INVITED REVIEW

Loss of epigenetic silencing of the DUX4 transcription factor gene in facioscapulohumeral muscular dystrophy

Jane E. Hewitt*

School of Life Sciences, Queen’s Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK

*To whom correspondence should be addressed. Tel: +44 1158230364; Fax: +44 1158230338; Email: jane.hewitt@nottingham.ac.uk

Abstract

Current genetic and molecular evidence best supports an epigenetic mechanism for facioscapulohumeral muscular dystrophy (FSHD), whereby de-repression of the D4Z4 macrosatellite array leads to aberrant expression of the DUX4 transcription factor in skeletal muscle. This de-repression is triggered by either array contraction or (more rarely) by mutation of the SMCHD1 (structural maintenance of chromosomes flexible hinge domain containing 1) gene. Activation of DUX4 targets, including germline genes and several mammalian retrotransposons, then drives pathogenesis. A direct role for DUX4 mRNA in suppression of nonsense-mediated decay pathways has recently been demonstrated and may also contribute to muscle pathology. Loss of D4Z4 repression in FSHD is observed as hypomethylation of the array accompanied by loss of repressive chromatin marks. The molecular mechanisms of D4Z4 repression are poorly understood, but recent data have identified an Argonaute (AGO)-dependent siRNA pathway. Targeting this pathway by exogenous siRNAs could be a therapeutic strategy for FSHD.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an inherited muscle disorder with a prevalence of ∼1:20 000 (1). Symptoms typically present during the second decade (although a small number of cases show infantile onset) with weakness of the face and/or shoulder girdle muscles. The disease then usually progresses to include the humeral and proximal lower extremity muscles, and the foot dorsiflexors; 20% of individuals require a wheelchair later in life (1). Compared with females, males have an earlier disease onset and tend to be more severely affected (2). Respiratory and cardiac muscle involvement is rare, and life expectancy is not significantly reduced. Non-muscular symptoms associated with FSHD include retinal vascular disease and high-frequency hearing loss (1). Genetic and molecular evidence best supports an epigenetic disease mechanism that results in inappropriate expression of the transcription factor DUX4.

FSHD Is Caused by De-repression of the D4Z4 Macrosatellite

FSHD1: contraction-dependent FSHD

The most common genetic form of FSHD (FSHD1, accounting for ∼95% of cases) shows autosomal dominant inheritance (although there is a high frequency of sporadic cases owing to de novo mutations) and is caused by a contraction of the D4Z4 macrosatellite repeat array on chromosome 4q35 (3,4). Unaffected individuals carry 11–100 D4Z4 repeat units whereas nearly all FSHD1 cases have a shortened array with 1–10 units (5), at least one repeat being required for pathology (6). The array contraction occurs via a mitotic rearrangement in early embryonic development (7). There is an inverse relationship between the residual D4Z4 array length and clinical phenotype (1); for a given array length, the disease severity can vary considerably even within the same family. Up to 10–20% of mutation carriers are asymptomatic or develop the disease late in life (8–10).

Molecular diagnosis of FSHD1 typically uses Southern blot analysis using EcoRI digestion and the p13E-11 probe (1), the sizes of the fragments indicating the number of 3.3-kb repeat units within the arrays. Chromosome 4q alleles can usually be differentiated from a highly similar array on 10q by digestion with chromosome-specific restriction enzymes (1). In the majority of cases, this is sufficient to identify a contracted, FSHD-associated 4q allele. Some patients have more complex repeat arrays that are not readily differentiated as being of 4q or 10q origin or have deletions that also remove the p13E-11 region, and these cases require a more detailed molecular analysis to confirm a diagnosis of FSHD1 (11,12).
FSHD2: contraction-independent FSHD

Approximately 5% of patients have no D4Z4 contraction but are clinically indistinguishable from FSHD1 (13). The apparent mode of inheritance is variable; while the majority of cases are apparently sporadic, dominant and recessive patterns have also been described (13). It is now clear that FSHD1 and FSHD2 show similar epigenetic changes at D4Z4 and share a common disease pathway (3). Exome sequencing of FSHD2 individuals recently identified causative mutations in the SMCHD1 (structural maintenance of chromosomes flexible hinge domain containing 1) gene (14).

Diagnosis of FSHD2 can be problematic, with no diagnostic short 4q D4Z4 array it can be misdiagnosed as a form of limb-girdle muscular dystrophy (15,16). Sequencing of SMCHD1 has been used to confirm diagnosis in some cases (17,18), but this can be challenging as it is a large gene with 48 exons (14) and most mutations appear to be private (19). Genetic heterogeneity also seems likely in FSHD2, although no locus apart from SMCHD1 has yet been identified.

FSHD Is Caused by a Failure of Epigenetic Repression of DUX4

Each D4Z4 repeat contains a 1.2-kb open-reading frame encoding the DUX4 homeodomain protein (20–22). DUX4 genes are defined by their two closely spaced homeobox sequences and found only in mammals (23,24). Although some proposed disease mechanisms favour cis-acting epigenetic effects, whereby D4Z4 deletions result in aberrant expression of genes proximal to the array (25), a direct role of DUX4 in FSHD is the model most consistently supported by genetic and molecular data (5,26). This model does retain an important epigenetic component as changes in D4Z4 methylation and chromatin modifications in FSHD allow the aberrant (and pathological) expression of DUX4 in skeletal muscle.

D4Z4 contractions are only pathogenic in particular chromosomal contexts (27,28). Of the two distinct subtelomere variants (4qA and 4qB), FSHD is associated only with D4Z4 contractions on 4qA chromosomes (29,30). On this 4qA variant, the D4Z4 array terminates in a block of 68-bp β-satellite DNA that is absent in the 4qB subtelomere (29). A highly similar macrosatellite array on chromosome 10q26, which arose by an ancestral duplication of the 4qter region onto 10qter (29), is not associated with FSHD (27,28).

Haplotype analysis of 4q uncovered the reason for the specificity of this association (28). An SSLP proximal to D4Z4 and the A/B alleles are used to define these haplotypes, for example ‘4qA161’ refers to a 4q allele with a 161-bp SSLP and an A-type telomere. The association of only these “permissive” haplotypes with FSHD is due to linkage disequilibrium between proximal polymorphisms and variants within a polyadenylation signal (PAS) in the 68-bp β-satellite DNA (31,32). On the permissive 4qA haplotypes, this PAS sequence is ATTAAA (a non-canonical signal that is common in humans). On chromosome 10q, the sequence is ATCAAATA or ATTTAAA, neither of which are functional, whereas non-permissive chromosomes of the ‘4qB’-type lack the PAS entirely owing to the absence of the β-satellite DNA (33). Although FSHD2 is not associated with D4Z4 contractions, all affected individuals carry at least one 4q allele with a permissive haplotype (14). The length of this permissive 4q allele tends to be shorter than the average in controls, although longer than the FSHD1 threshold of 10 repeat units (34).

Although FSHD1 had been invariably been associated only with the 4q locus, Lemmers et al. described a pedigree in which the disease-causative allele is a short D4Z4 array on chromosome 10qter (33). Owing to a de novo translocation event, the disease-associated 10q array ends distally in 4qter sequence complete with the ATTAAA sequence. This combination of a shortened 10q D4Z4 array in combination with a functional PAS is therefore sufficient for the development of FSHD and argues strongly against a primary mechanism that acts via a cis effect of D4Z4 contraction on more proximal 4q35 genes.

The shared phenotype for both types of FSHD is explained by aberrant DUX4 expression owing to epigenetic changes that are triggered either by an array contraction in FSHD1 or by a defect in D4Z4 repression in FSHD2 (Fig. 1). Miss-expression of the DUX4 transcription factor is then pathogenic (33,35), although...
there may also be a contribution to the disease directly from the mRNA (36). This model explains the requirement for at least one D4Z4 unit and the permissive 4q haplotypes. Non-penetrant carriers of FSHD1 mutations could be explained by maintenance of sufficient array repression in these individuals for DUX4 levels to remain below a disease threshold. Cases of apparent FSHD without a permissive haplotype need to be evaluated in detail to exclude more complex rearrangements in 4qter (12,37) or mutations in genes that phenocopy FSHD (15,16).

Epigenetic changes in FSHD release D4Z4 repression

D4Z4 arrays, which are normally heavily methylated in somatic tissues, show hypermethylation in both types of FSHD (34). In FSHD1, this is confined to the contracted array, whereas FSHD2 patients show hypomethylation on all four 4q and 10q alleles (15,25,38–40). These changes are seen in a wide range of somatic cells including peripheral blood mononuclear cells, fibroblasts, muscle and saliva (14,34,41,42). The best studies of methylation in FSHD have combined Southern blotting with methylation-sensitive restriction enzymes as this allows differential analysis of the 4q and 10q loci (34,38–40), but this approach limits the number of CpGs that can be analysed. PCR-based techniques such as bisulphite sequencing, while allowing fine-scale analyses, are compromised by cross-hybridization of primers to the 10q duplicon and a large number of poorly mapped D4Z4-related sequences in some of their amplicons (41,45), a re-

D4Z4 methylation levels contribute to variation in FSHD severity

The level of DNA methylation at D4Z4 approximately correlates inversely with disease severity and penetrance (39,45). Furthermore, independent of FSHD status, shorter arrays on either 4q or 10q have lower methylation levels (34). This was studied in detail using a methylation-sensitive FseI site in the proximal repeat unit (known to be representative of methylation across the entire array) to investigate the correlation between array length and methylation in a large study containing 254 controls, 186 FSHD1 individuals and 74 SMCHD1 mutation carriers (19). Note that FSHD2 shows digenic inheritance and SMCHD1 mutation carriers only develop FSHD2 if they also carry a relatively short permissive 4qA allele (14).

Analysis of this FseI site produces a measure of the average methylation across all four 4q and 10q loci. In controls and the subset of FSHD1 individuals with arrays of 1–6 repeat units, there was strong correlation between FseI methylation and the cumulative repeat array sizes, whereas in SMCHD1 mutation carriers, the methylation level was reduced by ∼50% but still showed a correlation with overall D4Z4 copy number. In individuals with arrays of 7–10 units, FSHD1 individuals (but not unaffected family members carrying the same short arrays) were less methylated than would be predicted from the total D4Z4 copy number (19). This reflects the relationship seen between array length and clinical severity, where FSHD1 mutation carriers with alleles of 7–10 units show poor correlation and a higher frequency of non-penetrant mutation carriers (8,10). The methylation data indicate that, for these longer disease alleles, intrinsic differences between individuals in the efficacy of epigenetic D4Z4 silencing contribute to clinical variation and non-penetrance (19).

The Complex Landscape of DUX4 Transcription

DUX4 protein-coding transcripts are produced in FSHD muscle cells

Although early attempts to identify DUX4 transcripts failed, most likely for technical reasons (20,43,59–62), there is now substantial evidence for a complex pattern of endogenous transcription of this locus (31,32,35,63). Many differentially spliced and
un-spliced transcripts, microRNAs and antisense transcripts have been identified (32,64,65). The DUX4 coding region is contained within a single exon, likely due to its origin via retrotransposition of a processed mRNA (24,56). Unusually, the gene subsequently gained introns in the 3′ untranslated region (31,32,35). Most documented cases of intron gain in retrogenes have occurred in the 5′ UTR, probably because introns distal to the stop codon usually destabilize mRNA transcripts via nonsense-mediated decay (NMD) (66).

Full-length DUX4 transcripts (DUX4f) encode the complete DUX4 protein. The introns in the 3′ UTR are alternatively spliced to generate two DUX4f transcripts (31,32). A further alternative transcript uses a splice donor site within the coding region. If translated, this splice form (DUX4s) would encode a protein retaining the homeodomains but lacking the C-terminal 264 amino acid transactivation domain.

DUX4s transcripts are produced in both control and FSHD muscle. DUX4f mRNA and DUX4 protein expression is not seen in most somatic tissues but is consistently observed in a low-proportion (~1 out of every 1000) FSHD muscle cells (33,35). Although DUX4f transcripts have also been reported in control muscle, these are much less frequently observed and are present at much lower levels (67), suggesting that there is a threshold above which DUX4f levels cause FSHD.

A switch in splicing from DUX4s to DUX4f therefore accompanies the change in D4Z4 chromatin structure in FSHD. The full-length transcript is stabilized by the functional PAS on permissive 4q haplotypes leading to production of the DUX4 protein and triggering development of disease pathology (Fig. 1). Forced expression of DUX4 in cells induces apoptosis and abnormal myotube formation (63,68,69); similar processes may account for the muscle loss seen in FSHD.

Robust expression of both DUX4f mRNA and protein is seen in normal testis (32,35) where it may have a role in germ cell differentiation, in line with its endogenous DUX4 expression in testis, that are not normally expressed in muscle (73,74). DUX4 also binds and activates transcription of several mammalian retrotransposons, including MaLRs, ERVL and ERVK elements, which may also contribute to the development of FSHD (73,74).

The DUX4f mRNA may also play a direct role in FSHD

This location of introns in the 3′ UTR of DUX4 suggests that transcript levels may be regulated by a NMD; this is supported by recent data showing that NMD is an endogenous suppressor of DUX4 RNA levels in muscle cells (36). RNA-seq analysis of myoblast cells overexpressing DUX4 transcripts showed an increased abundance of RNAs with premature termination codons owing to defective NMD. The mechanism of NMD suppression appears to involve increased degradation of UPF1 in response to transcription of the DUX4 3′ UTR (36). This crosstalk also provides a potential positive autoregulatory loop, whereby increased transcription of DUX4 transcript reduces its degradation by NMD. Such local repression of NMD could lead to stabilization of transcript and protein levels, facilitating the observed spreading of DUX4 expression between adjacent nuclei in a muscle fibre (36,67,75). The concomitant increased abundance of RNAs containing premature termination codons in muscle cells may also contribute to disease pathology (36).

Transcriptional silencing of DUX4 by an argonaute-dependent siRNA pathway

A cluster of small RNAs originating upstream of the DUX4 mRNA (32) appears to function in D4Z4 silencing (76). Transcriptional silencing by siRNA is well established as a mechanism for gene regulation (77). Exogenous targeting of this upstream region of D4Z4 by siRNA led to increased H3K9me2 methylation and silencing of DUX4 expression in FSHD cells; this was shown to act via an Argonaute (AGO)-dependent pathway (76). Knockdown of DICER or AGO2 in non-FSHD muscle cells (each carrying permissive D4Z4 arrays with 13 repeat units) resulted in the expression of DUX4, consistent with a role in endogenous silencing of D4Z4. Interestingly, reduction of DICER or AGO2 in a cell line carrying a permissive haplotype with 74 D4Z4 units did not result in release of D4Z4 repression or DUX4 expression (76). This is reminiscent of the limited efficacy of SMCHD1 haploinsufficiency, where longer arrays are less susceptible to loss of repression (14).

Genome-wide ChIP-on-chip studies and RNA sequencing of chromatin complexes in somatic cells also identified D4Z4 as a potential target for AGO-mediated silencing (78,79). AGO proteins also modulate RNA polymerase II elongation rate, thereby affecting alternative splicing (78); it will be interesting to see whether they also contribute to the switch in splicing from DUX4s to DUX4f in FSHD. Finally, and perhaps most importantly for the field, the discovery of this D4Z4 silencing mechanism highlights siRNA oligonucleotide targeting of this region as a promising therapeutic strategy in FSHD (76).

Conflict of Interest statement. None declared.

Funding

Funding from the Muscular Dystrophy UK for work on FSHD is gratefully acknowledged.

References

rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum. Mol. Genet., 2, 2037–2042.


