Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence

Silvère M. van der Maarel¹, Rabi Tawil² and Stephen J. Tapscott³

Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) has an unusual pathogenic mechanism. FSHD is caused by deletion of a subset of D4Z4 macrosatellite repeat units in the subtelomere of chromosome 4q. Recent studies provide compelling evidence that a retrotransposed gene in the D4Z4 repeat, DUX4, is expressed in the human germline and then epigenetically silenced in somatic tissues. In FSHD, the combination of inefficient chromatin silencing of the D4Z4 repeat and polymorphisms on the FSHD-permissive alleles that stabilize the DUX4 mRNAs emanating from the repeat result in inappropriate DUX4 protein expression in muscle cells. FSHD is thereby the first example of a human disease caused by the inefficient repression of a retrogene in a macrosatellite repeat array.

Facioscapulohumeral muscular dystrophy

Facioscapulohumeral dystrophy (FSHD) was first described by two French neurologists in the late nineteenth century [1]. The disease was named after its clinical presentation to distinguish it from the well known Duchenne type of muscular dystrophy. Indeed, the core phenotype of FSHD involves progressive weakness and wasting of the proximal elements on chromosome 4 (facio), shoulder and upper arm (scapulohumeral) muscles (for a detailed clinical description, see Box 1). There is currently no treatment available for FSHD (for outstanding questions, see Box 2) [2].

The major form of FSHD (FSHD1: MIM # 158900) is autosomal dominantly transmitted with linkage to the subtelomere of chromosome 4q [3]. The FSHD1 locus was mapped in 1990 [4], the first genetic condition to be mapped with polymorphic microsatellite repeats, and the nature of the genetic defect was resolved in 1993 [5]. Nevertheless, its pathogenic mechanism remains somewhat elusive. This review focuses on recent developments that establish FSHD as the first example of a human disease caused by the inefficient repression of a retrogene array.

FSHD is a repeat disease

FSHD1 is caused by a contraction of the highly polymorphic D4Z4 macrosatellite repeat in chromosome 4q (Box 3, Figure 1) [6]. The D4Z4 macrosatellite repeat is located approximately 40–60 kb proximal to the telomere repeat and varies between 11 and >100 copies of D4Z4 units in the unaffected population [6,7]. The D4Z4 repeat unit is defined as a 3.3 kb KpnI fragment, and multiple units are ordered head-to-tail to form the D4Z4 repeat array. Most patients with FSHD1 have a partial and internal deletion of the repeat array leaving only 1–10 units on one of their chromosomes 4. It is believed that at least one unit of D4Z4 is necessary to develop FSHD, as monosomy of 4q does not cause FSHD [8].

A contraction of the D4Z4 repeat array only predisposes to the disease; this contraction needs to occur on a specific chromosomal background [9]. Soon after the discovery of the D4Z4 repeat on chromosome 4, it was established that the subtelomere of chromosome 10q is almost identical to that of chromosome 4q and that it also contains a highly homologous and equally polymorphic repeat array [10,11]. A considerable proportion of individuals in the population carry 4q or 10q chromosome ends with repeat arrays that have apparently been entirely or partially transferred between both chromosomes [7,12,13]. However, contracted repeat arrays on chromosome 10 have until recently never been shown to cause FSHD [14,15].

The observation of linkage of the disease with chromosome 4 and the absence of linkage with chromosome 10 led to the hypothesis that interplay of D4Z4 with other, more proximal elements on chromosome 4 could explain the chromosome 4 specificity of the disease. In this scenario, either by spreading [16] or looping [17,18] mechanisms, the D4Z4 repeat contraction would affect the transcriptional regulation of proximal chromosome 4-specific genes. Indeed, closely located genes with high myopathic potential were reported to be transcriptionally upregulated [16] in FSHD, such as FRG1, FRG2 and ANT1. Variability between studies possibly owing to biological differences between the tissues sampled and technical differences between the various studies, however, prevented a consensus agreement on any single mechanism [19].

Homologous is not identical

The D4Z4 repeat and its homolog are located in the subtelomere of chromosomes 4 and 10. Subtelomeres are unusual domains, showing a relatively high level of plasticity...
and resulting in the frequent transfer of sequences between homologous and nonhomologous chromosome ends [20].

To understand the chromosome 4 linkage with the disease, genetic studies were undertaken leading to the identification of the mechanism of D4Z4 rearrangements and to the identification of large polymorphisms in the subtelomere of chromosome 4q [21–23]. Because approximately half of new FSHD cases arise as a consequence of a postzygotic rearrangement of the repeat leading to somatic mosaicism for the D4Z4 repeat contraction, these mosaic cases were instrumental for studying the genetic mechanism [23]. It became apparent that the preferred template for these rearrangements was the sister chromatid, instead of the homologous chromosome, making a frequent occurrence of sequence transfers between homologous chromosomes or between chromosomes 4 and 10 unlikely [23]. In addition, a large polymorphism was identified on chromosome 4 that involved the presence or absence of a β-satellite repeat immediately distal to the D4Z4 repeat array, identifying two major haplotypes of chromosome 4 called 4A (with β-satellite repeat) and 4B (without β-satellite repeat) [21]. On chromosome 10, only the A variant was identified. Whereas 4A and 4B chromosomes are almost equally common in the population, FSHD chromosomes seemed to be exclusively of the 4A type [22], which was later confirmed in independent studies [24,25]. Thus, it was concluded that attributes specific to 4A, presumably polymorphisms, confer permissiveness to the D4Z4 repeat.

This led to attention being focused on identifying allele-specific polymorphisms associated with FSHD. Initially, a simple sequence length polymorphism was identified immediately proximal to the D4Z4 repeat and was instrumental in our understanding of the genetic basis of FSHD [9,26]. Studies in patients and control individuals from different populations showed that during recent human evolution, there were probably only four events in which sequences were transferred between chromosomes 4 and 10. Subsequent detailed genetic studies of the FSHD locus led to the identification of additional polymorphisms subdividing chromosome 4 into at least 17 genetically distinct subtelomeric variants and chromosome 10 into 8 subtelomeric variants. Intriguingly, contractions in only three genetically almost identical chromosome 4 subtelomeres, the common variant 4A161 and the rare variants 4A159 and 4A168, caused FSHD, whereas contractions in other 4q subtelomeres were not associated with disease [9,26–28].

This finding provided strong evidence that genetic factors in the subtelomere of chromosome 4 contribute to FSHD pathology, and indeed detailed sequence analysis of the first and last repeat unit of the array of the most common chromosomal backgrounds identified 4A161-specific sequence variants [26,28]. Thus, not only outside the D4Z4 repeat, but also within the repeat array, it was shown that the permissive chromosomes contained specific sequence variants not shared by other (nonpermissive) chromosome ends [26,28]. In summary, these detailed genetic studies revealed that each chromosome maintained specific polymorphisms and that specific sequences in the FSHD-permissive variants of 4A confer permissiveness to this repeat.

The D4Z4 unit and its transcriptional landscape

The sequence of the D4Z4 repeat contains the open reading frame (ORF) of a double-homeobox transcription factor, DUX4 (Figure 1) [29,30]. The DUX4 ORF is in a single...
exon, whereas other members of the double-homeobox family have multiple introns, indicating that DUX4 was inserted into the genome as a retrotransposed mRNA from an intron containing the DUX gene, possibly either DUXC or, less likely, Duxbl [31–33]. In contrast to the many pseudogenes retrotransposed to our genome, the DUX4 retrogene maintained a conserved ORF [31]. However, it is unclear whether the conservation of the ORF was a consequence of a conserved functional role propagated to all repeats by concerted evolution.

Although initial attempts to identify DUX4 mRNA expression in normal development or disease were unsuccessful [34], a major advance in understanding FSHD was the identification of polyadenylated mRNA containing the DUX4 ORF using RT-PCR [35]. Initially identified only in FSHD muscle samples, the polyadenylation site of the DUX4 mRNA was mapped to the region immediately telomeric to the last D4Z4 repeat, a region previously cloned from a phage clone containing the D4Z4 repeat and flanking sequences and called pLAM1 [5]. It was proposed that the contraction of the D4Z4 array results in the transcription of the DUX4 retrogene [35]; however, the abundance of the DUX4 mRNA and protein was extremely low.

In addition, a later study [34] identified D4Z4 and DUX4 transcripts in both FSHD and control muscle. Random priming of RNA identified both sense and antisense transcripts throughout the D4Z4 region. Regions of lower transcript abundance correlated with the presence of siRNA- or miRNA-sized fragments, and it was suggested that these bidirectional transcripts and small RNAs contribute to the heterochromatin suppression of this region [34](see below). In addition, several splice forms of a DUX4 polyadenylated mRNA were identified that used the pLAM1 polyadenylation site, but these DUX4 transcripts were also at extremely low abundance [34]. Several groups demonstrated that relatively high levels of DUX4 expression were pathologic to muscle cells and other cell types [36–39]. Therefore, if DUX4 could be shown to be expressed at sufficient levels in FSHD, then it was likely to be a major cause of the muscle pathology.

**Chromatin studies**

A third important clue that could explain how a repeat contraction can cause disease came from chromatin studies of the D4Z4 array. The D4Z4 repeat is GC-rich and contains sequences often residing in heterochromatic domains of the
It was therefore postulated that normally the D4Z4 repeat is in a relatively closed chromatin configuration and that, as a consequence of repeat contraction, it would adopt a more open chromatin configuration (Figure 1). DNA methylation studies and studies of histone modifications and other chromatin factors supported this hypothesis [17,41,42]. Normally, the D4Z4 repeat is densely DNA methylated; FSHD chromosomes experience an approximate 30–40% reduction of DNA methylation at specific sites tested in D4Z4. In addition, chromatin immunoprecipitation (ChIP) studies showed that the D4Z4 repeat is normally occupied by both transcriptionally repressive as well as permissive histone modifications, whereas in FSHD chromosomes, there is a relative loss of repressive histone modifications. These changes in chromatin structure are restricted to the D4Z4 repeat and do not seem to spread proximally. ChIP studies also identified the losses or gains of other chromatin factors such as HP1γ, the cohesin complex, YY1 (losses) and CTCF (gain) at D4Z4 of disease alleles [16,42,43]. Overall, the data support a model in which D4Z4 in FSHD chromosomes adopts a relatively open chromatin structure facilitating the transcriptional activity of the repeat and possibly affecting the processing of the different D4Z4 transcripts.

Interestingly, similar changes in the chromatin structure of D4Z4 were also identified in a small cohort of patients whose disease status could not be confirmed by standard molecular diagnostic tests [42,44,45]. These patients, now classified as FSHD2, have normal, but compared to the general population, smaller-sized D4Z4 repeat arrays; disease alleles of FSHD2 patients show similar changes in D4Z4 chromatin structure as those of FSHD1 patients [45]. In contrast to FSHD1 where the relative relaxation of the D4Z4 chromatin structure seems restricted to the contracted allele, the D4Z4 repeats on both chromosomes 4 and 10 seem to be affected in patients with FSHD2. Other repeat structures in the genome of these patients seem to be normally structured [46] and the cause for this change in chromatin structure of D4Z4 in patients with FSHD2 is currently not known. Also in common with FSHD1, patients with FSHD2 have at least one permissive (4A161) chromosome [45]. Thus, it seems that patients with FSHD1 and FSHD2 share the commonality of a relative chromatin relaxation of D4Z4 on the genetic 4A161 background.

A new developmental model for FSHD

Very recently, genetic studies directly demonstrated the requirement of the DUX4 polyadenylation site for FSHD [28], and molecular studies have produced a new developmental model for the disease that is consistent with the extremely low abundance of the mRNA and protein [47] (Figure 2). Together, these studies explain the apparent discrepancies in previous models of FSHD and provide compelling support for the expression of DUX4 as a major cause of FSHD.

Figure 2. A unifying mechanism for FSHD. Upon (a) contraction of the D4Z4 repeat (FSHD1) or by (b) a yet unknown mechanism (FSHD2), the D4Z4 repeat array (triangles) adopts a more open chromatin configuration (orange > green dots) leading to the leaky expression of DUX4 mRNA. On permissive chromosomes, this mRNA is stabilized owing to the presence of a canonical polyadenylation signal immediately distal to the D4Z4 repeat array. (c) Nonpermissive chromosomes do not have this polyadenylation signal and therefore DUX4 mRNA becomes rapidly degraded. The DUX4 mRNA encodes for a nuclear double-homeobox protein that when expressed in muscle induces apoptosis.
Meticulous genetic analysis of patients with unusual hybrid D4Z4 repeat structures containing units with sequence signatures consistent with those originating from chromosomes 4 and 10 revealed a common last portion of the D4Z4 repeat array and flanking pLAM1 sequence [28]. This old-fashioned positional cloning strategy strongly argued that the distal end of the repeat array and flanking pLAM1 sequences are crucially important for the development of FSHD [28]. Further corroborating this finding was the identification of a FSHD family in which the disease segregated with a contracted D4Z4 allele of chromosome 10. Importantly, the last part of this disease-associated repeat array was replaced by permissive chromosome 4 sequences. The identification of this family in which FSHD segregates with chromosome 10 essentially confirms the importance of the distal end of the repeat and pLAM1 sequences and precludes a prominent role for other proximal candidate genes on chromosome 4.

Further genetic studies of the distal end of the D4Z4 repeat array and flanking sequences allowed almost perfect separation of permissive and nonpermissive chromosome ends based on the identification of consistent polymorphisms in the region sequenced [28]. One noticeable difference between permissive chromosomes 4 and nonpermissive chromosomes 10 was the presence of a DUX4 polyadenylation signal on chromosome 4, whereas on chromosome 10 this polyadenylation signal was lost because of the presence of independent polymorphisms [28]. Indeed, when transfecting the crucial region in murine C2C12 muscle cells, stable DUX4 transcripts making efficient use of the DUX4 polyadenylation signal could only be identified when constructs derived from permissive chromosomes with a polyadenylation signal were transfected [28]. Interestingly, a study of the recent hominoid evolution of the 4q subtelomere shows that the permissive chromosome end is ancestral to all 4q chromosome ends, and the data are consistent with an evolutionary pressure to eliminate the third DUX4 exon in pLAM1 [26]. These genetic studies clearly demonstrated the requirement for the polyadenylation site utilized by DUX4 mRNA and, therefore, strongly implicated DUX4 protein as a cause of FSHD. However, as noted above, although DUX4 mRNA was detected in FSHD muscle, it was still at extremely low abundance.

Low abundance mRNA in a population of cells could reflect either a small amount of mRNA in all cells or an abundant amount of mRNA in just a few cells. RT-PCR amplification of DUX4 mRNA in small pools of 100 or 600 differentiated FSHD muscle cells identified relatively abundant transcripts in a subset of the pools [47]. The frequency of positive pools suggested that approximately 1 in 1000 FSHD muscle cell nuclei were expressing an abundant amount of DUX4 mRNA. Immunodetection confirmed that approximately 0.1% of nuclei in cultured FSHD muscle cells expressed an abundant amount of protein. In addition, the DUX4-expressing FSHD muscle nuclei had characteristics consistent with DUX4 induced toxicity, i.e., an aggregation of nuclear DUX4 protein that occurs coincident with DUX4-induced apoptosis. Therefore, the very low abundance of DUX4 mRNA in FSHD muscle represented relatively abundant amounts of DUX4 mRNA and protein in a small subset of the nuclei, probably leading to dysfunction or death of those DUX4-expressing nuclei.

As a retrogene, DUX4 was viewed by some as a dead gene, or pseudogene, brought ‘back to life’ by the contraction of the D4Z4 array, but the conservation of the ORF suggested otherwise [32–34]. Indeed, DUX4 mRNA and protein was shown to be highly expressed in the testes of unaffected individuals, most probably in the germline, both from the permissive 4A161 allele as well as from the nonpermissive alleles [47]. It is interesting to note that the germline DUX4 mRNA transcripts from chromosome 10, which lacks the polyadenylation site that is present on the permissive chromosomes, use an alternative polyadenylation site approximately 6 kb telomeric to the end of the D4Z4 array. Use of this alternative polyadenylation was restricted to germline tissues and not identified in somatic tissues. The inefficient chromatin-mediated repression, either related to the contraction of the array in FSHD1 or through unknown mechanisms in FSHD2, results in the occasional escape from repression in muscle cells, and possibly other somatic cells. In this model, the muscle cell nuclei would be lost over time in FSHD because of the inappropriate expression of DUX4 protein (Figure 2).

Concluding remarks

These recent studies substantiate a developmental model of FSHD that explains many of the previously unexplained mysteries of this human disease. First, genetic studies demonstrate the requirement for the DUX4 polyadenylation site in the pLAM1 region of the permissive alleles, indicating that DUX4 mRNA is crucial for FSHD [28]. Second, molecular studies show decreased density of repressive chromatin modifications in both FSHD1 and FSHD2 [6], indicating that DUX4 mRNA is more likely to be expressed. Third, RNA and protein studies showed an occasional escape from the inefficient chromatin repression leading to high levels of DUX4 expression in a small number of nuclei in FSHD muscle cells [47]. Fourth, abundant expression of DUX4 in testis, most probably the germline cells, indicates that this retrogene might have a normal role in germ cell development [47]. Finally, if a retrogene has subsumed a normal role in germ cell biology, then repression of that gene in somatic cells needs to co-opt regulatory mechanisms distinct from the evolved enhancers and promoters of the parental gene. Therefore, the repression of DUX4 in somatic cells is probably a mechanism adopted from other loci, such as the mechanism of silencing retrotransposons and other repetitive elements, and is not highly evolved for the DUX4 locus. In this case, co-opted mechanisms might not be sufficiently robust to avoid disease, as is evident by the association of FSHD with the contracted D4Z4 array.

This new unifying and substantiated model of FSHD has one additional profound and as yet unexplored implication. DUX4 arose from the retrotransposition of a parental DUX4 mRNA, possibly either DUXC or, less likely, Duxbl [31–33]. Both DUXC and Duxbl are expressed in the germ-line, a requirement for introducing a retrogene into the population, but neither is present in primates [33]. There-
fore, primates have retained the retrogene and lost the parental gene, suggesting a selective advantage of the retrogene. Primates have sacrificed upper extremity and facial muscle mass for the advantage of an upright posture and highly expressive facial muscles. Although this remains highly speculative, it is interesting to suggest that the DUX4 retrogene might have been retained in preference to the parental gene because inefficient chromatin repression results in sufficient expression in skeletal muscle to modulate facial and upper extremity muscle mass, even in individuals without the FSHD deletion. If this is correct, then FSHD is a hypermorphic phenotype for traits that are crucial for primate evolution, and the mystery of FSHD still has the potential to lead us to new understandings of human biology.

Acknowledgments

Our research is supported by the Fields Center for FSHD and Neuromuscular Research, the National Institutes of Health (NINDS PO1NS069539, NIAMS R01AR045203 and NIAMS R21AR059966), the Muscular Dystrophy Association (173202), the Prinzes Beatrix Fonds (WAR08-14), the Shaw Family Foundation, the FSH Society, the Dutch FSHD Foundation and the Pacific Northwest Friends of FSH Research.

References

5 van Deutekom, J.C. et al. (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum. Mol. Genet. 2, 2037–2042
9 Lemmers, R.J. et al. (2007) Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. 81, 884–894
18 Bodega, B. et al. (2009) Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. BMC Biol. 7, 41
22 Lemmers, R.J. et al. (2002) Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat. Genet. 32, 235–236
27 Spurlock, G. et al. (2010) Confirmation that the specific SSLP microsatellite allele 4qA161 segregates with facioscapulohumeral muscular dystrophy (FSHD) in a cohort of multiplex and simplex FSHD families. Muscle Nerve 42, 820–821
46 de Greef, J.C. et al. (2007) Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. *Neurology* 69, 1018–1026