CMLS, Cell. Mol. Life Sci. 61 (2004) 557–566
1420-682X/04/050557-10
DOI 10.1007/s00018-003-3285-3

CMLS Cellular and Molecular Life Sciences

Review

Molecular basis of facioscapulohumeral muscular dystrophy

R. Tupler[a, b, *] and D. Gabellini[a]

[a] Program in Gene Function and Expression, University of Massachusetts Medical School, Lazare Medical Research Building, 364 Plantation Street, Rm. 660, Worcester, Massachusetts 01605 (USA)
[b] Biologia Generale e Genetica Medica, via Forlanini 14, Universitá degli Studi di Pavia, 27100 Pavia (Italy),
Fax: +39 0382 525030, e-mail: rossella.tupler@umassmed.edu

Received 16 July 2003; received after revision 26 August 2003; accepted 1 September 2003

Abstract. Facioscapulohumeral muscular dystrophy (FSHD), the third most common myopathy, is an autosomal dominant disease with an insidious onset and progression. Almost all FSHD patients carry deletions of an integral number of tandem 3.3 kb repeats, termed D4Z4, located on chromosome 4q35. In FSHD patients a deletion of the integral number of D4Z4 repeats generates a fragment that is usually smaller than 35 kb (fewer than 11 repeats), whereas in normal controls the size usually ranges from 50 to 300 kb (between 11 and 150 units). D4Z4 is a repetitive element with heterochromatic features. Recently, 4q35 genes located upstream of D4Z4 have been found to be inappropriately overexpressed specifically in FSHD muscle. An element within D4Z4 has been shown to behave as a silencer that provides a binding site for a transcriptional repressing complex. These results suggest a model in which deletion of D4Z4 leads to the inappropriate transcriptional derepression of 4q35 genes, resulting in disease.

Key words. FSHD; transcriptional derepression; repetitive DNA; muscle-specific transcription.

Facioscapulohumeral muscular dystrophy: an introduction

Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common hereditary myopathies, with an incidence of 1 in 20,000 in the general population [1]. FSHD is characterized by the progressive wasting of a highly selective set of muscle groups. The onset of FSHD involves the weakening of certain facial muscles, such that patients may not be able to close their eyes, puff up their cheeks or smile. As the disease progresses, the muscular weakness spreads to limb-girdle muscles, such as scapula fixators, and deltoid muscles; as a consequence patients are unable to lift their arms. Abdominal muscles also become affected, leading to a characteristic posture in which the patient's stomach protrudes due to an abnormal inward curvature of the spine. Eventually, the disease extends to pelvic girdle muscles, thereby affecting the individual's ability to walk. A notable characteristic feature of the disease is that the muscle weakness displays an asymmetric distribution, which does not correlate with the handedness of the individual [2]. The chronology of disease progression is unpredictable; for example, long periods of stability can be followed by sudden and dramatic worsening. In addition, there is a wide variability in clinical spectrum among patients with the disease, ranging from subjects with very mild muscle weakness, who are almost unaware of being affected, to patients who are wheelchair dependent. This variability in disease penetrance is exemplified by a set of monozygotic male twins, shown in figure 1, who carry the same genetic mutation but are affected by FSHD to dramatically different extents [3]. Electromyographic and histological analysis reveals nonspecific myopathic changes, and therefore diag-

* Corresponding author.

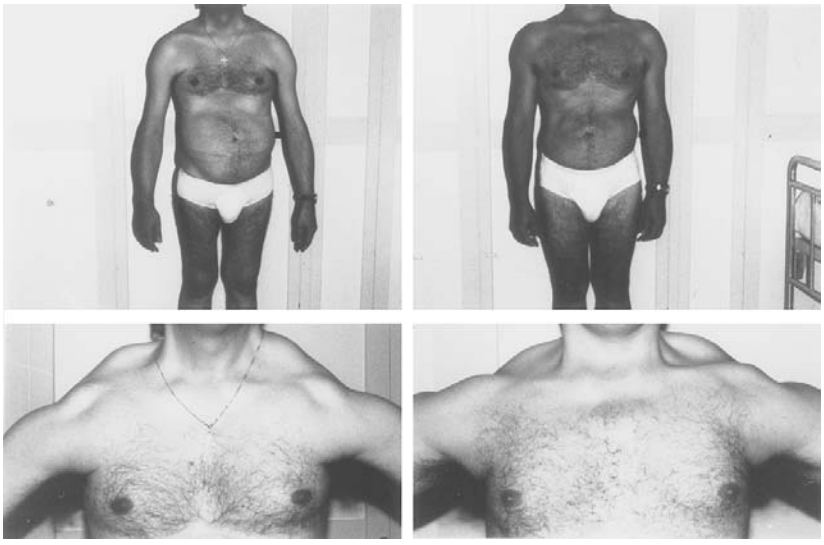

Figure 1. A set of monozygotic twins showing different clinical expression of FSHD. In the upper panels, remarkable differences of muscle groups are visible. On the left, the more severely affected twin shows a marked reduction in muscle mass of biceps and quadriceps; his stomach protrudes because of weakness in the abdominal muscles. In the lower panels, weakness of scapula fixators is revealed during arm abduction by the elevation of *trapezius*. It is evident that the more severely affected twin, on the left, is not able to lift his arms.

nosis of FSHD has been mainly based on clinical phenotype [4], although a molecular diagnosis is now available (see below).

FSHD is a genetic disorder that is transmitted in an autosomal dominant fashion. FSHD1, which defines the major form of FSHD, genetically maps to chromosome 4q35 [5–8]. At least 5% of FSHD families are not linked to the genetic locus on chromosome 4, suggesting that at least one additional genetic locus associated with FSHD is present in the genome [5, 9–12]. However, to date, a second putative genetic locus has not yet been defined for the non-4q linked form(s) of FSHD.

FSHD1 has been associated with DNA rearrangements characterized by an allele with a shorter *EcoR*I fragment as detected by Southern analysis [13]. The polymorphic genomic region detected consists of an array of tandemly repeated 3.3-kb segments (hereafter referred to as D4Z4 repeats), such that the variation in the size of *EcoR*I fragments is due to variability in the number of D4Z4 repeats [14]. The number of D4Z4 repeats varies from 11 to 150 in the general population, whereas less than 11 repeats are present in sporadic and familial FSHD patients.

The D4Z4 unit, as outlined in figure 2, consists of several known sequence motifs: Lsau, hhspm3 and a double homeobox [15–17]. Lsau is a middle repetitive element associated with heterochromatic regions of the genome [18], and hhspm3 is a low-copy GC-rich element [19]. Homeobox sequences, which are highly conserved during evolution [20], encode precisely defined protein domains, known as homeodomains, that are characteristic of some gene regulatory proteins, homeobox proteins, coordinating the expression of sets of genes during development. The D4Z4 sequence also contains a putative promoter with a TACAA box located 149 bp upstream of an 1173 bp ORF (open reading frame) and is predicted to encode a gene known as DUX4, which has two homeodomains [21]. However, the predicted gene has no introns and lacks a polyadenylation site. Furthermore, no D4Z4-specific transcript has been so far identified either in muscles or other tissues, suggesting that DUX4 is most likely a pseudogene.

D4Z4 is part of a family of repetitive elements, known as the 3.3-kb family, which are present at 1q12 and on the short arm of acrocentric chromosomes in the human genome [22]. Interestingly, a study on the evolutionary distribution of 3.3-kb repeats revealed that in Old World monkeys 4qter contains the ancestral copy of this repeat, whereas in great apes D4Z4-like sequences become dispersed among chromosomes. In particular, hybridization data suggest that the copy number and organization of the 3.3-kb family are similar in chimpanzee and humans [23, 24].

These data collectively suggest that D4Z4 is a heterochromatic element belonging to a family of repeats, whose complexity and chromosomal dispersion have increased during recent evolution. This phenomenon may be related to the presence of highly repetitive DNA within and adjacent to the D4Z4 array that may facilitate recombination and translocation events, which may be linked to a functional role of these repetitive elements.

Molecular diagnosis

Molecular diagnosis of FSHD1 is based on the analysis of the D4Z4 polymorphic alleles detected by Southern analysis using the probe p13E-11 [13]. Normal subjects carry p13E-11 *Eco*RI alleles larger than 35 kb originating from chromosome 4, whereas alleles shorter than 35 kb are present in the majority of either de novo or familial FSHD patients [25], as schematically depicted in figure 3. Consistently, de novo deletions transmitted from an affected parent to her/his offspring cosegregate with the disorder [26]. New mutations account for a surprisingly high percentage of FSHD patients (10–33%) [27, 28]. This high incidence can be partly explained by the presence of parental mosaicism for 4q short alleles that has been reported in 40% of de novo cases [26, 29–32].

A complication of molecular testing by Southern analysis is caused by the presence of a similar polymorphic region recognized by the probe p13E-11 located at the subtelomeric region of chromosome 10q [12, 33]. This region at 10q26 shares numerous similarities with the 4q35 subtelomeric region. The repeat element at 10q is 98%

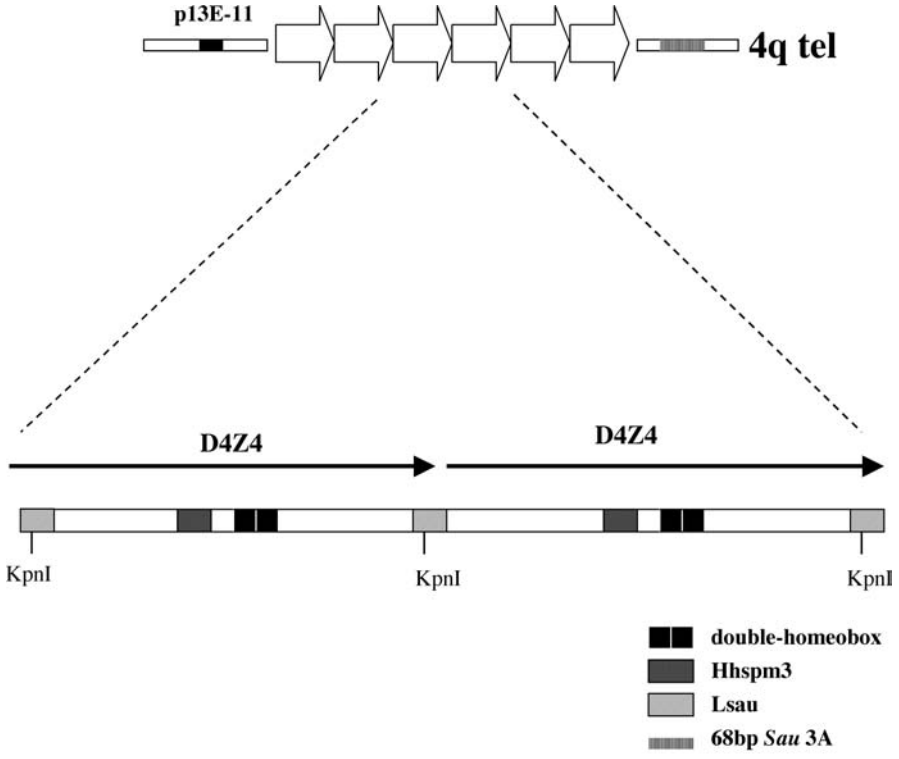

Figure 2. Schematic representation of an array of 3.3 tandemly repeated units (D4Z4) as reported by Hewitt et al. (1994). The position of sequence motifs are shown. The internal repeated units vary in number, and the most proximal repeated unit is located adjacent to p13E-11.

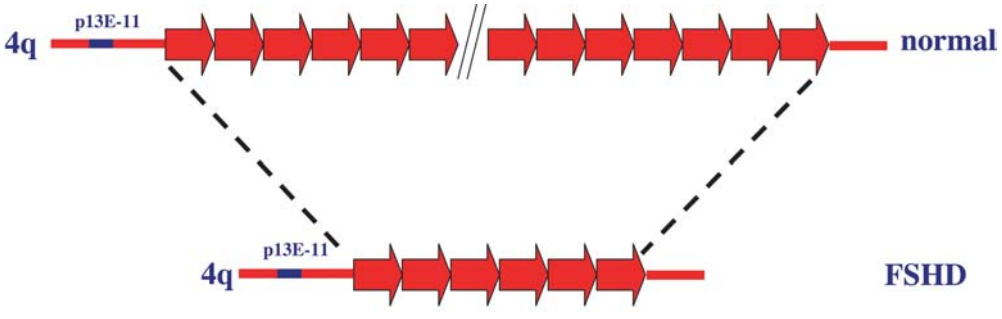

Figure 3. A model depicting the chromosome 4q35 allele of a normal individual (top) and a FSHD patient (bottom) at the D4Z4 locus. Allele sizes are detected by Southern hybridization of the probe p13E-11 on *EcoR1* and *Eco*RI/*Bln*I digested genomic DNA. The number of 3.3-kb repeated elements (arrow) is critical to the disease onset. It varies between 11 and 150 units within the normal population, whereas FSHD patients carry fewer than 11 repeats.

homologous to D4Z4, and the size of 10q *Eco*RI alleles varies between 11 and 300 kb. Moreover, 10% of these alleles are shorter than 35 kb [12, 31], overlapping the 4q alleles, and clear-ly this can complicate molecular diagnosis of FSHD. However, the presence of a *Bln*I restriction site within the 3.3-kb element uniquely associated with chromosome 10q allows the discrimination between 4q and 10q alleles [33]. As a result, Southern blot hybridization of *Eco*RI and *Eco*RI/*Bln*I digested genomic DNA is currently used for the molecular diagnosis of FSHD [25].

Due to the homology between the chromosome 10q and 4q repeats, these sequences are sometimes rearranged. The presence of chromosome 10-type repeats on chromosome 4 has been observed in 10% of the Dutch population. The reverse configuration of chromosome 4-type repeats on chromosome 10 is equally frequent [34, 35], and hybrid repeats, consisting of both 4-type and 10-type repeat units, have also been identified. These data demonstrate that recombination between 4q35 and 10q26 occurs frequently, and show the dynamic characteristics of the 4q-10q homologous domain [36, 37]. Most important, these observations confirm that FSHD is not caused by alteration of the specific sequence composition of the D4Z4 repeats but rather by their structural organization in the 4q35 region.

Recently, Lemmers and colleagues [38] described a DNA region distal to D4Z4, and also present at 10q26, that shows a biallelic polymorphism in the population. Interestingly, they have found that one of the two alleles, 4qA, appears to be uniquely associated with FSHD, thus pointing to a functional role of this allele in the causation of the disease. This is an intriguing hypothesis that warrants further study. Pertinent to this question will be the verification of the clinical status of six subjects carrying D4Z4 alleles shorter than 35 kb that are equally associated with 4qA and 4qB alleles [35].

The FSHD region on chromosome 4q35 contains genes and pseudogenes

Although it is well established that deletions of D4Z4 are causally related to FSHD, it is not clear how this triggers the disease. It has long been speculated that such deletions may alter the expression of genes located within or nearby the repeats. Nonetheless, the search for a FSHD candidate gene has been quite frustrating. Bengtsson et al. [39] demonstrated that D4Z4 lies adjacent to a subtelomeric sequence, which is within 5–14 kb of a telomeric repeat, and little if any DNA exists between the telomere and the D4Z4 array [40]. No genes have been described or predicted distal to the array. Therefore the gene search has been extended to the region proximal to the repeat.

Numerous approaches such as exon-trapping and complementary DNA (cDNA) selection have been applied to the 4q35 region with very little success [41]. Only recently a gene named *FRG1* (*FSHD* Region Gene 1) has been identified within this region, although it is located approximately 100 kb centromeric to the repeated units [34]. The gene encodes a 258-amino acid protein whose function has not yet been elucidated. Although the FRG1 protein does not share significant overall homology to any known protein, it does contain a high proportion of charged amino acids and a so-called fascin domain (catalogued as CDD1776 in the protein database at www.ncbi.nlm.nih.gov/Structure/cdd/). Interestingly, multiple FRG1-related sequences are present in the genome, even though many are considered pseudogenes. The copy at 4q35, however, appears to be conserved throughout evolution, as homologues have been identified in mouse [42], *Fugu rubripes* and *Caenorhabditis elegans* [43].

Another novel gene, called *FRG2*, (*FSHD* Region Gene 2), has been mapped about 37 kb proximal to D4Z4. FRG2 expression was observed in differentiating myoblasts obtained from FSHD patients, but no detectable transcript was present in normal controls [R. R. Frants and S. van der Maarel, personal communication]. Multiple FRG2-related sequences are present in the human genome. As with FRG1, its biological function remains unknown.

A putative gene has been also identified that maps 80 kb centromeric to D4Z4 on chromosome 4q35 [44]. This gene, TUBB4Q, is part of the β-tubulin supergene family and encodes a putative protein of 434 amino acids that shares 87% homology with β2-tubulin. Notably, TUBB4Q-related sequences are present on multiple chromosomes. However, attempts to search for a TUBB4Q-specific transcript(s) in muscles or other tissues have failed, suggesting that TUBB4Q is most likely a pseudogene.

The 4q35 subtelomere containing D4Z4 shows typical features of heterochromatic regions and shares strong similarities with other subtelomeres [45]. The distal subdomain on 4qter comprises 25 kb and contains repetitive sequences typically found at many chromosome ends, whereas the proximal domain is largely comprised of D4Z4 repeats and its size may vary from 50 to more than 500 kb due to the D4Z4 polymorphism. This region has proven to be a very complex region of the genome, and sequences closely related to loci on distal 4q35 are dispersed throughout the human genome [15, 16, 41].

Interestingly, phylogenetic analysis demonstrates that similar to humans, multiple copies of FRG1 and TUBB4Q are present in great apes, and show a striking similarity to the distribution of the 3.3-kb repeat family in primates [46, 47]. This is consistent with previous observations suggesting that this region underwent further amplification and dispersion in the great ape lineage [23]. Duplications associated with 4q35 could be a direct result of the close proximity to the telomere. However, the func-

tion of the region has not yet been determined. It would be intriguing to determine whether this phenomenon correlates with the acquisition of specialized functions developed during recent evolution, such as mimicking abilities or brachiation.

Expression of 4q35 genes is upregulated in FSHD muscle

Multiple lines of evidence, such as the lack of candidate genes and the causative association of the deletion of repetitive elements with the disease, indicate that FSHD is not the result of a classical mutation within a protein-coding gene. Instead, the genomic organization of the 4q35 subtelomeric region strongly argues for its role in control of gene expression.

The hypothesis that genes within this region are not normally expressed laid the basis of a novel working model which proposed that the inappropriate silencing of 4q35 genes was due to a positional effect. One immediate prediction of the model was that deletions of D4Z4 would place 4q35 genes under the control of telomeric heterochromatin, thereby causing their transcriptional silencing [13, 15, 16, 48]. However, several observations did not fit with this hypothesis. First, the 100-kb region proximal to the repeat does not contain genes expressed in normal tissues [49]. Second, complete deletion of a 200-kb region, including the D4Z4 repeat, of one of the two copies of chromosome 4 had no phenotypic consequences, indicating that FHSD is not the result of reduced gene transcription within 4q35 [50]. Finally, increasing evidence indicates that it is the expansion of repeated elements or long tandemly arrayed sequences – not the reduction of these elements – that is associated with suppression of transcription [51–53].

An alternative model for the molecular basis of FSHD proposed instead that D4Z4 deletion might induce abnormal transcriptional activity of the 4q35 genes. This hypothesis was fostered by several observations. First, as described above, D4Z4 displays features common to heterochromatic elements, in that it has a high CG content [15] and is hypermethylated in normal tissues [54]. Thus one prediction is that reduction of D4Z4 elements derepresses gene transcription. Second, all approaches aimed at identifying genes within the region, such as exon trapping, cDNA selection or bioinformatic analysis, failed, strongly leading to the conclusion that the 4q35 region is either gene poor, or that genes within this region are not expressed in most fetal and adult normal tissues and, therefore, are under-represented in most commercial cDNA libraries.

To test the latter hypothesis, a differential messenger RNA (mRNA) screen on normal and FSHD muscles was carried out [55]. The results, although complex, have highlighted a clear scenario in which the transcriptional profiling of FSHD dystrophic muscle is profoundly changed compared with that of normal muscle. Intriguingly, several transcriptional regulators were found to be either up- or downregulated in the affected muscle tissue, suggesting that a transcription regulation defect might be the underlying basis of the FSHD dystrophic process. This hypothesis has been further strengthened by recent work showing that the expression level of known 4q35 genes such as FRG2, FRG1 and ANT1 is specifically elevated in FSHD dystrophic muscles [56].

D4Z4 contains a transcriptional repression element that is bound by a multi-protein repressor complex

The specific alteration of 4q35 gene expression in FSHD-affected muscle raised the question of whether D4Z4 was directly involved in transcriptional control of 4q35 genes. To test this possibility, the interaction between D4Z4 and nuclear proteins was analyzed using an electrophoretic mobility shift assay (EMSA). This approach showed that one region of D4Z4 supported formation of a specific complex [56]. The binding site was further mapped to a 27-bp sequence called the D4Z4 binding element (DBE). Biochemical studies demonstrated that the DBE binds a multi-protein complex in vitro and in vivo comprising of YY1, HMGB2 and nucleolin, termed the D4Z4 repressing complex (DRC) [56].

These three proteins possess interesting features. YY1 is a complex protein that is involved in repressing and activating a number of promoters. It interacts with numerous key regulatory proteins, suggesting that these interactions are important for determining which particular function of YY1 is displayed at specific promoters [57]. Furthermore, the activity of YY1 is modulated by histone deacetylases and histone acetyltransferases [58]. These two groups of enzymes modify histones that, in turn, alter chromatin structure. Interestingly, the DBE contains a putative YY1 recognition sequence (CCATN [59]). EMSA experiments indicated that YY1 has an intrinsic DBE binding activity and suggested that YY1 is the most likely DRC component to directly interact with the DBE [56].

HMGB2 is a member of one of the three families of high mobility group (HMG) proteins (for reviews, see [60–62]). The HMGB family comprises the ubiquitous HMGB1, HMGB2, and HMGB3 proteins, of which HMGB1 is the most-studied member. HMGB1 binds DNA in a non-sequence specific manner, interacting with the minor groove of DNA and bending the DNA segment. HMGB1 also binds with high affinity to unusual DNA structures, such as four-way junctions and DNA bulges, and can be recruited by a variety of sequence-specific DNA binding proteins that facilitate the assembly of multiprotein complexes on DNA. It is therefore considered to

have an 'architectural' function. HMGB2 is extremely similar to HMGB1 (greater than 80% amino acid identity). It has been shown that HMGB1 and HMGB2 are completely interchangeable in vitro: they both bind to HOX proteins [63], steroid hormone receptors [64] and RAG1 recombinase [65], and both enhance the transcription and recombination activities of their binding partners. However, knockout experiments indicate that the two proteins have nonredundant biological functions in vivo [66, 67]. Furthermore, a yeast two-hybrid screen identified SP100 as an interactor of HMGB2; SP100 in turn interacts with HP1, the cardinal heterochromatin-binding protein, raising the possibility that HMGB2 might be involved in the organization and/or maintenance of heterochromatic regions [68].

The third component of the DRC is nucleolin, which is an abundant protein in the nucleolus. The presence of several structural domains in nucleolin allows its interaction with different proteins and RNA sequences. Nucleolin has been implicated in chromatin structure, ribosomal RNA (rRNA) transcription, rRNA maturation, ribosome assembly and nucleo-cytoplasmic transport (for a review, see [69]).

All three DRC proteins bear versatile activities and can be envisioned to constitute a dynamic structure at the 4q35 locus that can control gene expression at different levels and extents. In particular, the observation that in FSHD muscle genes mapping at 4q35 were specifically upregulated raised the question of whether deletion of D4Z4 elements, such as in FSHD muscle, might induce a reduction of DRC components at 4q35, resulting in increased activity of genes in cis. To address whether the level of the DRC components influenced transcription of 4q35 genes, antisense experiments to decrease intracellular levels of DRC components were performed. These experiments showed that depletion of YY1, HMGB2 or nucleolin resulted in overexpression of the 4q35 gene *FRG2*, which was been shown to be silent in normal cells and tissues [56]. These experiments might recapitulate the molecular events observed at 4q35 in FSHD muscle, where contraction of the D4Z4 repeat is expected to reduce the levels of DRC components. These observations support a model in which deletion of D4Z4, as happens in FSHD patients, causes inappropriate transcriptional activation of 4q35 genes in cis, resulting in disease.

A new model for FSHD

For a long time, FSHD has been a hereditary disorder in search of a genetic defect. The recent results suggest a novel model for the molecular basis of FSHD (figure 4A). In normal individuals, the presence of a threshold number of D4Z4 repeats leads to repression of 4q35 genes by virtue of a DNA-bound multi-protein complex that actively suppresses gene expression. In FSHD patients, deletion of an integral number of D4Z4 repeats reduces the number of bound repressor complexes and consequently decreases (or abolishes) transcriptional repression of 4q35 genes. The model suggests that deletion of repeated elements in the subtelomeric region of 4q may act in cis on neighboring genes by derepressing their transcription and thus starting a cascade of events which ultimately lead to FSHD. This idea is also in agreement with the observation that haploinsufficiency of distal 4q does not cause FSHD [50].

The derepression model is consistent with the clinical and genetic features of FSHD as follows: (i) gene overexpression provides a molecular explanation for the autosomal-dominant transmission of FSHD; (ii) the observation that 4q35 gene overexpression is muscle specific can explain the tissue specificity of the disease; (iii) D4Z4 deletion may result in variegated expression of 4q35 genes in FSHD muscle cells by a mechanism similar to position effect observed in *Drosophila* and yeast; this stochastic variation in gene expression in muscle cells can explain the asymmetric involvement of muscles, the great variability of clinical expression between and within families, and the apparent threshold effect whereby there is a requirement for the deletion of a certain number of copies of D4Z4; (iv) an increased reduction of D4Z4 repeats might extend the derepression proximally down the chromosome to include other genes that may also contribute to the nonmuscle aspects of the disease, such as retinopathies and mental retardation; (v) moreover, in addition to overexpression of 4q35 genes, factors such as allelic variability of 4q35 genes, gender and environment that might interfere with the activity of DRC can explain the large variability in disease onset and clinical severity observed in FSHD patients.

The proposed model might also explain the disease onset in the 5–10% of FSHD cases that do not carry D4Z4 deletions and are therefore not currently considered to be linked to the 4q35 locus. One possibility is that mutations in either the DBE sequence or one of the genes encoding a DRC component might lower the activity of the repressing complex, triggering the derepression of cis genes and eventually the disease (fig. 4B). Because association of FSHD with the 4q35 locus is based solely on the detection of D4Z4 deletions, mutations in the DBE would be undetectable by current diagnostic methods and would therefore be inappropriately classified as non-4q35-linked cases. Furthermore, this model can explain sporadic and familial FSHD cases that do not carry D4Z4 deletions. It is plausible to speculate that point mutations occurring in a particular domain of one of the DRC components might interfere with a tissue-specific activity of the protein. This possibility can fit both dominant and recessive modes of inheritance. Consistent with this hypothesis, sporadic cases might be due not only to de novo dominant mutations, but also to recessive alleles present in the parents.

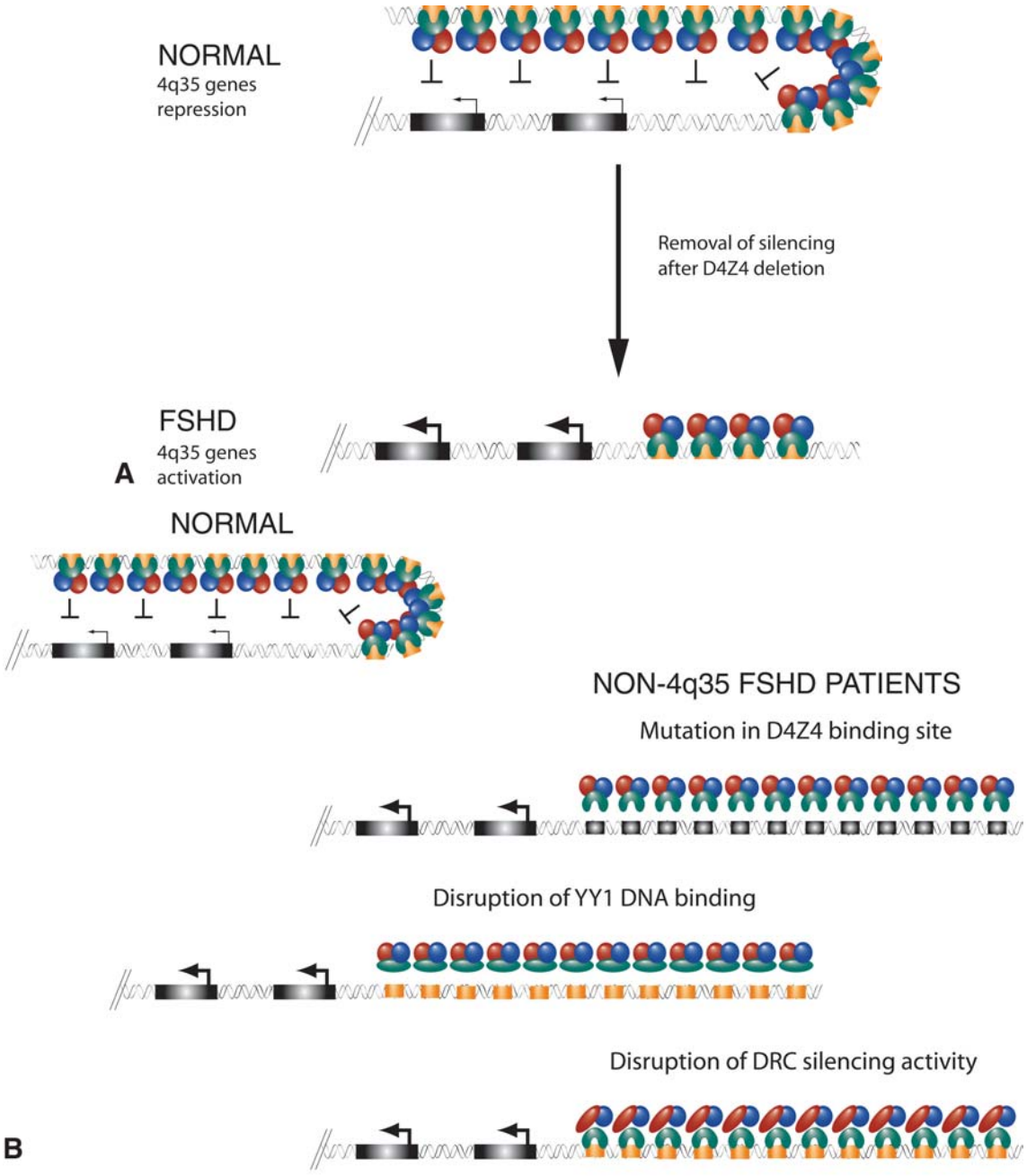

Figure 4. A model for the molecular basis of FSHD. (*A*) In normal individuals, the presence of a threshold number of D4Z4 repeats leads to repression of 4q35 genes by virtue of a DNA-bound multi-protein complex that actively suppresses gene expression. In FSHD patients, deletion of an integral number of D4Z4 repeats reduces the number of bound repressing complexes and consequently decreases (or abolishes) transcriptional repression of 4q35 genes. As a result, these genes are inappropriately overexpressed, ultimately leading to disease onset and progression.(*B*) The proposed model might also explain the disease onset in the 5–10% of FSHD cases that are currently not classified as being linked to the 4q35 locus. A possibility is that mutations in either the DBE sequence or one of the genes encoding a DRC component might lower the activity of the repressing complex, triggering the derepression of cis genes and eventually the disease.

Perspectives

The derepression of 4q35 genes could lead to a dystrophic phenotype through different pathways depending on the biological features of the overexpressed genes. One can envision that the pathological effect might be linked to overexpression of one or more genes that are critical regulators of muscle cell biology. Gene overexpression in FSHD might trigger a series of events resulting in the profound transcriptional misregulation previously reported in the FSHD dystrophic muscle that eventually leads to muscle wasting. Another possibility is that

gene overexpression causes a toxic effect due to protein overloading, or that one of the genes involved alters some component of structural organization of the muscle cell. To address these fundamental questions two critical points need to be clarified. First, although gene activation observed in FSHD dystrophic muscle involves a set of genes located at 4q35, it is not yet clear which gene(s) are primarily involved in FSHD pathogenesis. Human genome sequence analysis has uncovered only a limited number of *bona fide* and predicted genes within 4q35. These data are consistent with the hypothesized heterochromatic nature of this chromosomal region and might have hampered the identification of novel genes or expressed sequence tags (ESTs). Therefore, it will be of primary interest to reexamine the 4q35 region for possible candidate genes that may have been missed by traditional approaches.

Second, as has been already discussed, the natural history of FSHD is elusive. The pathology of muscle tissue affected by FSHD is nonspecific, sharing similarities with other muscular dystrophies. Diagnosis for FSHD is now mainly supported by the molecular analysis of DNA. The very interesting aspect of FSHD is the peculiar involvement of specific muscle groups. This observation suggests that the molecular defect underlying FSHD might affect diverse muscles to different extents.

Is this related to the fiber type composition of each muscle? Is this due to the different physiology of each muscle group? Is this linked to muscle embryonic origin or development? How important are environmental factors in determining muscle sufferance in FSHD? Animal models reproducing certain features of FSHD will be a valuable and abundant source of biological material to study the biological effects of 4q35 gene overexpression in muscle.

Altered control of gene expression is an intriguing feature of the pathogenic mechanism causing FSHD. Gene activation appears to be muscle specific, and involves only a subset of genes located at 4q35 [R. Tupler, unpublished observations]. Several factors or molecular mechanisms may contribute either independently or cooperatively to this phenomenon. For example, the components of the DRC might have a specific level of activity only in muscle cells; therefore, muscle cells may be particularly sensitive to the loss of repressor function. Alternatively, other regulatory proteins could confer tissue-specific activity to the DRC. It is also plausible that tissue-specific transcription factors could promote the expression of 4q35 genes in muscle tissue; as a consequence, in the absence of a functional repressor, these genes would be expressed.

D4Z4 deletions might also affect 4q35 gene expression through modification of higher-order chromatin structure. Patterns of gene expression that are established and maintained during cellular differentiation not only result from the targeting of stage-specific transcription factors, but can also be influenced by the long-range organization of chromatin, which establishes ‘open’ or ‘closed’ conformations that are permissive or refractory to transcription. It is possible that the DRC mediates a ‘closed’ higher-order chromatin structure. Consistent with this possibility, the DRC contains a structural chromatin factor, HMGB2, and is bound to the DNA via a sequence-specific DNA-binding factor, YY1, which is, as expected, implicated in both transcriptional activation and repression, depending on the context. Moreover, multiple YY1 binding sites are present in repetitive DNA. These features of the DRC fulfill the criteria proposed by Gasser [70] to define heterochromatin-type repression. Therefore, FSHD can be considered a model for studying control of tissue-specific gene expression in humans.

In conclusion, FSHD pathogenesis provides a new paradigm for dominantly inherited genetic diseases by demonstrating that repetitive elements, which represent a large portion of the human genome, can play an important role in gene regulation. FSHD can be a useful model for studying muscle cell biology and also offers the possibility of studying the role of chromatin conformation and nuclear position in control of muscle-specific gene expression.

Acknowledgments. We apologize in advance to those whose work we were not able to cover due to space constraints. We thank G. Perini, L.W. Ranum and Sara Evans for critically reading the manuscript. This work was supported by a development grant from the Muscular Dystrophy Association to D.G., as well as grants from the National Institutes of Health (R21 NS43973), the Muscular Dystrophy Association, the Ministry of Italian University and Research, the Italian National Research Council and Telethon (GP284Y01) to R.T.

1 Padberg G. (1982) Facioscapulohumeral disease. Thesis, Leiden University
2 Brouwer O. F., Padberg G. W., Ruys C. J., Brand R., de Laat J. A. and Grote J. J. (1991) Hearing loss in facioscapulohumeral muscular dystrophy. Neurology **41:** 1878–1881
3 Tupler R., Barbierato L., Memmi M., Sewry C. A., De Grandis D., Maraschio P. et al. (1998) Identical de novo mutation at the D4F104S1 locus in monozygotic male twins affected by facioscapulohumeral muscular dystrophy (FSHD) with different clinical expression. J. Med. Genet. **35:** 778–783
4 Munsat T. (1994) Facioscapulohumeral dystrophy and the scapuloperoneal syndrome. In: pp. 1220–1232, Myology. Andrew G. and Franzini-Armstrong C. (eds.), McGraw-Hill, New York
5 Wijmenga C., Padberg G. W., Moerer P., Wiegant J., Liem L., Brouwer O. F. et al. (1991) Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics **9:** 560–565
6 Mathews K. D., Mills K. A., Bosch E. P., Ionasescu V. V., Wiles K. R., Buetow K. H. et al. (1992) Linkage localization of facioscapulohumeral muscular dystrophy (FSHD) in 4q35. Am. J. Hum. Genet. **51:** 428–431
7 Sarfarazi M., Wijmenga C., Upadhyaya M., Weiffenbach B., Hyser C., Mathews K. et al. (1992) Regional mapping of fa-

cioscapulohumeral muscular dystrophy gene on 4q35: combined analysis of an international consortium. Am. J. Hum. Genet. **51:** 396–403
8 Upadhyaya M., Lunt P., Sarfarazi M., Broadhead W., Farnham J. and Harper P. S. (1992) The mapping of chromosome 4q markers in relation to facioscapulohumeral muscular dystrophy (FSHD). Am. J. Hum. Genet. **51:** 404–410
9 Gilbert J. R., Stajich J. M., Speer M. C., Vance J. M., Stewart C. S., Yamaoka L. H. et al. (1992) Linkage studies in facioscapulohumeral muscular dystrophy (FSHD). Am. J. Hum. Genet. **51:** 424–427
10 Iqbal Z., Roper H., Pericak-Vance M. A., Hung W. Y., Delong R., Cummings W. J. et al. (1992) Genetic heterogeneity in facioscapulohumeral disease. Am. J. Hum. Genet. Suppl. **51:** A191
11 Gilbert J. R., Stajich J. M., Wall S., Carter S. C., Quitt H., Vance S. M. et al. (1993) Evidence for heterogeneity in facioscapulohumeral muscular dystrophy (FSHD). Am. J. Hum. Genet. **53:** 401–408
12 Bakker E., Wijmenga C., Vossen R. H., Padberg G. W., Hewitt J., van der Wielen M. et al. (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve **2:** S39–S44
13 Wijmenga C., Hewitt J. E., Sandkuijl L. A., Clark L. N., Wright T. J., Dauwerse H. G. et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat. Genet. **2:** 26–30
14 van Deutekom J. C., Wijmenga C., van Tienhoven E. A., Gruter A. M., Hewitt J. E., Padberg G. W. 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
15 Hewitt J. E., Lyle R., Clark L. N., Valleley E. M., Wright T. J., Wijmenga C. et al. (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum. Mol. Genet. **3:** 1287–1295
16 Winokur S. T., Bengtsson U., Feddersen J., Mathews K. D., Weiffenbach B., Bailey H. et al. (1994) 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 Res. **2:** 225–234
17 Lee J. H., Goto K., Matsuda C. and Arahata K. (1995) Characterization of a tandemly repeated 3.3-kb KpnI unit in the facioscapulohumeral muscular dystrophy (FSHD) gene region on chromosome 4q35. Muscle Nerve **2:** S6–S13
18 Meneveri R., Agresti A., Marazzi A., Saccone S., Rocchi M, Archidiacono N. et al. (1993) Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. Gene **123:** 227–234
19 Zhang X. Y., Loflin P. T., Gehrke C. W., Andrews P. A. and Erlich M. (1987) Hypermethylation of human DNA sequence in embryonic carcinoma cells and somatic tessues but not in sperm. Nucleic Acid Res. **15:** 9429–9449
20 Gehring W. J., Muller M., Affolter M., Percival-Smith A., Billeter M., Quian J. Q. et al. (1990) The structure of the homeodomain and its functional implications. Trends Genet. **6:** 323–329
21 Gabriels J., Beckers M. C., Ding H., De Vriese A., Plaisance S., van der Maarel S. M. et al. (1999) 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:** 25–32
22 Lyle R., Wright T. J., Clark L. N. and Hewitt J. E. (1995) The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics **28:** 389–397
23 Clark L. N., Koehler U., Ward D. C., Wienberg J. and Hewitt J. E. (1996) 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:** 180–189
24 Winokur S. T., Bengtsson U., Vargas J. C., Wasmuth J. J., Altherr M. R., Weiffenbach B. et al. (1996) The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. Hum. Mol. Genet. **5:** 1567–1565.
25 Lunt, P. W. (1998) 44th ENMC International Workshop: Facioscapulohumeral Muscular Dystrophy: Molecular Studies 19–21 July 1996, Naarden, The Netherlands. Neuromuscul. Disord. **8:** 126–130
26 Griggs R. C., Tawil R., Storvick D., Mendell J. R. and Altherr M. R. (1993) Genetics of facioscapulohumeral muscular dystrophy: new mutations in sporadic cases. Neurology **43:** 2369–2372
27 Padberg G. W., Brouwer O. F., de Keizer R. J., Dijkman G., Wijmenga C., Grote J. J. et al. (1995) On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. Muscle Nerve **2:** S73–S80
28 Zatz M., Marie S. K., Passos-Bueno M. R., Vainzof M., Campiotto S., Cerqueira A. et al. (1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. Am. J. Hum. Genet. **56:** 99–105
29 Weiffenbach B., Dubois J., Storvick D., Tawil R., Jacobsen S. J., Gilbert J. et al. (1993) Mapping the facioscapulohumeral muscular dystrophy gene is complicated by chromosome 4q35 recombination events. Nat. Genet. **4:** 165–169
30 Upadhyaya M., Osborn M., Maynard J., Altherr M., Ikeda J. and Harper P. S. (1995) Towards the finer mapping of facioscapulohumeral muscular dystrophy at 4q35: construction of a laser microdissection library. Am. J. Med. Genet. **60:** 244–251
31 Bakker E., Van der Wielen M. J., Voorhoeve E., Ippel P. F., Padberg G. W., Frants R. R. et al. (1996) Diagnostic, predictive and prenatal testing for facioscapulohumeral muscular dystrophy: diagnostic approach for sporadic and familial cases. J. Med. Genet. **33:** 29–35
32 van der Maarel S. M., Deidda G., Lemmers R. J., van Overveld P. G., van der Wielen M., Hewitt J. E. et al. (2000) De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. Am. J. Hum. Genet. **66:** 26–35
33 Deidda G., Cacurri S., Piazzo N. and Felicetti L. (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). J. Med. Genet. **33:** 361–365
34 van Deutekom J. C., Lemmers R. J., Grewal P. K., van Geel M., Romberg S., Dauwerse H. G. et al. (1996) Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. Hum. Mol. Genet. **5:** 581–590
35 van Overveld P. G., Lemmers R. J., Deidda G., Sandkuijl L., Padberg G. W., Frants R. R. et al. (2000) Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum. Mol. Genet. **9:** 2879–2884
36 Cacurri S., Piazzo N., Deidda G., Vigneti E., Galluzzi G., Colantoni L. et al. (1998) Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric Kpn repeat units implicated in facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. **63:** 181–190
37 Lemmers R. J., van der Maarel S. M., van Deutekom J. C., van der Wielen M. J ., Deidda G., Dauwerse H. G. et al. (1998) Inter- and intrachromosomal sub-telomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. Hum. Mol. Genet. **7:** 1207–1214
38 Lemmers R. J., de Kievit P., Sandkuijl L., Padberg G. W., van Ommen G. J., Frants R. R. et al. (2002) Facioscapulohumeral

muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat. Genet. **32:** 235–236
39 Bengtsson U., Altherr M. R., Wasmuth J. J. and Winokur S. T. (1994) High resolution fluorescence in situ hybridization to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. Hum. Mol. Genet. **3:** 1801–1805
40 Altherr M. R., Bengtsson U., Markovich R. P., Winokur S. T. (1995) Efforts toward understanding the molecular basis of facioscapulohumeral muscular dystrophy. Muscle Nerve **2:** S32–S38
41 van Deutekom J. C., Hofker M. H., Romberg S., van Geel M., Rommens J., Wright T. J. et al. (1995) Search for the FSHD gene using cDNA selection in a region spanning 100 kb on chromosome 4q35. Muscle Nerve **2:** S19–S26
42 Grewal P. K. , van Deutekom J. C., Mills K. A., Lemmers R. J., Mathews K. D., Frants R. R. et al. (1997) The mouse homolog of FRG1, a candidate gene for FSHD, maps proximal to the myodystrophy mutation on chromosome 8. Mamm. Genome **8:** 394–398
43 Grewal P. K., Todd L. C., van der Maarel S., Frants R. R. and Hewitt J. E. (1998) FRG1, a gene in the FSHD muscular dystrophy region on human chromosome 4q35, is highly conserved in vertebrates and invertebrates. Gene **216:** 13–19
44 Van Geel M., van Deutekom J. C., van Staalduinen A., Lemmers R. J., Dickson M. C., Hofker M. H. et al. (2000) Identification of a novel beta-tubulin subfamily with one member (TUBB4Q) located near the telomere of chromosome region 4q35. Cytogenet. Cell. Genet. **88:** 316–321
45 Flint J., Thomas K., Micklem G., Raynham H., Clarck K., Dogett N. A. et al. (1997) The relationship between chromosome structure and function at a human telomeric region. Nat. Genet. **15:** 252–256
46 Grewal P. K., van Geel M., Frants R. R., de Jong P. and Hewitt J. E. (1999) Recent amplification of the human FRG1 gene during primate evolution. Gene **227:** 79–88
47 Van Geel M. (2001) Facioscapulohumeral muscular dystrophy, a complex human disease entangled in dynamic molecular evolution. Thesis, Leiden University
48 Fisher J. and Upadhyaya M. (1997) Molecular genetics of facioscapulohumeral muscular dystrophy (FSHD). Neuromuscul. Disord. **7:** 55–62
49 van Deutekom J. C., Bakker E., Lemmers R. J., van der Wielen M. J., Bik E., Hofker M. H. et al. (1996) Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counseling and etiology of FSHD1. Hum. Mol. Genet. **5:** 1997–2003
50 Tupler R., Berardinelli A., Barbierato L., Frants R., Hewitt J. E., Lanzi G. et al. (1996) Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. J. Med. Genet. **33:** 366–370
51 Dorer D. R. and Henikoff S. (1994) Expansion of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. Cell **77:** 993–1002
52 Garrick D., Fiering S., Martin D. I. and Whitelaw E. (1998) Repeat-induced gene silencing in mammals. Nat. Genet. **18:** 56–59
53 Saveliev A., Everett C., Sharpe T., Webster Z. and Festenstein R. (2003) DNA triplet repeats mediate heterocromatin-protein-1-sensitive variegated gene silencing. Nature **422:** 909– 913
54 Tsien F., Sun B., Hopkins N. E., Vedanarayanan V., Figlewicz D., Winokur S. et al. (2001) Methylation of the FSHD syndrome-linked subtelomeric repeat in normal and FSHD cell cultures and tissues. Mol. Genet. Metab. **74:** 322–331
55 Tupler R., Perini G., Pellegrino M. A. and Green M. R. (1999) Profound misregulation of muscle specific gene expression in facioscapulohumeral muscular dystrophy. Proc. Natl. Acad. Sci. USA **96:** 12650–12654
56 Gabellini D., Green M. R. amd Tupler R. (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell **110:** 339–348
57 Thomas M. J. and Seto E. (1999) Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene **236:** 197–208
58 Yao Y. L., Yang W. M. and Seto E. (2001) Regulation of transcription factor YY1 by acetylation and deacetylation. Mol. Cell. Biol. **21:** 5979–5991
59 Yant S. R., Zhu W., Millinoff D., Slighton J. C., Goodman M. and Gunucio D. L. (1995) High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. Nucleic Acid Res. **23:** 4353–4362
60 Bustin M. (1999) Regulation of DNA-dependent activities by the functional motifs of the High-Mobility-Group chromosomal proteins. Mol. Cell. Biol. **19:** 5237–5246
61 Bianchi M. E. and Beltrame M. (2000) Upwardly mobile proteins. Workshop: the role of HMG proteins in chromatin structure, gene expression and neoplasia. EMBO Rep. **1:** 109–114
62 Agresti A. and Bianchi M. E. (2003) HMGB proteins and gene expression. Curr. Opin. Genet. Dev. **13:** 170–178
63 Zappavigna V., Falciola L., Helmer-Citterich M., Mavilio F. and Bianchi M. E. (1996) HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. EMBO J. **15:** 4981–4991
64 Boonyaratanakornkit V., Melvin V., Prendergast P., Altmann M., Ronfani L., Bianchi M. E. et al. (1998) High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. Mol. Cell. Biol. **18:** 4471–4487
65 Aidinis V., Bonaldi T., Beltrame M., Santagata S., Bianchi M. E. and Spanopoulou E. (1999) The RAG1 homedomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2. Mol. Cell. Biol. **19:** 6532–6542
66 Calogero S., Grassi F., Aguzzi A., Voigthlander T., Ferrier P., Ferrari S. et al. (1999) The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycoemia in new born mice. Nat. Genet. **22:** 276–280
67 Ronfani L., Ferraguti M., Croci l., Ovitt C. E., Scholer H. R. Consalez G. G. et al. (2001) Reduced fertility and spermatogenesis defects in mice lacking chromosomal protein Hmgb2. Development **128:** 1265–1273
68 Lehming N., Le Saux A., Shuller J. and Ptashme M. (1998) Chromatin components as part of a putative transcriptional repressing complex. Proc. Natl. Acad. Sci. USA **95:** 7322–7326
69 Ginisty H., Sicard H., Roger B. and Bouvet P. (1999) Structure and functions of nucleolin. J. Cell. Sci. **112:** 761–772
70 Gasser S.M. (2001) Positions of potential: nuclear organization and gene expression. Cell **104:** 639–642