing ChromHMM and ChIP-seq signal tracks.

Figure 3. Insight a functional trans interaction relevant for the atrophic phenotype

(A) Algorithm workflow exemplified: image datasets are processed first segmenting nuclei in 2D and detecting spots in each slice, then spots are 3D reconstructed within nucleus and all the measurements are done.
(B) Representative nuclei of 3D DNA multicolor FISH in CN (up) and FSHD (bottom). Nuclei are counterstained in DAPI (blue), Atrogin1 locus (red), 4q35.1 region (green) and 10q26.3 region (pink). The scale bar is indicated.
(C) Bar plot show the percentage of interprobe distances below the specified cutoff, set at the distance that included the sum of lowest ten percentile of measurements for the interactions and distance between Atrogin1 from 4q35.1 and 10q26.3 probes in CN and FSHD. n=3. Approximately measurements derived from 200-300 nuclei for each condition are used. Fisher one-tailed test $p<0.05$.

(D) Frequency distributions from the nuclear centroid of Atrogin1 (red), 4q35.1 (green) and 10q26.3 (pink) spots in CN (left) and FSHD (right). n=3. Approximately measurements derived from 200-300 nuclei for each condition are used. Atrogin1/4q35.1 and Atrogin1/10q26.3 Wilcoxon two-tailed test $p<0.001$ in both CN and FSHD.

(E) Frequency distributions from the nuclear centroid of 4q35.1 interacting with Atrogin1 (light green) and 4q35.1 not interacting (dark green) in CN (left) and FSHD (right). n=3. Approximately measurements derived from 200-300 nuclei for each condition are used. 4q35.1 interacting/not interacting Man U two-tailed test $p<0.05$ in both CN and FSHD.

(F) Bar plots showing enrichment of RNA Pol II at Atrogin1 promoter (left) and in an internal region (right) assessed by ChIP-PCR experiment in both CN and FSHD. n=3. S.e.m. is indicated.

(G) Bar plot showing expression levels of Atrogin1 gene during CN and FSHD differentiation (MB, myoblasts, MT2, myotubes day 2, MT4, myotubes day 4, MT6, myotubes day 6). Data were normalized on *Gapdh* expression. n=3. S.e.m. is indicated. MT2 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.01$; MT4 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.05$.

Figure 4. From bottom to top: Atrogin1 interactome deregulation in FSHD and recovery of Atrogin1 expression through a new 4q D4Z4 array

(A) Genome browser snapshot of the Atrogin1 locus depicting 4C normalized coverage tracks in CN and FSHD as well as chromatin segmentation tracks.

(B) Circus plots depicting all the interactions mediated by the Atrogin1 promoter region in CN and FSHD. Cis interaction are depicted in yellow, trans interactions between Atrogin1 and chr4 in red and chr10 in green. Other trans interactions are depicted in grey.

(C) Venn diagrams of Atrogin1 interacting regions and associated protein-coding genes in CN and FSHD.

(D) Pie charts showing percentages of genes from control Atrogin1 interactome and FSHD Atrogin1 interactome that present or not statistically significant chromatin switches as assessed with chromDiff software.

(E) Gene ontology analysis of genes of FSHD Atrogin1 interactome that shows chromatin state switch. Significant enriched terms were clustered according to semantic similarity and represented as heatmaps.

(F) Representative nuclei of a 3D DNA multicolor FISH in two FSHD2 cell lines, FSHD2-1 (up) and FSHD-2 (bottom). Nuclei are counterstained in DAPI (blue), Atrogin1 locus (red) and 4q35.1 region (green). The scale bar is indicated.

(G) Bar plot show the percentage of interprobe distances below

the specified cutoff, set at the distance that included the sum of lowest ten percentile of measurements for the interactions and distance between Atrogin1 from 4q35.1 probes in CN, FSHD1 and FSHD2. n=3. Approximately measurements derived from 100-200 nuclei for each condition are used. FSHD1/CN Fisher one-tailed test $p<0.05$; FSHD2/CN Fisher one-tailed test $p<0.05$.

(H) Bar plots showing expression levels of Atrogin1 gene in CN and FSHD MB after transfection with a control BAC (Ctrl BAC) and BAC containing a piece of 4q D4Z4 array (BAC D4Z4n). Data were normalized on *Gapdh* expression. n=at least 3. S.e.m. is indicated. CN MB Ctrl BAC/BAC D4Z4n Paired one-tailed t test $p<0.05$; FSHD MB Ctrl BAC/BAC Paired one-tailed t test D4Z4n $p<0.01$.

Supplementary figures legend

Supplementary Figure S1.

(A) Scatter plots of Pearson's correlation coefficient of samples for the 4q-D4Z4 4C showing reproducibility between biological samples of CN (C5 and C6) (left) and FSHD (F5 and F6) (right) and bar plots showing percentage of mapped reads in cis and in trans for the 4q-D4Z4 viewpoint.

(B) Scatter plots of Pearson's correlation coefficient of samples for the 10q-D4Z4 4C showing reproducibility between biological samples of CN (C5 and C6) (left) and FSHD (F5 and F6) (right) and bar plots showing percentage of mapped reads in cis and in trans for the 10q-D4Z4 viewpoint.

Supplementary Figure S2.

(A) Genome browser snapshot of whole chr4 with the ChromHMM tracks in CN and FSHD myoblasts and myotubes with 4q- and 10q-4C coverage.

(B) Genome browser snapshot of whole chr10 with the ChromHMM tracks in CN and FSHD myoblasts and myotubes with 4q- and 10q-4C coverage.

(C) Genome browser snapshot of FRG1 gene region with the ChromHMM tracks in CN and FSHD myoblasts and myotubes with 4q- and 10q-4C coverage.

Supplementary Figure S3.

(A) Clustered heatmap of genome-wide enrichment signal of all the chip-seq datasets in myoblasts we generated and used for each replicate (left). Samples were clustered according to Pearson correlation coefficient. Clustered heatmap of genome-wide enrichment signal of the chip-seq datasets we generated and corresponding datasets from ENCODE (right).

(B) Clustered heatmap of genome-wide enrichment signal of all the chip-seq datasets in myotubes we generated and used for each replicate (left). Samples were clustered according to Pearson correlation coefficient. Clustered heatmap of genome-wide enrichment signal of the chip-seq datasets we generated and corresponding datasets from ENCODE (right).

(C) Clustered heatmap of genome-wide enrichment signal of all the chip-seq datasets we generated and used, depicting merged biological replicates for each histone mark and myoblasts (left) and myotubes (right).

(D) PCA representation of genome-wide enrichment signals for all the histone marks we analyzed in myoblasts (left) and myotubes (right).

Supplementary Figure S4.

(A) Scheme of the approach used to detect overlap between genomic regions interacting with 4q sslp as detected by 4C-seq and several features such as chromatin states, CpG islands, genes and Lamin-Associated Domains (LADs). GAT (Genomic

Association Tester) [61] was used to perform 10,000 random permutations and compute the significance of overlap between 4C interacting regions and above-mentioned features/intervals. Results of significant overlap between features is given as Fold Enrichment of observed versus Expected overlap corrected for FDR empirical p value.

(B) Overlap enrichment of selected chromatin states and common interacting regions (Common), CN-specific regions (CN) and FSHD-specific regions (FSHD) of 4q sslp 4C.

(C) Overlap enrichment of CpG islands, Refseg protein coding genes, lincRNAs and LADs and common interacting regions (Common), CN-specific regions (CN) and FSHD-specific regions (FSHD) of 4q sslp 4C.

Supplementary Figure S5.

(A) Diagram of the genomic region analyzed in the chromosome conformation capture (3C) experiments, indicating the PvuII restriction sites (thin vertical lines); the arrowheads indicate the primer positions, the numerical series in Atrogin1 locus and * in the D4Z4 repeats; crosslinking frequency is expressed as the ratio of polymerase chain reaction (PCR) performed on 3C samples relative to bacterial artificial chromosome (BAC) controls between the fixed PvuII fragment * (D4Z4 repeats) and the Atrogin1 locus.

n=3. S.e.m. is indicated. CN MB One-way ANOVA $p<0.001$; FSHD MB One-way ANOVA $p<0.001$.

(B) Representative nucleus of a 3D DNA multicolor FISH in CN. Nuclei are counterstained in DAPI (blue), D4Z4 repeat (green), 4q35.1 region (red) and 10q26.3 region (pink). The scale bar is indicated.

(C) Box & whiskers plot showing the distribution of distances of 4q and 10q D4Z4 spots from 4q35.1 and 10q36.3 regions.

(D) Representative nuclei of 3D DNA multicolor FISH in CN and FSHD. Nuclei are counterstained in DAPI (blue), 4q35.1 region (green) and 10q26.3 region (pink). The scale bar is indicated.

(E) Bar plot show the percentage of interprobe distances below the specified cutoff, set at the distance that included the sum of lowest ten percentile of measurements for the interactions and distance between 4q35.1 from 10q26.3 probes in CN and FSHD. n=3. Approximately measurements derived from 200-300 nuclei for each condition are used.

(F) Frequency distributions from the nuclear centroid of 10q26.3 interacting with Atrogin1 (light pink) and 10q26.3 not interacting (dark pink) in CN (left) and FSHD (right). n=3. Approximately measurements derived from 200-300 nuclei for each condition are used.

(G) Frequency distributions from the nuclear centroid of Atrogin1 interacting with 4q35.1 (light red) or 10q26.3 (orange) and Atrogin1 not interacting (dark red) in CN (left) and FSHD (right). n=3. Approximately measurements derived from 200-300 nuclei for each condition are used.

Supplementary Figure S6.

(A) Bar plots showing percentage of input enrichment of RNA Pol II at DBE and two regions for positive and negative binding of RNA PolII through ChIP experiment in both CN and FSHD. n=3. S.e.m. is indicated. For PolII Ctrl + FSHD/CN $p<0.05$.

(B) Bar plots showing expression levels of DBE-T, DUX4-fl, MyoD, MYHC2, Myogenin genes during CN and FSHD1 differentiation (MB, myoblasts, MT2, myotubes day 2, MT4, myotubes day 4, MT6, myotubes day 6). Data were normalized on *Gapdh* expression. n=3. S.e.m. is indicated. For DBE-T MT6 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.05$; for DUX4-fl MT4 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.05$.

(C) Bar plots showing expression levels of DBE-T, DUX4-fl, MyoD, MYHC2, Myogenin genes during CN and FSHD2 differentiation (MB, myoblasts, MT2, myotubes day 2, MT4, myotubes day 4, MT6, myotubes day 6). Data were normalized on *Gapdh* expression. n=3. S.e.m. is indicated. For DBE-T MT6 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.05$; for DUX4-fl MT4 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.05$.

(D) Bar plots showing expression levels of Atrogin1 gene during CN, FSHD6 (an FSHD1 cell line) (left) and FSHD2 (right) differentiation (MB, myoblasts, MT2, myotubes day 2, MT4, myotubes day 4, MT6, myotubes day 6). Data were normalized on *Gapdh* expression. n=3. S.e.m. is indicated. For FSHD6

MT2 and MT4 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.001$; MT6 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.01$; For FSHD2 MT4 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.001$; MT6 FSHD/CN Two-way ANOVA and Bonferroni post-tests $p<0.05$;

Supplementary Figure S7.

(A) Representative field of an immunofluorescence in CN and FSHD transfected with plasmid containing DUX4 (left). Nuclei are counterstained in DAPI (blue) and DUX4 protein (green). The scale bar is indicated. Bar plot showing the percentage of transfected cells in CN and FSHD. n=3. S.e.m. is indicated.

(B) Bar plots showing expression levels of Atrogin1, MYHC2, Myogenin, DUX4-ORF, DUX4 Ctrl + (gene positively regulated by DUX4) and DUX4 Ctrl – (gene not regulate dby DUX4) genes in CN and FSHD MB after transfection with a control plasmid (Mock) and a plasmid containing DUX4 (DUX4). Data were normalized on *Gapdh* expression. n=3. S.e.m. is indicated.

Supplementary Figure S8.

(A) Scatter plots showing reproducibility between biological samples of CN (C5 and C6) and FSHD (F5 and F6)(left) and bar plots showing percentage of mapped reads in cis and in trans for all the samples with Atrogin1 viewpoint.

(B) Gene ontology analysis of genes of CN Atrogin1

interactome that shows chromatin state switch. Significant enriched terms were clustered according to semantic similarity and represented as heatmaps.

(C) Gel electrophoresis showing PCR products. DNA extracting from control and transfected CN and FSHD cells were amplified using primers that recognize portions of BAC vector to validate transfection.

(D) Bar plots showing expression levels of Myogenin and FOXO3 genes in CN and FSHD MB after transfection with a control BAC (Ctrl BAC) and BAC containing a piece of 4q D4Z4 array (BAC D4Z4n). Data were normalized on *Gapdh* expression. n=at least 3. S.e.m. is indicated.

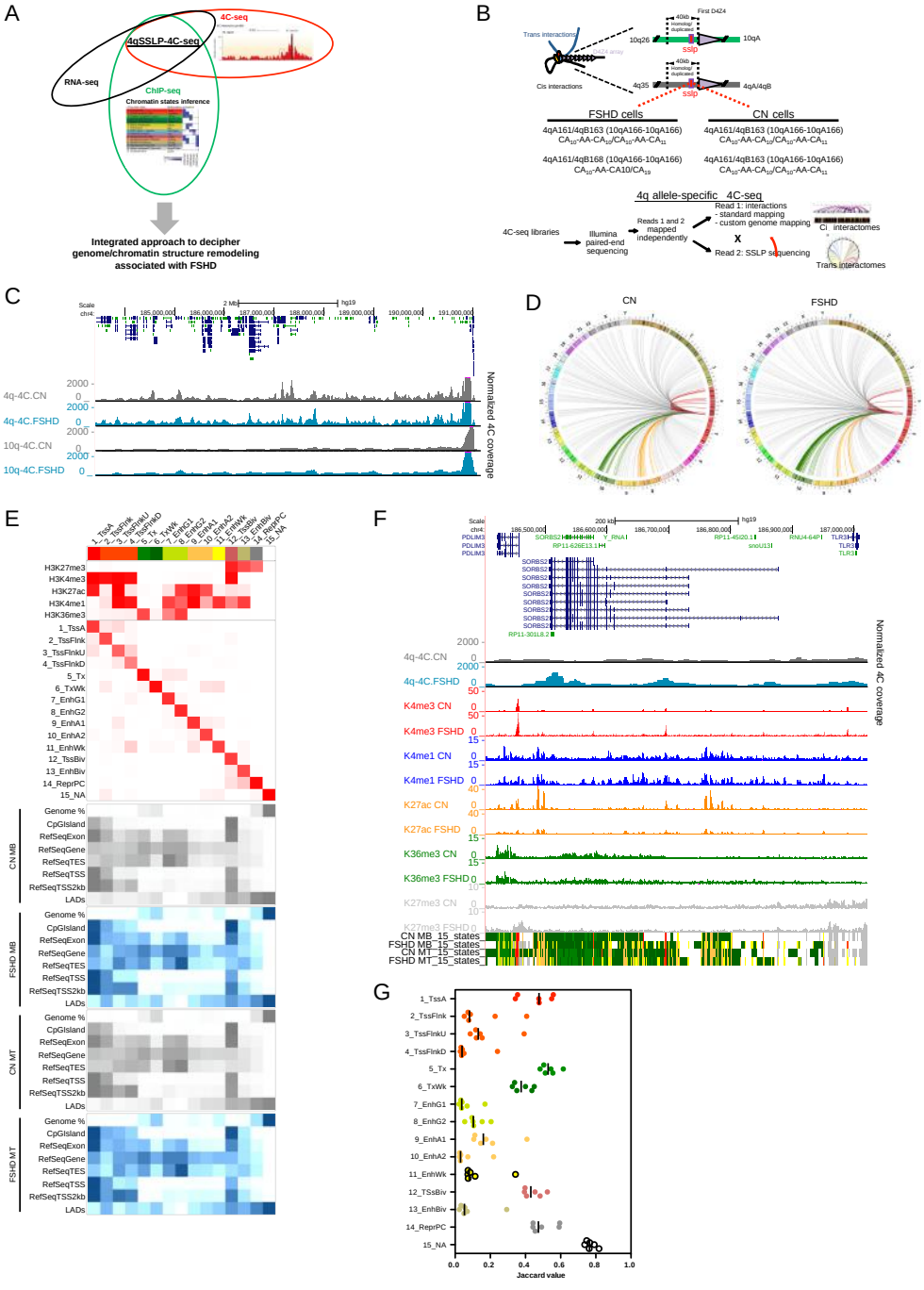


Figure 1. Definition of 4q specific D4Z4 interactome that is epigentically deregulated in FSHD

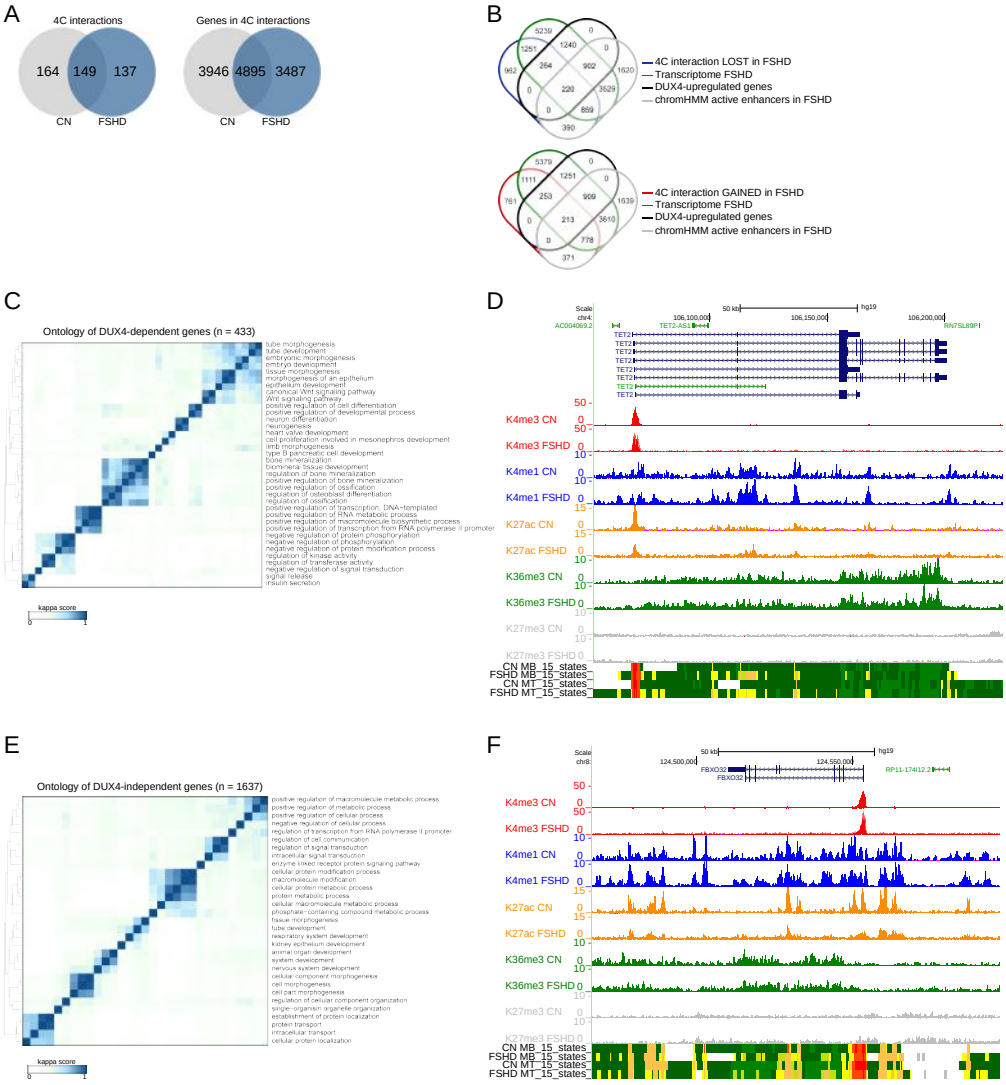


Figure 2. FSHD deregulated interactions show genes enriched for active enhancers that are also DUX4 indipendent

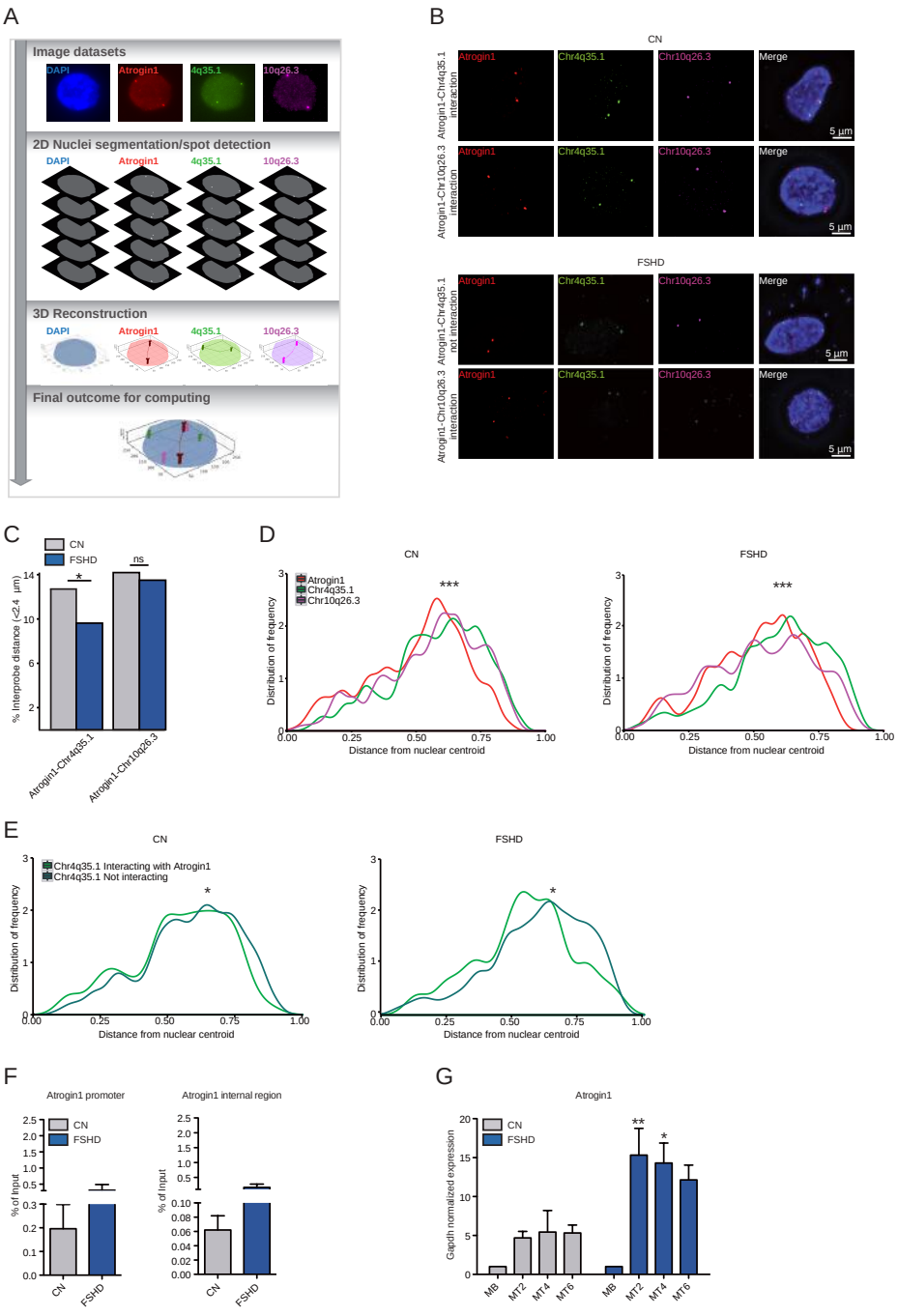


Figure 3. Insight a functional trans interaction relevant for the atrophic phenotype

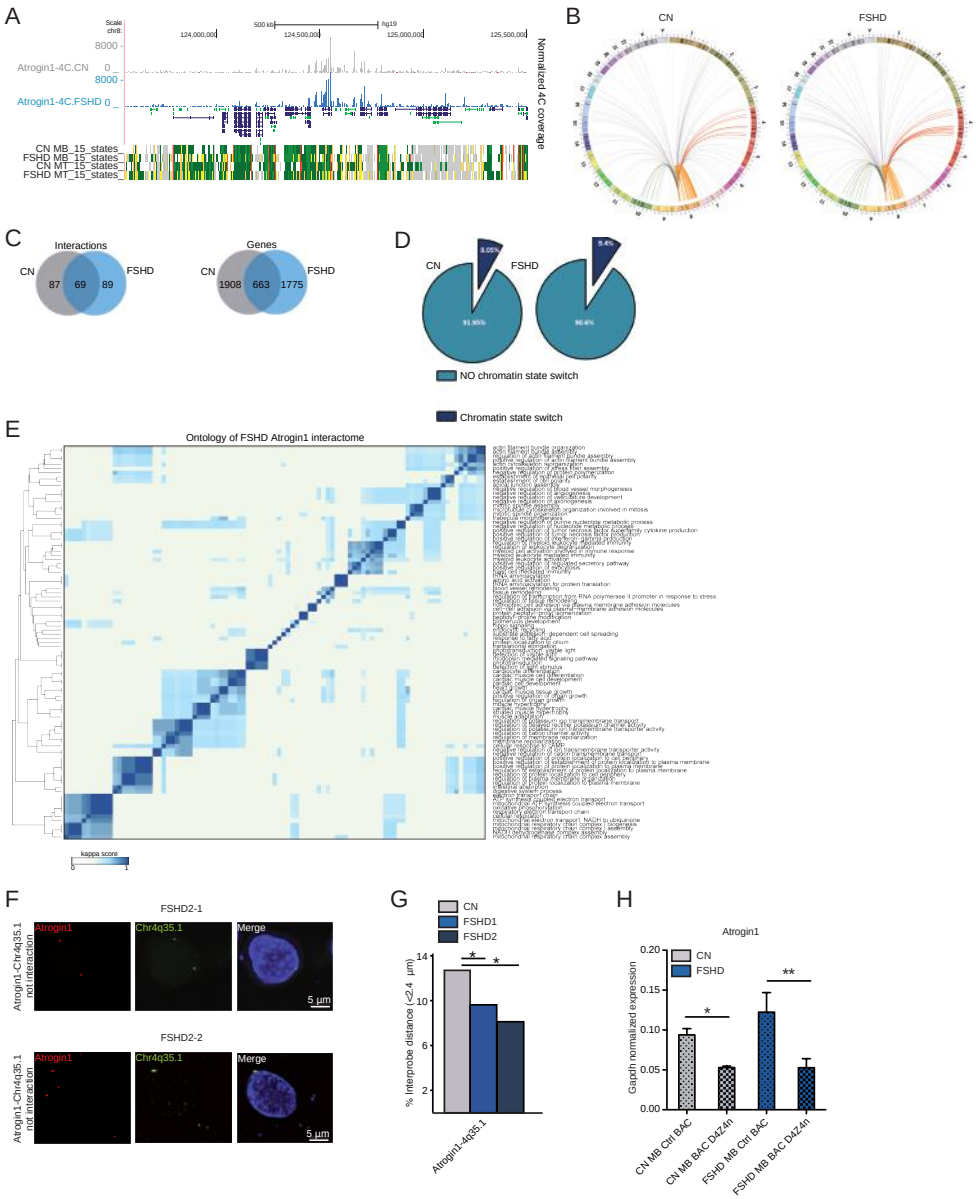
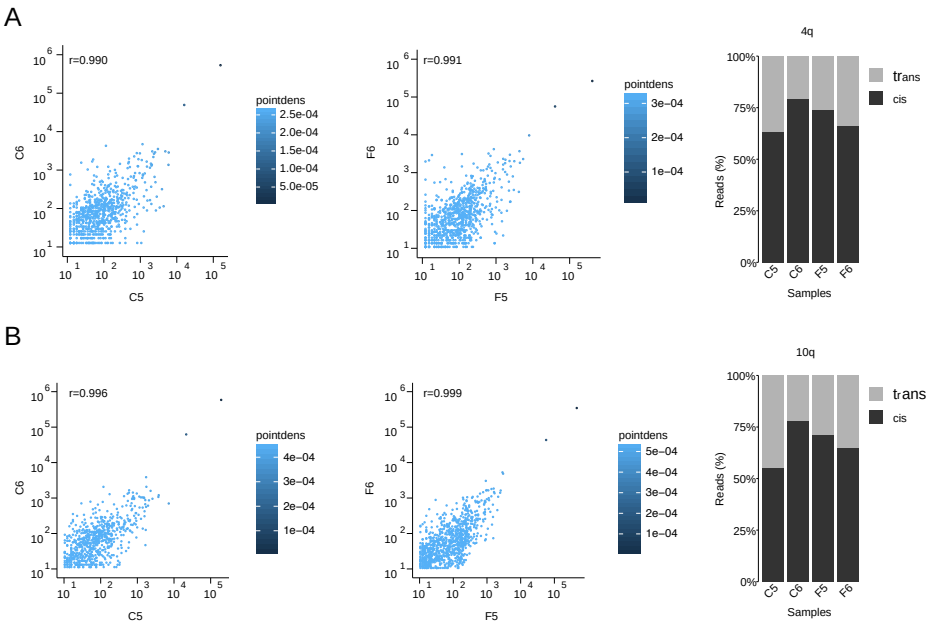
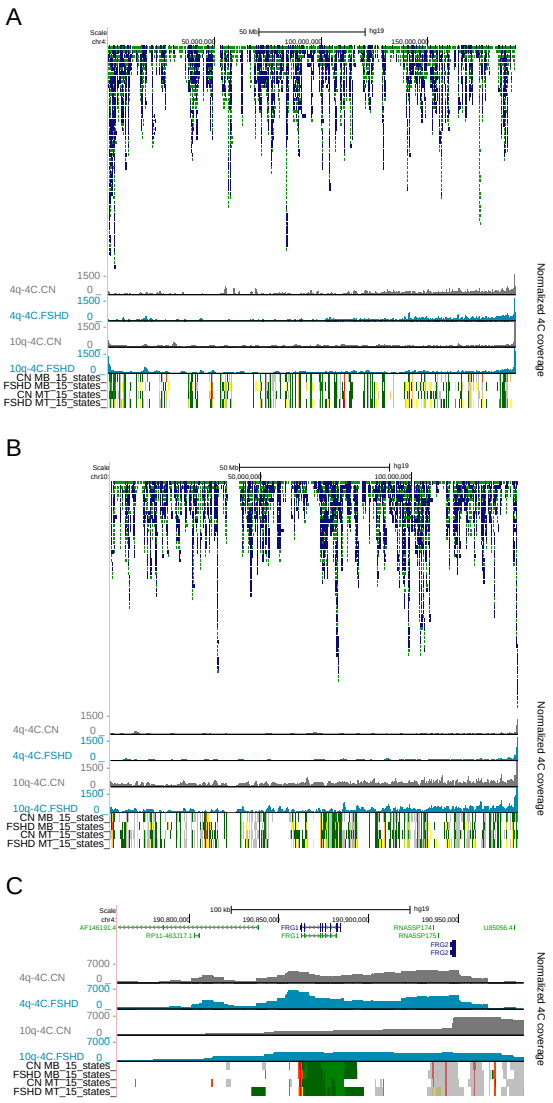


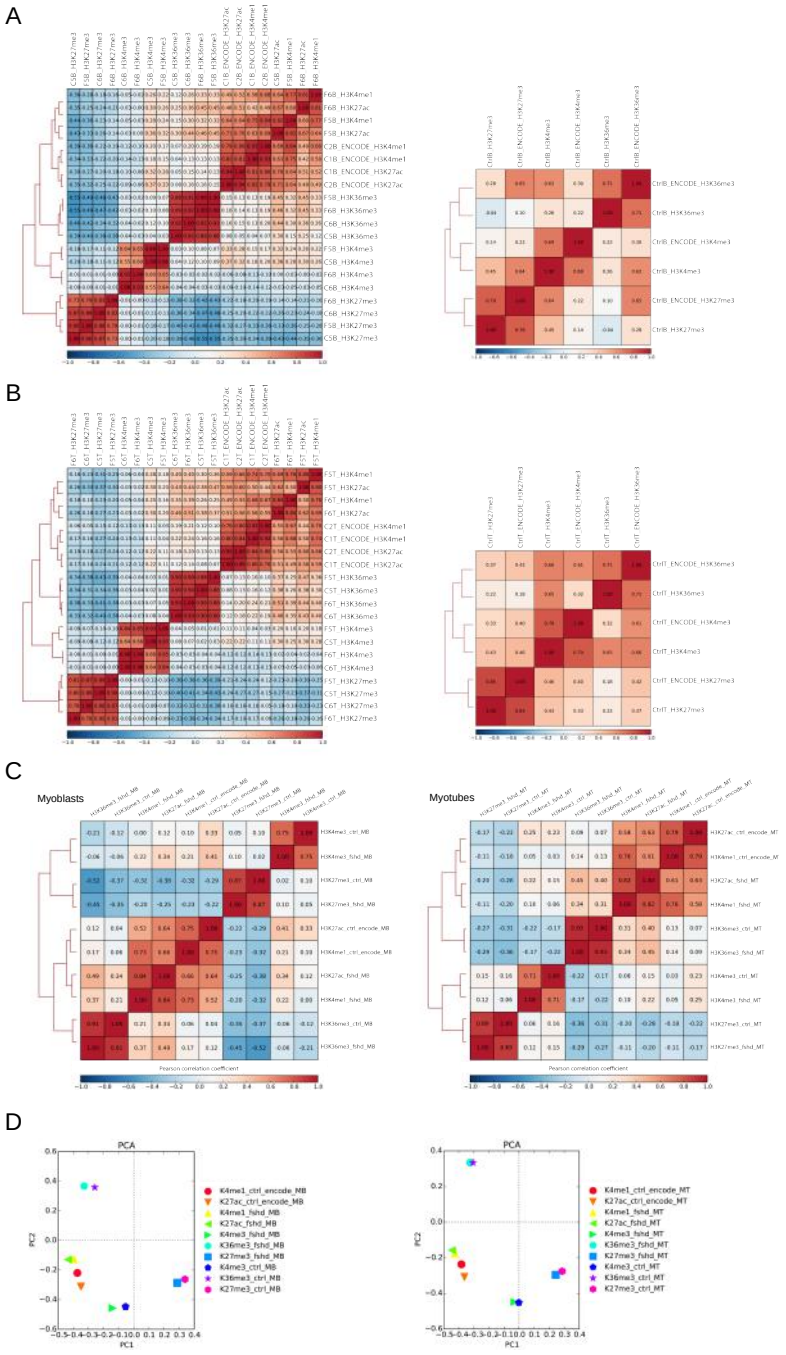
Figure 4. From bottom to top: Atrogin1 interactome deregulation in FSHD and recovery of Atrogin1 expression through a new 4q D4Z4 array



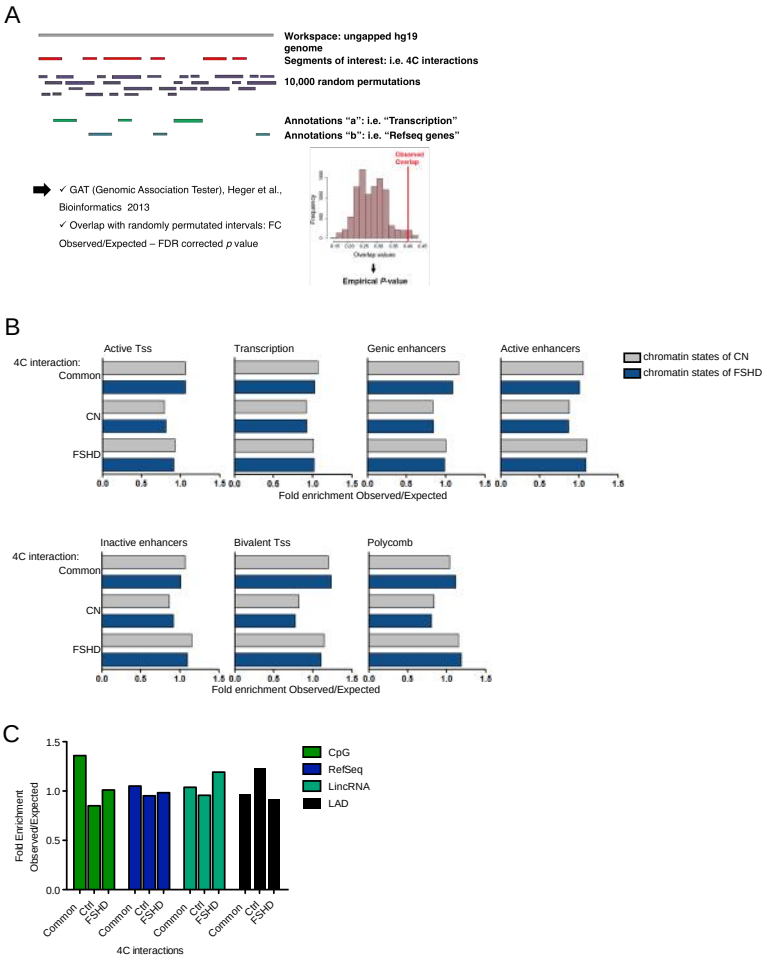
Supplementary Figure S1.



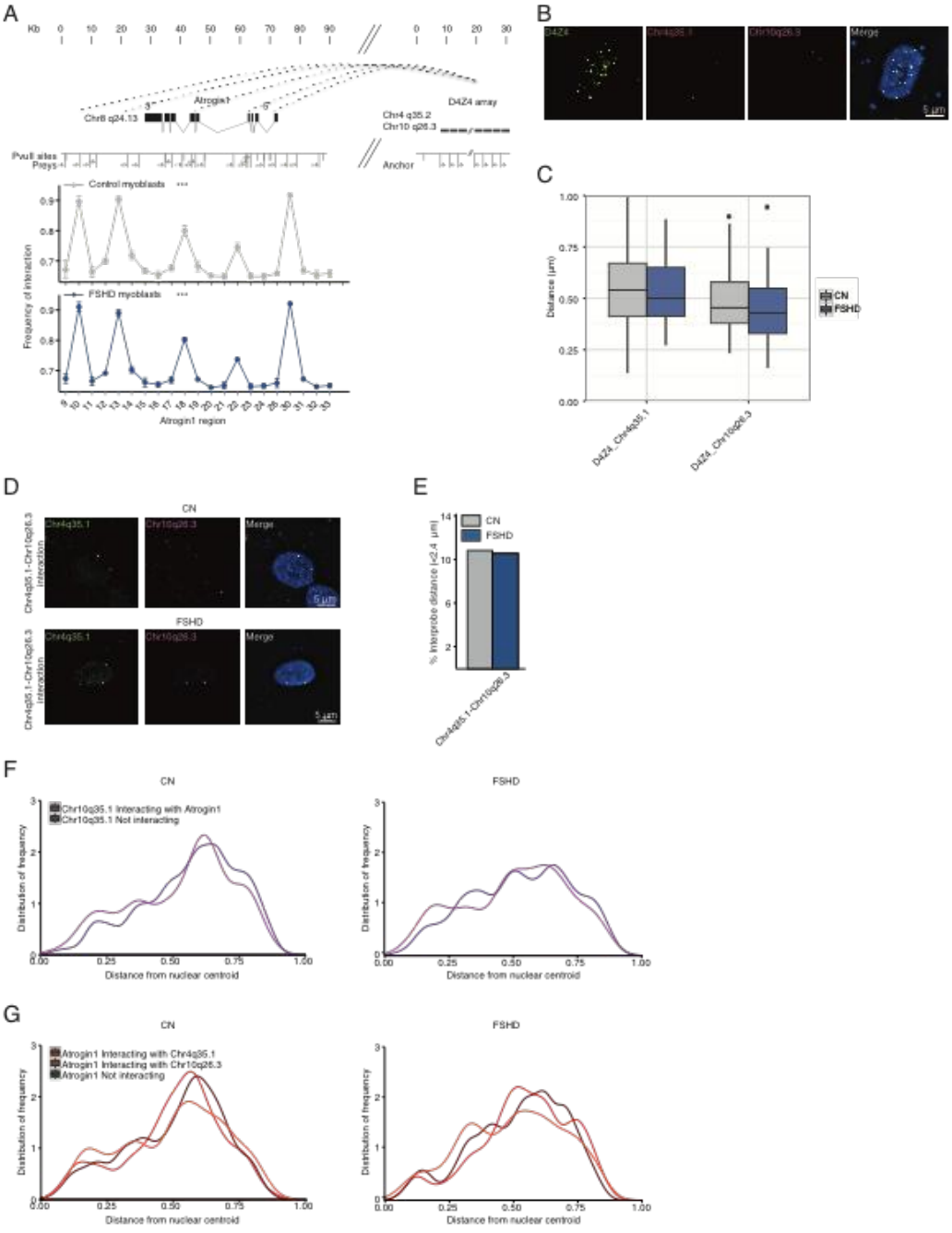
Supplementary Figure S2.



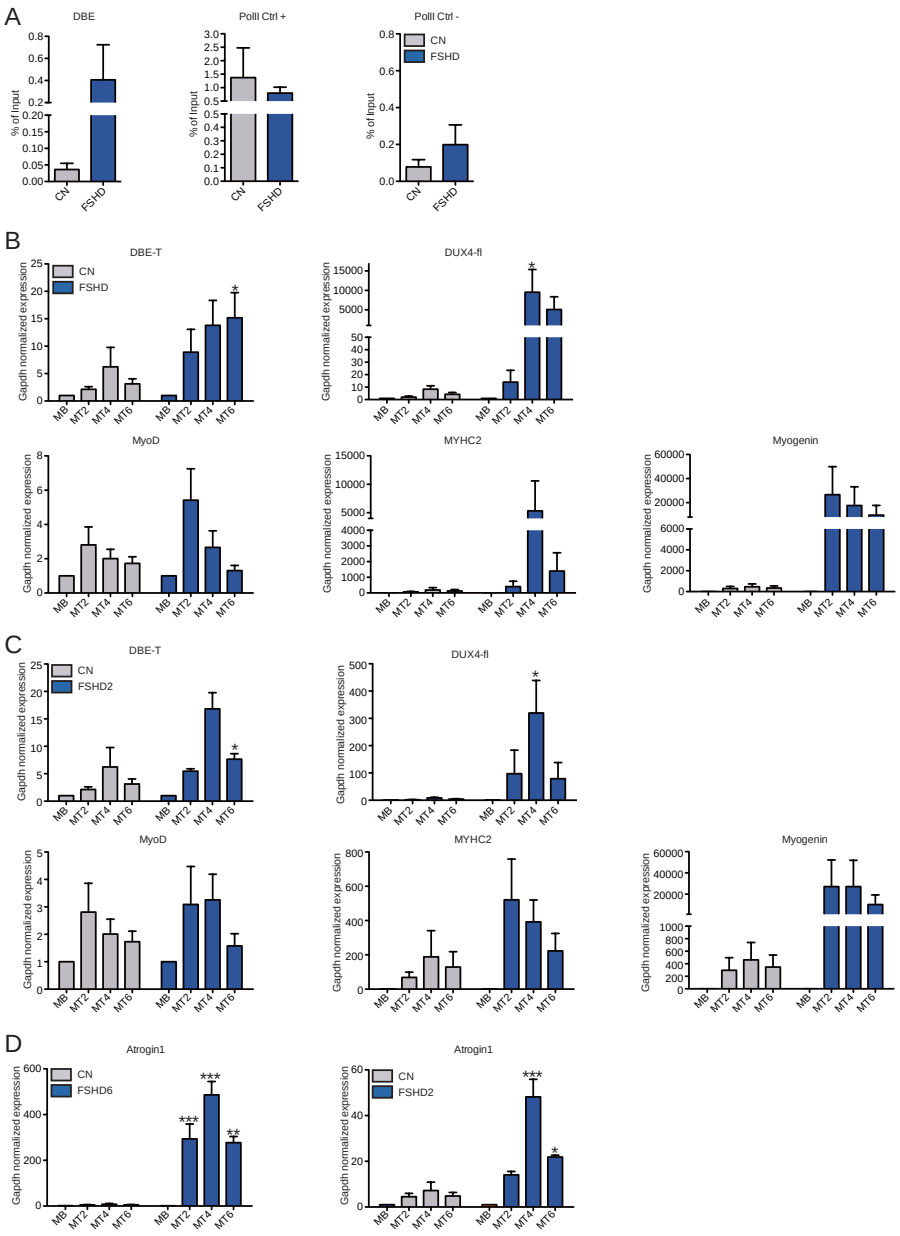
Supplementary Figure S3.



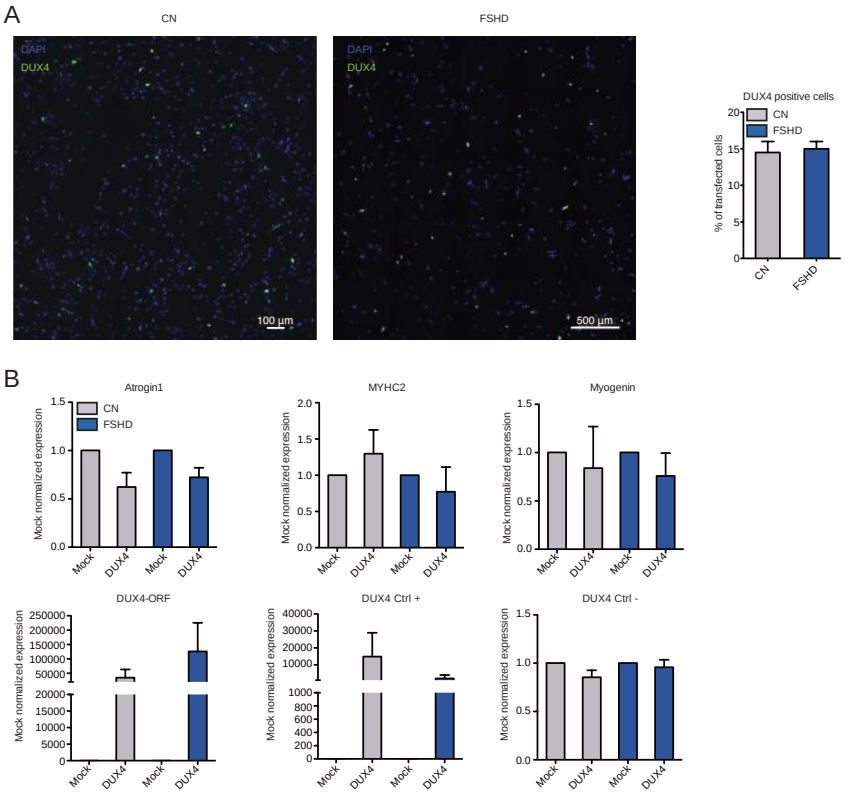
Supplementary Figure S4.



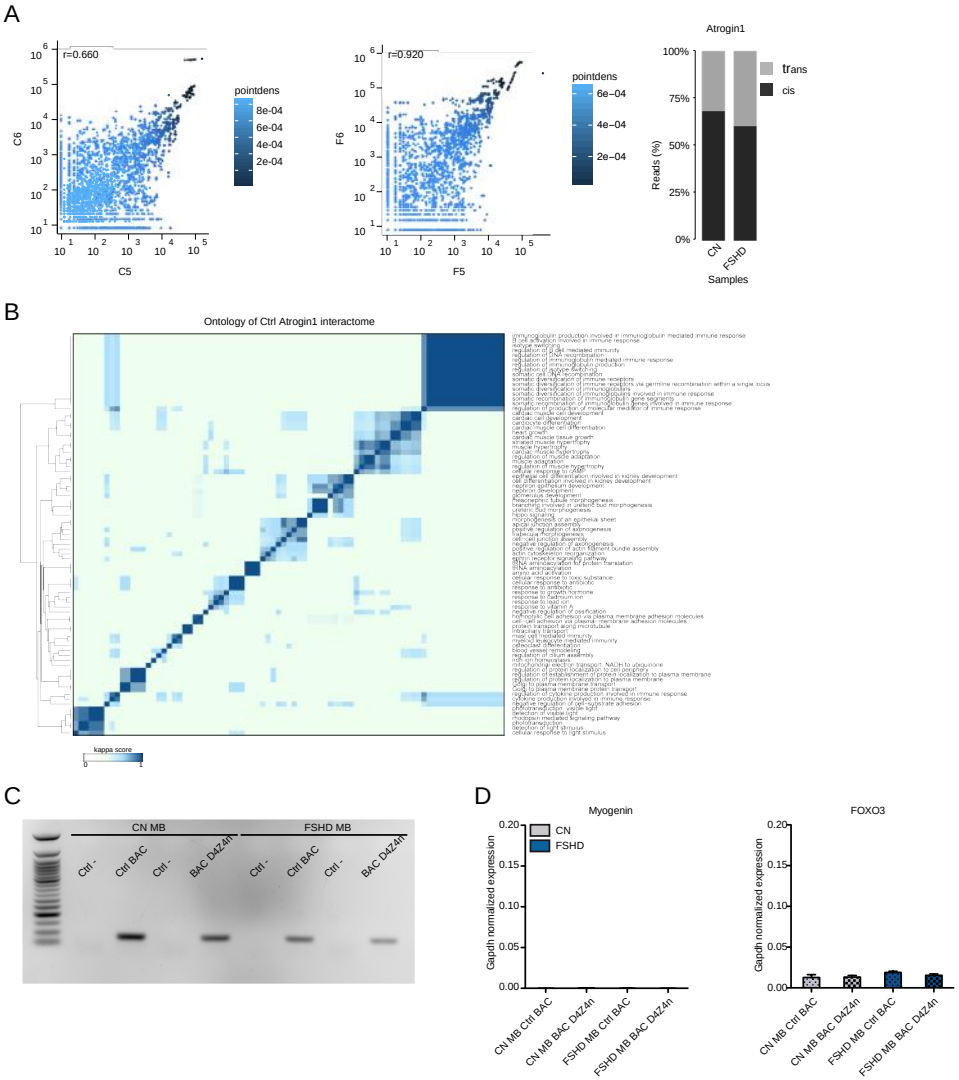
Supplementary Figure S5.



Supplementary Figure S6.



Supplementary Figure S7.



Supplementary Figure S8.

Sample Provenience	Sample type	Sample ID	Sex	Age at biopsy	Allele size (Kb)
Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA	Human primary myoblasts from healthy donors	CN5	M	43	65
Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA	Human primary myoblasts from healthy donors	CN6	F	42	65
Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA	Human primary myoblasts from FSHD1 patients	FSHD5	F	45	18
Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA	Human primary myoblasts from FSHD1 patients	FSHD6	M	52	22
Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA	Human primary myoblasts from FSHD2 patients	FSHD2-1	F	62	54
Fields Center for FSHD of the Rochester Medical Center Dept. of Neurology, New York, USA	Human primary myoblasts from FSHD2 patients	FSHD2-2	M	26	44
Telethon BioBank of the C. Besta Neurological Institute, Milano, Italy	Human primary myoblasts from healthy donors	CN2	M	22	NA
Telethon BioBank of the C. Besta Neurological Institute, Milano, Italy	Human primary myoblasts from healthy donors	CN3	M	35	NA
Telethon BioBank of the C. Besta Neurological Institute, Milano, Italy	Human primary myoblasts from FSHD1 patients	FSHD1	M	27	22
Telethon BioBank of the C. Besta Neurological Institute, Milano, Italy	Human primary myoblasts from FSHD1 patients	FSHD4	M	33	30
University of Massachusetts Medical School Wellstone center for FSH Muscular Dystrophy Research, Wellstone Program & Dept. of Cell & Developmental Biology, Worcester, MA USA	Human primary immortalized myoblasts from healty donors	CN im1	M	46	>40
University of Massachusetts Medical School Wellstone center for FSH Muscular Dystrophy Research, Wellstone Program & Dept. of Cell & Developmental Biology, Worcester, MA USA	Human primary immortalized myoblasts from FSHD patients	FSHD im1	M	42	18

Supplementary Table 1.

Sample ID	Sex	Race	Age at biopsy	No. of Pre-served 4qA Units	Allele 4$_1$				Allele 4$_2$				Allele 10$_1$				Allele 10$_2$				Hypo-methylated
					Del Size (Kb)	A/B	SSLP	SSLP seqence	Del Size (Kb)	A/B	SSLP	SSLP sequence	Del Size (Kb)	A/B	SSLP	SSLP sequence	Del Size (Kb)	A/B	SSLP	SSLP sequence	
CN5	M	WH	43	18	65	A	161	CA10AACA10	101	B	163	CA10AACA11	118	A	166	CA9AACA8	127	A	166	CA9AACA8	N
CN6	F	WH	42	74	250	A	161	CA10AACA10	65	B	163	CA10AACA11	95	A	166	CA9AACA8	100	A	166	CA9AACA8	N
FSHD5	F	WH	45	4	18	A	161	CA10AACA10	27	B	163	CA10AACA11	33	A	166	CA9AACA8	126	A	166	CA9AACA8	N
FSHD6	M	WH	52	5	22	A	161	CA10AACA10	99	B	168	CA19	72	A	166	CA9AACA8	78	A	166	CA9AACA8	N

Supplementary Table 2.

Sample ID	4C SSLP reading primer sequences with barcodes	4C bait reading primer sequences with barcodes	Sequencing setting	Run	No. of reads	No. of reads passing filter	No. of trimmed paired-end reads	No. of reads with SSLP					No. of reads after pooling technical and biological replicates				No. of aligned reads				No. of reads with DpnII sites				% Cis reads			
								Total	4q	Allele 4.	Allele 4.	10q	4q	Allele 4.	Allele 4.	10q	4q	Allele 4.	Allele 4.	10q	4q	Allele 4.	Allele 4.	10q	4q	Allele 4.	Allele 4.	10q
CN5	TTGACTGGTGGAGTTCTGGTTTCAGC	CGTGATGATGCCTGGCTTAATCAAGCTT	75 PE	Run1	4,989	3,754	2,835	0,677	0,306	0,137	0,061	0,368	1,449	3,876	0,573	2,547	0,787	0,477	0,310	1,454	0,575	0,347	0,228	0,961	68,064	66,295	65,737	61,144
				Run2	6,037	4,682	2,845	0,641	0,245	0,148	0,089	0,396																
				Run3	5,121	3,939	2,431	0,589	0,223	0,132	0,091	0,347																
				Run4	4,513	3,540	2,307	0,685	0,275	0,164	0,111	0,407																
CN6	GGAACTGGTGGAGTTCTGGTTTCAGC	ACATCGGATGCCTGGCTTAATCAAGCTT	75 PE	Run1	8,880	3,152	2,705	0,615	0,175	0,110	0,066	0,439																
				Run2	10,538	2,848	2,472	0,049	0,010	0,010	*0.007163	0,032																
				Run3	5,976	3,467	3,041	0,088	0,042	0,024	0,018	0,046																
				Run4	7,978	3,990	2,671	0,781	0,270	0,163	0,106	0,510																
FSHD5	TGACATGGTGGAGTTCTGGTTTCAGC	GCCTAAGATGCCTGGCTTAATCAAGCTT	75 PE	Run1	4,812	3,904	2,423	0,792	0,280	0,170	0,110	0,512	5,268	1,634	0,470	3,431	0,937	0,705	0,232	1,849	0,779	0,575	0,204	1,076	66,475	67,066	64,675	63,163
				Run2	0,002	0,001	0,001	*0.000078	*0.000031	*0.000018	*0.000013	*0.000045																
				Run3	0,002	0,001	0,001	*0.000137	*0.000055	*0.000027	*0.000019	*0.000082																
				Run4	6,481	5,101	3,585	1,436	0,586	0,344	0,242	0,850																
FSHD6	GCGGACGGTGGAGTTCTGGTTTCAGC	ATTGGCGATGCCTGGCTTAATCAAGCTT	75 PE	Run1	4,038	2,851	2,350	0,706	0,186	0,186	*0.001386	0,518																
				Run2	3,283	2,411	2,170	0,575	0,200	0,154	0,046	0,375																
				Run3	3,871	2,895	3,606	0,786	0,395	0,313	0,071	0,583																
				Run4	4,807	3,293	2,936	1,012	0,297	0,297	*0.001508	0,713																

Supplementary Table 3.

Sample ID	4C bait reading primer sequences with barcodes	4C mate primer sequences with barcodes	Sequencing setting	Run	No. of reads	No. of reads passing filter	No. of trimmed paired-end reads	No. of reads after pooling technical replicates	No. of aligned reads	No. of reads with DpnII sites	% Cis reads
CN5	CTGATCATATTGCACCCTGGGGGAAGCT	CGTACGAGACCTACCTGCACCAACTG	75 PE	Run1	7.868	1.965	1.723	2.759	1.982	0.718	57.532
				Run2	5.748	1.177	1.036				
CN8	GATCTGATATTGCACCCTGGGGGAAGCT	GGCCACAGACCTACCTGCACCAACTG	75 PE	Run1	9.109	2.301	1.957	1.957	1.463	0.457	64.877
				Run2	6.274	0.266	10.135509				
FSHD5	ATTGGCATATTGCACCCTGGGGGAAGCT	TTTCACAGACCTACCTGCACCAACTG	75 PE	Run1	7.484	2.103	1.413	2.167	1.589	0.533	64.537
				Run2	5.243	1.135	0.754				
FSHD6	TGGTCAATATTGCACCCTGGGGGAAGCT	GGACGGAGACCTACCTGCACCAACTG	75 PE	Run1	5.942	1.656	1.496	3.499	2.713	0.734	64.494
				Run2	5.971	2.165	2.002				

Supplementary Table 4.

Name	Sequence
9	GGGGTCTTACCCCACTCCTA
11	CCCTGGATGGGCTTTTTAGT
13	TGTTGCCATCTTGAGTCCAA
15	GAGGATTTAAGGGGGAGTGG
17	AGAAGAAGGCCTGCACAGAG
19	CTCAGGGATGTGAGCTGTGA
21	TTACAGACGTGAGCCACCAC
23	CCAGTTTTGTGGCCTCATCT
26	GAGGCACCTTCAAAAGCAAG
31	TCTGGCCTCAAGTGATCCTC
33	ATGAGGTCAACAGGGGAGAA
*	CTCCCTCCTAACGTCCCTTC

Supplementary Table 5.

Name	Fw Sequence	Rv Sequence
Atrogin1 promoter	CACCGCTTCAAGTTTCCACC	CGTGGCTTTGTTTATGGGCT
Atrogin1 internal region	CAGTTGGTGCAGGTAGGTCT	AAAGACTGGGAGCAGGGTAC
DBE/DBE-T	AGGCCTCGACGCCCTGGGTC	TCAGCCGGACTGTGCACTGCGGC
PolII Ctrl +	TCCTCCTGTTTCATCCAAGC	CCTACTTTCTCCCCGCTTTT
PolII Ctrl -	GGTCAATGGCAGAAAAGGAAAT	CGCAGTTTGTGGGAATGATTC
GAPDH	CCCTTCATTGACCTCAACTACATG	TGGGATTTCCATTGATGACAAGC
Atrogin1	CTCATTCGGCAAAATCCAGT	GCGCTGAACATGAAAACAAA
DUX4-fl	CCCAGGTACCAGCAGACC	TCCAGGAGATGTAACTCTAATCCA
MyoD	GACGGCATGATGGACTACAG	GGGCGCCTCGTTGTAGTAG
MYHC2	GGACCAACTGAGTGAACTGAAA	TTGCCTCTTGATAACTGAGAC
Myogenin	TCAACCAGGAGGAGCGTGAC	TGTAGGGTCAGCCGTGAGCA
DUX4-ORF	GCGCAACCTCTCCTAGAAAC	AGCAGAGCCCGGTATTCTTC
RFLP2	CCCACATCAAGGAACTGGAG	TGTTGGCATCCAAGGTCATA
RPL13A	AACCTCCTCCTTTTCCAAGC	CAGTACCTGTTTAGCCACGA
FOXO3	CAGTGTGTTTCCGGTTCTGA	CATTTGGCAATGAGTGGAGA
BAC	CGATCCGGCGCGCCAATAGT	CGACCGATGCCCTTGAGAGCC

Supplementary Table 6.

References

1. Tawil, R., Van Der Maarel, S.M.: Facioscapulohumeral muscular dystrophy. Muscle & nerve **34**(1), 1-15 (2006). doi:10.1002/mus.20522

2. Lanzuolo, C.: Epigenetic alterations in muscular disorders. Comparative and functional genomics **2012**, 256892 (2012). doi:10.1155/2012/256892

3. Padberg, G.W., Lunt, P.W., Koch, M., Fardeau, M.: Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD **1**(4), 231-234 (1991).

4. Daxinger, L., Tapscott, S.J., van der Maarel, S.M.: Genetic and epigenetic contributors to FSHD. Current opinion in genetics & development **33**, 56-61 (2015). doi:10.1016/j.gde.2015.08.007

5. Sacconi, S., Salviati, L., Desnuelle, C.: Facioscapulohumeral muscular dystrophy. Biochimica et biophysica acta **1852**(4), 607-614 (2015). doi:10.1016/j.bbadis.2014.05.021

6. Tawil, R., Storvick, D., Feasby, T.E., Weiffenbach, B., Griggs, R.C.: Extreme variability of expression in monozygotic twins with FSH muscular dystrophy. Neurology **43**(2), 345-348 (1993).

7. Griggs, R.C., Tawil, R., McDermott, M., Forrester, J., Figlewicz, D., Weiffenbach, B.: Monozygotic twins with facioscapulohumeral dystrophy (FSHD): implications for

genotype/phenotype correlation. FSH-DY Group. Muscle & nerve. Supplement **2**, S50-55 (1995).

8. Neguembor, M.V., Gabellini, D.: In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy. Epigenomics **2**(2), 271-287 (2010). doi:10.2217/epi.10.8

9. Cabianca, D.S., Gabellini, D.: The cell biology of disease: FSHD: copy number variations on the theme of muscular dystrophy. The Journal of cell biology **191**(6), 1049-1060 (2010). doi:10.1083/jcb.201007028

10. van der Maarel, S.M., Tawil, R., Tapscott, S.J.: Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends in molecular medicine **17**(5), 252-258 (2011). doi:10.1016/j.molmed.2011.01.001

11. Wijmenga, C., Padberg, G.W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O.F., Milner, E.C., Weber, J.L., van Ommen, G.B., Sandkuyl, L.A., et al.: Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics **9**(4), 570-575 (1991).

12. Mathews, K.D., Mills, K.A., Bosch, E.P., Ionasescu, V.V., Wiles, K.R., Buetow, K.H., Murray, J.C.: Linkage localization of facioscapulohumeral muscular dystrophy (FSHD) in 4q35. American journal of human genetics **51**(2), 428-431 (1992).

13. Upadhyaya, M., Lunt, P., Sarfarazi, M., Broadhead, W., Farnham, J., Harper, P.S.: The mapping of chromosome 4q markers in relation to facioscapulohumeral muscular dystrophy

(FSHD). American journal of human genetics **51**(2), 404-410 (1992).

14. Lemmers, R.J., de Kievit, P., Sandkuijl, L., Padberg, G.W., van Ommen, G.J., Frants, R.R., van der Maarel, S.M.: Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nature genetics **32**(2), 235-236 (2002). doi:10.1038/ng999

15. Statland, J.M., Tawil, R.: Facioscapulohumeral Muscular Dystrophy. Continuum **22**(6, Muscle and Neuromuscular Junction Disorders), 1916-1931 (2016). doi:10.1212/CON.0000000000000399

16. van Deutekom, J.C., Bakker, E., Lemmers, R.J., van der Wielen, M.J., Bik, E., Hofker, M.H., Padberg, G.W., Frants, R.R.: Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. Human molecular genetics **5**(12), 1997-2003 (1996).

17. Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M., et al.: Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Human molecular genetics **3**(8), 1287-1295 (1994).

18. Winokur, S.T., Bengtsson, U., Feddersen, J., Mathews, K.D., Weiffenbach, B., Bailey, H., Markovich, R.P., Murray, J.C., Wasmuth, J.J., Altherr, M.R., et al.: The DNA rearrangement associated with facioscapulohumeral muscular dystrophy

involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology **2**(3), 225-234 (1994).

19. van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., Frants, R.R.: FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Human molecular genetics **2**(12), 2037-2042 (1993).

20. Lyle, R., Wright, T.J., Clark, L.N., Hewitt, J.E.: 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**(3), 389-397 (1995). doi:10.1006/geno.1995.1166

21. Deidda, G., Cacurri, S., Grisanti, P., Vigneti, E., Piazzo, N., Felicetti, L.: Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. European journal of human genetics : EJHG **3**(3), 155-167 (1995).

22. Bakker, E., Wijmenga, C., Vossen, R.H., Padberg, G.W., Hewitt, J., van der Wielen, M., Rasmussen, K., Frants, R.R.: The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a

homologue on 10qter. Muscle & nerve. Supplement(2), S39-44 (1995).

23. van Geel, M., Dickson, M.C., Beck, A.F., Bolland, D.J., Frants, R.R., van der Maarel, S.M., de Jong, P.J., Hewitt, J.E.: Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics **79**(2), 210-217 (2002). doi:10.1006/geno.2002.6690

24. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., Krom, Y.D., Klooster, R., Sun, Y., den Dunnen, J.T., Helmer, Q., Donlin-Smith, C.M., Padberg, G.W., van Engelen, B.G., de Greef, J.C., Aartsma-Rus, A.M., Frants, R.R., de Visser, M., Desnuelle, C., Sacconi, S., Filippova, G.N., Bakker, B., Bamshad, M.J., Tapscott, S.J., Miller, D.G., van der Maarel, S.M.: Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics **44**(12), 1370-1374 (2012). doi:10.1038/ng.2454

25. de Greef, J.C., Lemmers, R.J., van Engelen, B.G., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R., van der Maarel, S.M.: Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Human mutation **30**(10), 1449-1459 (2009). doi:10.1002/humu.21091

26. Cabianca, D.S., Casa, V., Gabellini, D.: A novel molecular mechanism in human genetic disease: a DNA repeat-derived

lncRNA. RNA biology **9**(10), 1211-1217 (2012). doi:10.4161/rna.21922

27. Casa, V., Gabellini, D.: A repetitive elements perspective in Polycomb epigenetics. Frontiers in genetics **3**, 199 (2012). doi:10.3389/fgene.2012.00199

28. Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., Gabellini, D.: A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell **149**(4), 819-831 (2012). doi:10.1016/j.cell.2012.03.035

29. Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., Belayew, A.: Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene **236**(1), 25-32 (1999).

30. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., Miller, D.G., Tapscott, S.J., Tawil, R., Frants, R.R., van der Maarel, S.M.: A unifying genetic model for facioscapulohumeral muscular dystrophy. Science **329**(5999), 1650-1653 (2010). doi:10.1126/science.1189044

31. Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Matteotti, C., van Acker, A.M., Leo, O., Figlewicz, D., Barro, M., Laoudj-Chenivesse, D., Belayew, A., Coppee, F., Chen, Y.W.: DUX4, a candidate gene of

facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proceedings of the National Academy of Sciences of the United States of America **104**(46), 18157-18162 (2007). doi:10.1073/pnas.0708659104

32. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., Miller, D.G.: Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics **6**(10), e1001181 (2010). doi:10.1371/journal.pgen.1001181

33. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., Tapscott, S.J.: DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell **22**(1), 38-51 (2012). doi:10.1016/j.devcel.2011.11.013

34. Bosnakovski, D., Xu, Z., Gang, E.J., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppee, F., Belayew, A., Perlingeiro, R.R., Kyba, M.: An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. The EMBO journal **27**(20), 2766-2779 (2008). doi:10.1038/emboj.2008.201

35. Block, G.J., Narayanan, D., Amell, A.M., Petek, L.M., Davidson, K.C., Bird, T.D., Tawil, R., Moon, R.T., Miller, D.G.: Wnt/beta-catenin signaling suppresses DUX4 expression and

prevents apoptosis of FSHD muscle cells. Human molecular genetics **22**(23), 4661-4672 (2013). doi:10.1093/hmg/ddt314
36. Tassin, A., Laoudj-Chenivesse, D., Vanderplanck, C., Barro, M., Charron, S., Ansseau, E., Chen, Y.W., Mercier, J., Coppee, F., Belayew, A.: DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? Journal of cellular and molecular medicine **17**(1), 76-89 (2013). doi:10.1111/j.1582-4934.2012.01647.x
37. Ferreboeuf, M., Mariot, V., Bessieres, B., Vasiljevic, A., Attie-Bitach, T., Collardeau, S., Morere, J., Roche, S., Magdinier, F., Robin-Ducellier, J., Rameau, P., Whalen, S., Desnuelle, C., Sacconi, S., Mouly, V., Butler-Browne, G., Dumonceaux, J.: DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Human molecular genetics **23**(1), 171-181 (2014). doi:10.1093/hmg/ddt409
38. Caruso, N., Herberth, B., Bartoli, M., Puppo, F., Dumonceaux, J., Zimmermann, A., Denadai, S., Lebosse, M., Roche, S., Geng, L., Magdinier, F., Attarian, S., Bernard, R., Maina, F., Levy, N., Helmbacher, F.: Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. PLoS genetics **9**(6), e1003550 (2013). doi:10.1371/journal.pgen.1003550
39. Stadler, G., Rahimov, F., King, O.D., Chen, J.C., Robin, J.D., Wagner, K.R., Shay, J.W., Emerson, C.P., Jr., Wright, W.E.: Telomere position effect regulates DUX4 in human

facioscapulohumeral muscular dystrophy. Nature structural & molecular biology **20**(6), 671-678 (2013). doi:10.1038/nsmb.2571

40. Mariot, V., Roche, S., Hourde, C., Portilho, D., Sacconi, S., Puppo, F., Duguez, S., Rameau, P., Caruso, N., Delezoide, A.L., Desnuelle, C., Bessieres, B., Collardeau, S., Feasson, L., Maisonobe, T., Magdinier, F., Helmbacher, F., Butler-Browne, G., Mouly, V., Dumonceaux, J.: Correlation between low FAT1 expression and early affected muscle in facioscapulohumeral muscular dystrophy. Annals of neurology **78**(3), 387-400 (2015). doi:10.1002/ana.24446

41. Puppo, F., Dionnet, E., Gaillard, M.C., Gaildrat, P., Castro, C., Vovan, C., Bertaux, K., Bernard, R., Attarian, S., Goto, K., Nishino, I., Hayashi, Y., Magdinier, F., Krahn, M., Helmbacher, F., Bartoli, M., Levy, N.: Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. Human mutation **36**(4), 443-453 (2015). doi:10.1002/humu.22760

42. Robin, J.D., Ludlow, A.T., Batten, K., Gaillard, M.C., Stadler, G., Magdinier, F., Wright, W.E., Shay, J.W.: SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy. Genome research **25**(12), 1781-1790 (2015). doi:10.1101/gr.190660.115

43. Bodega, B., Ramirez, G.D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli,

E., Ginelli, E.: 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 biology **7**, 41 (2009). doi:10.1186/1741-7007-7-41

44. Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., de Laat, W.: Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods **58**(3), 221-230 (2012). doi:10.1016/j.ymeth.2012.04.009

45. Denker, A., de Laat, W.: The second decade of 3C technologies: detailed insights into nuclear organization. Genes & development **30**(12), 1357-1382 (2016). doi:10.1101/gad.281964.116

46. Pellacani, D., Bilenky, M., Kannan, N., Heravi-Moussavi, A., Knapp, D.J., Gakkhar, S., Moksa, M., Carles, A., Moore, R., Mungall, A.J., Marra, M.A., Jones, S.J., Aparicio, S., Hirst, M., Eaves, C.J.: Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. Cell reports **17**(8), 2060-2074 (2016). doi:10.1016/j.celrep.2016.10.058

47. Jagannathan, S., Shadle, S.C., Resnick, R., Snider, L., Tawil, R.N., van der Maarel, S.M., Bradley, R.K., Tapscott, S.J.: Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Human molecular genetics (2016). doi:10.1093/hmg/ddw271

48. Bodine, S.C., Latres, E., Baumhueter, S., Lai, V.K., Nunez, L., Clarke, B.A., Poueymirou, W.T., Panaro, F.J., Na, E., Dharmarajan, K., Pan, Z.Q., Valenzuela, D.M., DeChiara, T.M., Stitt, T.N., Yancopoulos, G.D., Glass, D.J.: Identification of ubiquitin ligases required for skeletal muscle atrophy. Science **294**(5547), 1704-1708 (2001). doi:10.1126/science.1065874

49. Gomes, M.D., Lecker, S.H., Jagoe, R.T., Navon, A., Goldberg, A.L.: Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy. Proceedings of the National Academy of Sciences of the United States of America **98**(25), 14440-14445 (2001). doi:10.1073/pnas.251541198

50. Bodine, S.C., Baehr, L.M.: Skeletal muscle atrophy and the E3 ubiquitin ligases MuRF1 and MAFbx/atrogin-1. American journal of physiology. Endocrinology and metabolism **307**(6), E469-484 (2014). doi:10.1152/ajpendo.00204.2014

51. Calandra, P., Cascino, I., Lemmers, R.J., Galluzzi, G., Teveroni, E., Monforte, M., Tasca, G., Ricci, E., Moretti, F., van der Maarel, S.M., Deidda, G.: Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. Journal of medical genetics **53**(5), 348-355 (2016). doi:10.1136/jmedgenet-2015-103436

52. Bolger, A.M., Lohse, M., Usadel, B.: Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**(15), 2114-2120 (2014). doi:10.1093/bioinformatics/btu170

53. Langmead, B., Salzberg, S.L.: Fast gapped-read alignment with Bowtie 2. Nature methods **9**(4), 357-359 (2012). doi:10.1038/nmeth.1923

54. Raviram, R., Rocha, P.P., Muller, C.L., Miraldi, E.R., Badri, S., Fu, Y., Swanzey, E., Proudhon, C., Snetkova, V., Bonneau, R., Skok, J.A.: 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. PLoS computational biology **12**(3), e1004780 (2016). doi:10.1371/journal.pcbi.1004780

55. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A.: Circos: an information aesthetic for comparative genomics. Genome research **19**(9), 1639-1645 (2009). doi:10.1101/gr.092759.109

56. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., Manke, T.: deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research **42**(Web Server issue), W187-191 (2014). doi:10.1093/nar/gku365

57. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S.: Model-based analysis of ChIP-Seq (MACS). Genome biology **9**(9), R137 (2008). doi:10.1186/gb-2008-9-9-r137

58. Ernst, J., Kellis, M.: ChromHMM: automating chromatin-state discovery and characterization. Nature methods **9**(3), 215-216 (2012). doi:10.1038/nmeth.1906

59. Quinlan, A.R., Hall, I.M.: BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**(6), 841-842 (2010). doi:10.1093/bioinformatics/btq033

60. Yen, A., Kellis, M.: Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type. Nature communications **6**, 7973 (2015). doi:10.1038/ncomms8973

61. Heger, A., Webber, C., Goodson, M., Ponting, C.P., Lunter, G.: GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics **29**(16), 2046-2048 (2013). doi:10.1093/bioinformatics/btt343

62. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research **13**(11), 2498-2504 (2003). doi:10.1101/gr.1239303

63. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., Galon, J.: ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics **25**(8), 1091-1093 (2009). doi:10.1093/bioinformatics/btp101

64. Cortesi, A., Bodega, B.: Chromosome Conformation Capture in Primary Human Cells. Methods in molecular biology **1480**, 213-221 (2016). doi:10.1007/978-1-4939-6380-5_19

65. Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society **2**(2), 176-201 (1993). doi:10.1109/83.217222

Chapter 4 Summary, conclusions and future perspectives

Despite human genome composition could estimate for up to 66-69% of repetitive DNA [1], the functions of this fraction is still largely ignored. The development of highly sophisticated next generation sequencing (NGS) approaches allow to go deeper in repetitive elements functions understanding.

Indeed, over the past five years lots of reports showed that DNA repeats are a source of genetic variation [2] and evolution of fine epigenetic mechanisms of transcriptional regulation, suggesting that it's therefore becoming essential to look also at non-coding, repetitive portion of the genome, to understand both healthy and pathological phenotypes. In particular, it is recently reported intriguing evidences for a role for repeated genome in nuclear architecture, providing first suggestions of their potential role in higher-order chromatin packaging, linked to transcriptional regulation [3,4].

Nevertheless, there are emerging evidences of the peculiar role of repeats in regulating the epigenome. For instance, DNA repeats have been involved in chromosome structural organization, gene regulation, genome integrity, and evolution [5-10].

FSHD is an important human disease caused by alteration of repetitive sequences. The mechanisms through which repeat contraction at the D4Z4 locus on chromosome 4 leads to muscular dystrophy in specific muscles are still poorly understood. Despite it is demonstrated that FSHD arises as alteration of an epigenetic mechanism, none of the genes

reported accounts for all aspects of FSHD, and additional players are still actively sought. Indeed, it is increasingly evident that D4Z4 deletion or DUX4 overexpression are not sufficient *per se* to cause FSHD [11-13], suggesting that factors such as modifier genes, gender and environment may affect disease onset and severity.

All the above observations strongly claim for the development of novel approaches to achieve a global comprehension of the molecular determinants at the basis of FSHD manifestation, considering that until the cascade of molecular events following D4Z4 contraction will not be clear, the development of an efficacious therapy for FSHD patients will be delayed. Our goal was to dissect the epigenetic role of DNA repeats in myogenesis and FSHD muscular dystrophy, providing new tools for preclinical studies.

Our work hypothesized that the D4Z4 locus might play a critical role in establishing chromatin organization within the cell and that contraction of the D4Z4 locus might change the chromatin conformation and affect relevant functional characteristics of muscle cells.

We therefore assessed the role of D4Z4 tandem repeat located at 4q35 region in the coordination of transcriptional programs, through tridimensional interactions between non-contiguous elements in the genome of human primary muscle cells. The contraction of the array and the concomitant hypomethylation of the FSHD locus led to the deregulation of genome-wide

chromatin landscapes as well as to an altered 4q-D4Z4-mediated genomic interactome resulting in the induction of an atrophic signature in FSHD primary skeletal muscle cells. Moreover this peculiar mechanism of transcriptional regulation could explain the extreme phenotypic variability that patients exert, even belonging to the same family. We produced new insight into the molecular basis of FSHD.

A modified version of the 4C technology allowed us to identifiy the genomic regions that interact with the D4Z4 locus on chromosome 4 (while excluding those that interact with the D4Z4 on chromosome 10) in FSHD patients. Moreover the production of different ChIP-seq datasets enabled us decipher FSHD-associated chromatin landscapes using chromatin segmentation analysis tools such as ChromHMM and Chrom-diff. These analyses uncovered active enhancers as major variable states between controls and FSHD. The intersection of ChIP-seq datasets we generated with transcriptome datasets available for FSHD allowed us to discriminate DUX4-dependent and independent genes associated with active enhancers inside the deregulated interactions uncovered by 4qsslp4C in FSHD. Interestingly the first group (DUX4-dependent) showed processes associated to development, differentiation and RNA metabolism processes as described in the literature, instead the second group (DUX4-independent) unveiled novel possible target genes that included those related to the atrophic phenotype, characteristics of the pathology.

One of these is the atrophic marker Atrogin1. We provided evidences for the functional relevance of the trans interaction between Atrogin1 and 4q-D4Z4 proximal region. We indeed showed that Atrogin1 interaction was lost in FSHD leading to its transcriptional deregulation during muscle cell differentiation.

In this frame it is very interestingly to note that a “wild type” 4q-D4Z4 array containing at least 15 D4Z4 repeats reintroduced in FSHD cells led to an impaired upregulation of Atrogin1.

This suggests that 4q-D4Z4 array, after acquiring its epigenetic features is able to modify the genome structure, at least for the Atrogin1 locus, and somehow revert the FSHD phenotype to a “control-like” situation.

Finally, examining in parallel a 4C-seq analysis using the Atrogin-1 gene as a viewpoint, we showed that this locus displayed altered interacting partners in FSHD cells respect to control cells. This provides greater support for a 4q-D4Z4-dependent re-organization of chromatin higher order structure and in FSHD.

Surprisingly enough, the Atrogin1 interacting genes were indeed related to the atrophy and related pathways in which the protein is involved. This suggests at least in part a link between alteration of the chromatin structure and deregulated myogenic functions observed in the pathology.

We thus provide significant advance in understanding the mechanisms linking 3D genome organization and DNA repeats to dynamic programming of chromatin states in development

and disease. In particular, we provide a deeper understanding of the role of DNA repetitive elements in human myogenesis and disease manifestation.

Moreover, our work strongly outlines the relevance of epigenetics and repetitive elements in regulating genetic information and susceptibility to diseases.

Indeed, to date, most FSHD treatments involve attempts to physically improve functional impairment, with surgery used to alleviate both scapular fixation and ‘foot drop’ in patients [14].

However it is possible to perform surgery for winged scapula, although this technique is still controversial [15].

Also the role of exercise in maintaining or improving muscular force and/ or functional ability is still in doubt, because of the lack of controlled studies and because vigorous exercise could theoretically worsen muscular weakness inducing rhabdomyolysis [15]. However recent studies have shown that both strength training and aerobic exercise in FSHD have at least a short-term beneficial effect [16,17]. Moreover also neuromuscular electrical stimulation strength training seems to be effective [18].

Future randomized controlled studies are necessary to better define and broaden these recommendations.

An increasing knowledge of at least some of the molecular aspects of the pathology however led to more directed FSHD therapies.

A number of cell therapy-based and pharmacological strategies

have been tested in order to slow down or stop disease progression and treat secondary manifestations [15].

Transplantation of cultured myoblasts by intramuscular injection has been considered for a number of dystrophies. Because FSHD is characterized by selective muscle involvement, it is suggested that myoblasts derived from histologically unaffected FSHD muscles could be used for autologous cell therapy in FSHD, eliminating the need of immunosuppression [19]. However therapeutic trial, exploring the feasibility and efficacy of this approach is still ongoing.

Mesangioblast, another type of myogenic mesodermal stem cell, has been also considered for future autologous therapeutic trials. Mesangioblasts were shown to improve muscle morphology and function when injected intra-arterially in animal models of dystrophy. They can be delivered to whole body muscles through the circulation, while myoblasts have to be injected locally, and as such may represent a more effective alternative for autologous cell therapy.

The correct management of symptoms and of secondary manifestations is essential to slow down disease progression in order to conserve as long as possible functional autonomy and adequate quality of life.

In this frame different therapeutical approaches has been considered, unfortunately without significant effectiveness. Among these we could mention corticosteroids, since around 40% of FSHD muscle biopsies show T cell inflammatory

infiltrates [20-22]; β2 agonists because of their anabolic effect [23,24]; methionine and folic acid, which are supposed to increase methylation, since D4Z4 array shows hypomethylation in FSHD patients [25]; an antibody (MYO-029) against myostatin, that negatively controls satellite cell proliferation and differentiation [26,27]; stimulation of the oxidative stress response pathway, since oxidative stress plays an important role in determining DUX4 toxicity [28].

Other disease-specific therapies targeting FSHD molecular signatures have been developed. Indeed, the alterations in chromatin structure at the FSHD locus cause a gain of function mutation, and a logical therapeutic approach is thus to employ antisense strategies against specifically activated target genes. [29]. Most of them are focused on developing gene therapy.

Among these suppression of DUX4 gene expression has been tried by using siRNA or RNA-like antisense oligonucleotides (AONs) interfering with DUX4 mRNA processing and stability, or alter the splicing of the DUX4 transcript to reduce the amount of DUX4-fl, with the reported decrease in DUX4 expression found to normalize several FSHD deregulated genes [15,29]. A major issue in strategies targeting the DUX4 gene is the reported very low abundance of DUX4 expressing cells in culture (1/1,000 myoblasts) [30], however the DUX4 protein expressed from a single nucleus could spread in the cytoplasm of a fiber and reach neighboring nuclei, deregulating their gene expression. Thus, targeting DUX4 could block the spreading of

the deregulation cascade.

Moreover RNA interference strategies were designed to target the other FSHD candidate gene FRG1 [31,32], by using FRG1 - specific miRNAs and shRNAs in the myopathy mouse model overexpressing the FRG1 protein. This strategy resulted in a significant improvement of muscle histology, with increased muscle mass, reduced fat deposition and fibrosis, a decline in myofibre degeneration, and an overall improvement of muscle function and strength.

In addition to these current therapeutical strategies, our work adds a new layer a potential therapeutic intervention by highlighting unpredictable candidates (i.e. promoters, genes/non-coding RNAs, etc) derived among D4Z4 interacting regions. Our integrated multi-omics approach enabled us to uncover potential new candidates which could contribute to FSHD pathogenesis and provide novel biomarkers or therapeutic targets such as Atrogin1.

However the practice of medical genetics requires a clear, definite evaluation of the significance of genetic variations in patients, to provide prognostic information and genetic counseling. Here, we propose to tackle this issue from a novel and unbiased perspective.

Our work also raises the possibility to perform a gene-like therapy not acting on the recovery or direct regulation of a “diseased-gene” function but instead using a repetitive element portion of the genome with the ultimate goal to act on and

regulate the genomic structure mediated by this repeat.

This is a very interesting perspective that has to be further investigated in order to open a new era of therapeutic strategies targeting deregulated epigenomes in a third layer of complexity: 3D nuclear architecture.

References

1. de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., Pollock, D.D.: Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics **7**(12), e1002384 (2011). doi:10.1371/journal.pgen.1002384

2. Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., Moran, J.V.: LINE-1 retrotransposition activity in human genomes. Cell **141**(7), 1159-1170 (2010). doi:10.1016/j.cell.2010.05.021

3. Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., Noro, Y., Wong, C.H., de Hoon, M., Andersson, R., Sandelin, A., Suzuki, H., Wei, C.L., Koseki, H., Consortium, F., Hasegawa, Y., Forrest, A.R., Carninci, P.: Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nature genetics **46**(6), 558-566 (2014). doi:10.1038/ng.2965

4. Hall, L.L., Carone, D.M., Gomez, A.V., Kolpa, H.J., Byron, M., Mehta, N., Fackelmayer, F.O., Lawrence, J.B.: Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. Cell **156**(5), 907-919 (2014). doi:10.1016/j.cell.2014.01.042

5. Casa, V., Gabellini, D.: A repetitive elements perspective in Polycomb epigenetics. Frontiers in genetics **3**, 199 (2012). doi:10.3389/fgene.2012.00199

6. Kidwell, M.G., Lisch, D.R.: Transposable elements and host genome evolution. Trends in ecology & evolution **15**(3), 95-99 (2000).

7. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C.,

Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G.,

Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing, C.: Initial sequencing and analysis of the human genome. Nature **409**(6822), 860-921 (2001). doi:10.1038/35057062

8. Feschotte, C.: Transposable elements and the evolution of regulatory networks. Nature reviews. Genetics **9**(5), 397-405 (2008). doi:10.1038/nrg2337

9. Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., Rivera, M.N., Bardeesy, N., Maheswaran, S., Haber, D.A.: Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science **331**(6017), 593-596 (2011). doi:10.1126/science.1200801

10. Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., Verma, I.M.: BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature **477**(7363), 179-184 (2011). doi:10.1038/nature10371

11. Jones, T.I., Chen, J.C., Rahimov, F., Homma, S., Arashiro, P., Beermann, M.L., King, O.D., Miller, J.B., Kunkel, L.M., Emerson, C.P., Jr., Wagner, K.R., Jones, P.L.:

Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Human molecular genetics **21**(20), 4419-4430 (2012). doi:10.1093/hmg/dds284

12. Scionti, I., Greco, F., Ricci, G., Govi, M., Arashiro, P., Vercelli, L., Berardinelli, A., Angelini, C., Antonini, G., Cao, M., Di Muzio, A., Moggio, M., Morandi, L., Ricci, E., Rodolico, C., Ruggiero, L., Santoro, L., Siciliano, G., Tomelleri, G., Trevisan, C.P., Galluzzi, G., Wright, W., Zatz, M., Tupler, R.: Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. American journal of human genetics **90**(4), 628-635 (2012). doi:10.1016/j.ajhg.2012.02.019

13. Ricci, G., Scionti, I., Sera, F., Govi, M., D'Amico, R., Frambolli, I., Mele, F., Filosto, M., Vercelli, L., Ruggiero, L., Berardinelli, A., Angelini, C., Antonini, G., Bucci, E., Cao, M., Daolio, J., Di Muzio, A., Di Leo, R., Galluzzi, G., Iannaccone, E., Maggi, L., Maruotti, V., Moggio, M., Mongini, T., Morandi, L., Nikolic, A., Pastorello, E., Ricci, E., Rodolico, C., Santoro, L., Servida, M., Siciliano, G., Tomelleri, G., Tupler, R.: Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. Brain : a journal of neurology **136**(Pt 11), 3408-3417 (2013). doi:10.1093/brain/awt226

14. Tawil, R.: Facioscapulohumeral muscular dystrophy. Neurotherapeutics : the journal of the American Society for

Experimental NeuroTherapeutics **5**(4), 601-606 (2008). doi:10.1016/j.nurt.2008.07.005

15. Sacconi, S., Salviati, L., Desnuelle, C.: Facioscapulohumeral muscular dystrophy. Biochimica et biophysica acta **1852**(4), 607-614 (2015). doi:10.1016/j.bbadis.2014.05.021

16. van der Kooi, E.L., Lindeman, E., Riphagen, I.: Strength training and aerobic exercise training for muscle disease. The Cochrane database of systematic reviews(1), CD003907 (2005). doi:10.1002/14651858.CD003907.pub2

17. van der Kooi, E.L., Vogels, O.J., van Asseldonk, R.J., Lindeman, E., Hendriks, J.C., Wohlgemuth, M., van der Maarel, S.M., Padberg, G.W.: Strength training and albuterol in facioscapulohumeral muscular dystrophy. Neurology **63**(4), 702-708 (2004).

18. Colson, S.S., Benchortane, M., Tanant, V., Faghan, J.P., Fournier-Mehouas, M., Benaim, C., Desnuelle, C., Sacconi, S.: Neuromuscular electrical stimulation training: a safe and effective treatment for facioscapulohumeral muscular dystrophy patients. Archives of physical medicine and rehabilitation **91**(5), 697-702 (2010). doi:10.1016/j.apmr.2010.01.019

19. Vilquin, J.T., Marolleau, J.P., Sacconi, S., Garcin, I., Lacassagne, M.N., Robert, I., Ternaux, B., Bouazza, B., Larghero, J., Desnuelle, C.: Normal growth and regenerating ability of myoblasts from unaffected muscles of

facioscapulohumeral muscular dystrophy patients. Gene therapy **12**(22), 1651-1662 (2005). doi:10.1038/sj.gt.3302565

20. Munsat, T.L., Piper, D., Cancilla, P., Mednick, J.: Inflammatory myopathy with facioscapulohumeral distribution. Neurology **22**(4), 335-347 (1972).

21. Wulff, J.D., Lin, J.T., Kepes, J.J.: Inflammatory facioscapulohumeral muscular dystrophy and Coats syndrome. Annals of neurology **12**(4), 398-401 (1982). doi:10.1002/ana.410120415

22. Tawil, R., McDermott, M.P., Pandya, S., King, W., Kissel, J., Mendell, J.R., Griggs, R.C.: A pilot trial of prednisone in facioscapulohumeral muscular dystrophy. FSH-DY Group. Neurology **48**(1), 46-49 (1997).

23. Benson, D.W., Foley-Nelson, T., Chance, W.T., Zhang, F.S., James, J.H., Fischer, J.E.: Decreased myofibrillar protein breakdown following treatment with clenbuterol. The Journal of surgical research **50**(1), 1-5 (1991).

24. Maltin, C.A., Hay, S.M., McMillan, D.N., Delday, M.I.: Tissue specific responses to clenbuterol; temporal changes in protein metabolism of striated muscle and visceral tissues from rats. Growth regulation **2**(4), 161-166 (1992).

25. van der Kooi, E.L., de Greef, J.C., Wohlgemuth, M., Frants, R.R., van Asseldonk, R.J., Blom, H.J., van Engelen, B.G., van der Maarel, S.M., Padberg, G.W.: No effect of folic acid and methionine supplementation on D4Z4 methylation in patients with facioscapulohumeral muscular dystrophy. Neuromuscular

disorders : NMD **16**(11), 766-769 (2006). doi:10.1016/j.nmd.2006.08.005

26. Patel, K., Amthor, H.: The function of Myostatin and strategies of Myostatin blockade-new hope for therapies aimed at promoting growth of skeletal muscle. Neuromuscular disorders : NMD **15**(2), 117-126 (2005). doi:10.1016/j.nmd.2004.10.018

27. Wagner, K.R., Fleckenstein, J.L., Amato, A.A., Barohn, R.J., Bushby, K., Escolar, D.M., Flanigan, K.M., Pestronk, A., Tawil, R., Wolfe, G.I., Eagle, M., Florence, J.M., King, W.M., Pandya, S., Straub, V., Juneau, P., Meyers, K., Csimma, C., Araujo, T., Allen, R., Parsons, S.A., Wozney, J.M., Lavallie, E.R., Mendell, J.R.: A phase I/IItrial of MYO-029 in adult subjects with muscular dystrophy. Annals of neurology **63**(5), 561-571 (2008). doi:10.1002/ana.21338

28. Bosnakovski, D., Choi, S.H., Strasser, J.M., Toso, E.A., Walters, M.A., Kyba, M.: High-throughput screening identifies inhibitors of DUX4-induced myoblast toxicity. Skeletal muscle **4**(1), 4 (2014). doi:10.1186/2044-5040-4-4

29. Richards, M., Coppee, F., Thomas, N., Belayew, A., Upadhyaya, M.: Facioscapulohumeral muscular dystrophy (FSHD): an enigma unravelled? Human genetics **131**(3), 325-340 (2012). doi:10.1007/s00439-011-1100-z

30. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., Miller, D.G.: Facioscapulohumeral

dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics **6**(10), e1001181 (2010). doi:10.1371/journal.pgen.1001181

31. Wallace, L.M., Garwick-Coppens, S.E., Tupler, R., Harper, S.Q.: RNA interference improves myopathic phenotypes in mice over-expressing FSHD region gene 1 (FRG1). Molecular therapy : the journal of the American Society of Gene Therapy **19**(11), 2048-2054 (2011). doi:10.1038/mt.2011.118

32. Bortolanza, S., Nonis, A., Sanvito, F., Maciotta, S., Sitia, G., Wei, J., Torrente, Y., Di Serio, C., Chamberlain, J.R., Gabellini, D.: AAV6-mediated systemic shRNA delivery reverses disease in a mouse model of facioscapulohumeral muscular dystrophy. Molecular therapy : the journal of the American Society of Gene Therapy **19**(11), 2055-2064 (2011). doi:10.1038/mt.2011.153

Publications

Cesarini E, Mozzetta C, Marullo F, Gregoretti F, Gargiulo A, Columbaro M, **Cortesi A**, Antonelli L, Di Pelino S, Squarzoni S, Palacios D, Zippo A, Bodega B, Oliva G, Lanzuolo C.
Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes.
(2015) *J Cell Biol.* 9;211(3):533-51

Cortesi A, Bodega B.
Chromosome conformation capture in human primary cells. (2016) *Methods Mol Biol*.1480:213-21.

Cortesi A*, Pesant M*, Sinha S, Antonelli L, Gregoretti F, Oliva G, Soldà G. Bodega B.
Chromatin landscape of D4Z4 repeat interactome reveals a muscle atrophy signature in Facioscapulohumeral Dystrophy.
*Equal Contribution
Submitted