Rearrangements within the facioscapulohumeral muscular dystrophy locus: mechanism, timing and consequences.

Lemmers, R.

Citation

Lemmers, R. (2005, June 15). *Rearrangements within the facioscapulohumeral muscular dystrophy locus: mechanism, timing and consequences*. Retrieved from https://hdl.handle.net/1887/2699

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/2699

Note: To cite this publication please use the final published version (if applicable).

1.5 Discussion

Already 10 years have passed since FSHD was linked to D4Z4 contractions on 4q35. Since then, many candidate genes in the 4q35 region have been identified and characterized, but yet none of them has been proven to be causative of FSHD. Although still controversial, recent studies have indicated that D4Z4 contractions may give rise to gene deregulation in *cis* and *trans*.[48,144] This thesis describes new developments in the molecular diagnosis of FSHD and the identification and characterization of unusual FSHD genotypes. Extensive analysis of parents and *de novo* FSHD patients resulted in the elucidation of the mechanism and timing of mitotic D4Z4 rearrangements. The 4q35 region most probably has been subjected to several subtelomeric duplications because many 4q35 (pseudo)genes close to D4Z4 have homologs on subtelomeres of many other chromosomes. Furthermore, other duplications have been identified between the subtelomeres of chromosomes 4q and 10q, and 4p and 4q. Probably, these duplications initiated the translocation of 10-type D4Z4 repeats to chromosome 4 and *vice versa*. 4;10 D4Z4 translocations are hampering the molecular diagnosis of FSHD but, additionally, they might contribute to the refinement of the definition of the FSHD allele and possibly the elucidation of the molecular mechanism of FSHD.

1.5.1 Molecular diagnosis FSHD

Soon after FSHD was linked to D4Z4 contractions on chromosome 4q35, a diagnostic test was developed. This method was based on Southern blotting of *Eco*RI-digested genomic DNA from mononuclear blood cells followed by hybridization with probe p13E-11.[138] In general, this method has proven to be very reliable and has not changed in the last decade. However, soon after this diagnostic test was developed, an almost identical D4Z4 locus was discovered on the subtelomere of chromosome 10q26.[5] Initially, discrimination of 4 and 10- derived D4Z4 repeats by *Bln*I digestion seemed to be adequate for consistent FSHD genotyping (Figure 10a).[22] However, further analyses revealed 4;10 translocations of D4Z4 that gave rise to 4-type D4Z4 repeats on chromosome 10, 10-type D4Z4 repeats on chromosome 4 and also hybrid D4Z4 repeats (Section 1.2.5). In addition, FSHD alleles have been detected in which the region proximal to the contracted D4Z4 repeat, including probe p13E-11, was deleted (Chapter 6).[59] Finally, it was shown that the region distal to D4Z4 displays a bi-allelic polymorphism in the population, one of which is associated with FSHD alleles (4qA) while the other (4qB) is not (Chapters 3 and 4).[57,63] All these atypical conditions are complicating molecular diagnosis of FSHD, since they do not reveal a standard short 4-type D4Z4 repeat but can cause, especially in diagnostic settings using linear electrophoresis, false negative or false positive diagnoses. In other cases, the FSHD allele is recognized, but the detection method fails to recognize mosaicism for the mutation, particularly when the proportion of affected cells is greater than 40% (Chapter 8).[61]

Approximately 5% of all chromosomes 4 carry 10-type D4Z4 repeat units. These translocated repeats on chromosome 4 are often hybrid (Section 1.2.5). When the most proximal repeat unit of an FSHD-sized hybrid repeat array is 10-type, this repeat will emerge as a 10-type repeat because it disappears upon double digestion with *Eco*RI and *Bln*I. Incorrectly, the patient will get a false negative FSHD diagnosis (Figure 10c).[15] Alternatively, a large hybrid repeat can give rise to a small residual *Eco*RI/*Bln*I fragment when the most proximal D4Z4 units are 4-type followed by one or more 10-type repeat units. When this residual *Eco*RI/*Bln*I fragment (with p13E-11) accidentally co-migrates with a small 10-type D4Z4 repeat, this can lead to a false-positive FSHD diagnosis, because the non-pathogenic 10-type repeat appears as an FSHD allele (Figure 10b, Section 1.4). We have demonstrated that the additional use of a *Xap*I digestion can avoid mis-diagnosis for both allele constitutions (Chapter 9).[15,58]

Additionally, in about 5% of the Dutch population, 4-type D4Z4 repeat arrays have been detected on chromosome 10q26. FSHD-sized translocated alleles on chromosome 10 have never been detected and it is unknown if these alleles are pathogenic. If these alleles are non-pathogenic, like the short 10-type repeats on chromosome 10, they could complicate the molecular diagnosis. The determination of the chromosomal origin of D4Z4 repeats is discussed in section 1.4.

Sequence analysis of the region distal to D4Z4 from different DNA-sources revealed a bi-allelic variation, 4qA and 4qB.[121] Both alleles are almost equally common in the control population, but FSHD alleles are only associated with the 4qA-type allele (Chapter 3).[57] Recently, we have shown that D4Z4 contractions on 4qB-type alleles do not cause FSHD (Chapter 4).[63] However, these alleles are indistinguishable from the FSHD-causing 4qA alleles, using p13E-11 on *Eco*RI and *Eco*RI/*Bln*I digested DNA, and can cause false positive DNA diagnoses (Figure 10b). This situation might explain the observations in a family with an unrelated neuromuscular disorder, where an FSHD-sized D4Z4 repeat was detected that did not co-segregate with the disease.[3] Therefore, FSHD-sized alleles in apparently non-affected individuals need to be typed, using *Hin*dIII digested DNA and allelotyping probes 4qA and 4qB (Section 1.4).

FSHD alleles with a deletion of probe p13E-11 are more difficult to recognize, since these alleles are not visible in the standard molecular diagnostic setting, using probe p13E-11 (Figure 10c). To address this problem, new diagnostic methods were developed that were based on the standard linear gel electrophoresis-based method. Briefly, by using *Hin*dIII-digested DNA in combination with distal probe 4qA, these alleles can be visualized (Chapter 6 and Section 1.4).[59] Furthermore, these deletions can be detected using the *Bgl*II/*Bln*I dosage test that determines the number of chromosome 4- and 10-type D4Z4 repeats (Chapter 1.4).[114]

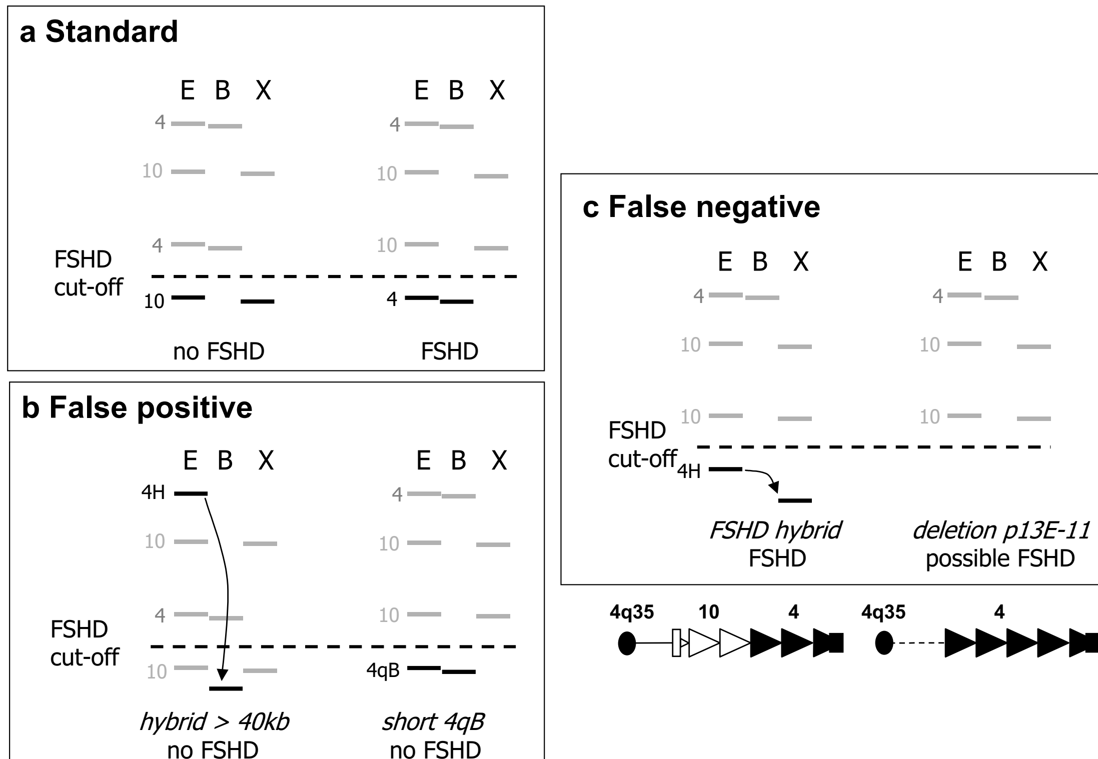

Figure 10 Summary of the most important pitfalls in the FSHD diagnostics. Each panel shows the D4Z4 genotyping using PFGE and the standard FSHD probe (p13E-11). Putative *Eco*RI/*Hin*dIII double digestion is depicted in lane E, *Eco*RI/*Bln*I double digestions (visualizing homogeneous chromosome 4 repeats arrays) in lane B and *Xap*I (visualizing homogeneous chromosome 10 repeats arrays) in lane X. In contrast to these illustrations, the D4Z4 repeat is highly polymorphic on chromosomes 4 and 10. The discussed D4Z4 allele in each panel is depicted in black and the less important D4Z4 alleles in gray. The dashed line represents the imaginary critical D4Z4 size for FSHD, which is positioned at 40 kb.

a) Standard genotypes of control (4-type D4Z4 > 40kb) and FSHD individuals (4-type D4Z4 < 40kb) are depicted in the upper left panel.

b) False positive genotypes that have been observed are depicted in the bottom left panel. Based on only lane E and B, the left genotype might be interpreted as having a hybrid FSHD allele. However, by the additional *Xap*I digestion it was shown that the short D4Z4 allele is in fact a homogeneous chromosome 10 repeat, not associated with FSHD. The residual fragment in lane B is derived from a > 40 kb hybrid repeat. The right genotype carries a homogeneous 4 repeat. However, because it resides on a 4qB chromosome it does not cause FSHD.

c) False negative genotypes that have been observed are depicted in the right panel. The left genotype has a hybrid FSHD allele, which could be demonstrated using the *Xap*I digestion: absence of homogeneous 10-type repeat in lane B. With respect to the p13E-11 probe the hybrid FSHD allele starts with 10-type D4Z4 units (composition of the hybrid FSHD allele is depicted below). The right genotype does not reveal an FSHD-sized repeat using linear gel electrophoresis. PFGE and the dosage test display only three D4Z4 alleles, which is suggestive of a p13E-11 deletion (composition of the p13E-11 deletion allele is depicted below). Additional tests should be performed to estimate the length of this D4Z4 repeat and the possible association with FSHD.

Unfortunately, recognition of D4Z4 mosaicism in FSHD patients appears to be almost impossible using LGE and necessitates the use of PFGE. In almost half of *de novo* FSHD cases, the D4Z4 contraction occurs during embryogenesis. Generally, an FSHD phenotype is presented when the proportion of the FSHD cells is larger than 40%, while <40% disease allele is mostly observed in non-patient carriers. Disease-presenting mosaicism was shown to account for half of these mitotic D4Z4 contractions.[115] Therefore, this phenomenon involves almost 25% of all *de novo* FSHD patients (Chapter 8).[62] The mosaic data suggest that mitotic D4Z4 contractions causing FSHD occur in the first few cell divisions after fertilization, which results in gonosomal mosaicism with a comparable proportion of FSHD cells in all tissues (Chapter 7).[61] If gonosomal mosaicism in FSHD patients is not recognized, the patient will be counseled as non-mosaic (FSHD-allele in all cells). The consequences of the failure to diagnose mosaicism are an overestimation of the recurrence risk for affected offspring and a possible underestimation of the phenotype of the mosaic patient's offspring, assuming similar dosage of FSHD cells in blood, muscle and germline cells.

Uncommon FSHD alleles obviously influence the specificity and sensitivity of the DNA test. As discussed in section 1.4, the specificity is defined as the proportion of unaffected individuals who have a negative test result, while the sensitivity is the proportion of FSHD patients with an FSHD allele. It is difficult to precisely estimate the occurrence of these atypical alleles and to which extent they hamper the molecular diagnosis of FSHD. A false positive FSHD test (specificity) can be caused by the interference of the residual *Eco*RI/*Bln*I fragment of a large hybrid repeat array in combination with a short 10-type repeat (Figure 10b). Additionally, FSHD-sized 4qB alleles can also mistakenly be interpreted as FSHD alleles. These two situations have been observed in 14 (hybrids on chromosome 4 or 10) and 6 (4qB) cases (2%) out of, in total, about 1000 genotypes.

To avoid false negative FSHD testing (Figure 10c) is even more important. Among our PFGE-analyzed patient group we have detected about 100 different FSHD alleles of which 4 initially were not recognized. Two of these FSHD alleles had a P13E-11 deletion and the other two were hybrid FSHD alleles that appeared as a short 10-type repeat. As discussed previously, the detection of mosaicism in FSHD patients presents a greater problem and requires the use of PFGE. As much as 20% of the *de novo* FSHD patients are somatic mosaic for the FSHD mutation, but in almost all of them mosaicism was not recognized using LGE (Chapter 8).[61]

Many methods have been developed, also for standard diagnostic setting with LGE, to elucidate the more complex FSHD genotypes. These methods have helped to improve both sensitivity and specificity of the FSHD molecular diagnosis, making them overall very reliable. The many requests to our laboratory for probes p13E-11, 4qA and 4qB, and the use of PFGE by more diagnostic laboratories are very promising and will improve the molecular diagnosis of FSHD worldwide.

1.5.2 Mitotic rearrangement of D4Z4 contractions

It has been estimated that about 10–30% of all FSHD cases are caused by a *de novo* D4Z4 contraction. In almost half of these cases the contraction occurred during embryogenesis in either patient or asymptomatic parent. It has been shown that mitotic D4Z4 rearrangements occur via an intrachromosomal gene conversion mechanism that is associated with a crossover in approximately 30% of the cases (Chapter 7).[62] Furthermore, the distribution of the *Pvu*II-RFLP on 4qA-type D4Z4 repeats suggests that rearrangements are polarized towards the distal site of the repeat (Section 1.3.7).

Rearrangement mechanism

In general rearrangements of mini- and macrosatellite repeats occur by gene conversion without crossover and only a few have been associated with a crossover (Section 1.3.2 and 1.3.3). In contrast, the subtelomeric D4Z4 repeat displays an unusually high frequency of gene conversion with crossover during early embryogenesis (Chapter 7).[62] We do not have an explanation for this phenomenon, but it might be related to the subtelomeric localization of D4Z4. It has been shown that human subtelomeres display elevated rates of mitotic recombination compared to interstitial chromosomal regions.[19] Possibly, the higher rearrangement frequency at telomeres is also associated with an increase of crossover-associated rearrangements.
For a further understanding of the high crossover rate in D4Z4, it might be interesting to compare the behavior of D4Z4 with other macrosatellite repeats, like DXZ4 and RS447 (Section 1.3.3). The composition and length of both repeats is very similar to D4Z4 and they also display mitotic instability.[33,34] However, since both DXZ4 and RS447 are located interstitial on chromosome Xq24 and chromosome 4p16.1, respectively, they might display different rearrangement features than D4Z4.

Timing of the mitotic D4Z4 rearrangement

Theoretically, unequal repeat rearrangements associated with crossovers occurring before the first zygotic cell division after fertilization will result in two cell populations of which one has a contracted and the other an expanded repeat (Figure 6b). This is exactly what we have observed in one of our mosaic FSHD patients. Parallel to this, gene conversion without crossover at this stage will result in two equally present cell populations of which one carries a rearranged D4Z4 repeat and the other the non-rearranged parental-sized repeat. This situation was observed in the majority of the mosaic FSHD patients (Figure 6b and 7a). Based on these results, it was hypothesized that D4Z4 rearrangements predominantly occur during first few zygotic divisions after fertilization (Section 1.3.5, Chapter 7).[62]
However, since only few cells of the early embryonic cells (32 and 64 cells stage) contribute to the embryo, the dosage of the *de novo* FSHD cells will also be determined by the contribution of the progeny cells (with and without D4Z4 contraction) to the embryo. That these stochastic

events contribute to the dosage of the FSHD cells is evident from the mosaic individuals from family 12 and Rf120 (Figure 7b), which display the mitotically contracted D4Z4 allele in 90% and 70% of their cells, respectively, which in theory could be maximally 50% if cell-distribution is equal. To further identify the role of these stochastic events, the dosage of FSHD cells after a mitotic D4Z4 rearrangement needs to be compared in different tissues of the same mosaic individual. Recently, the distribution of the different cell populations in a mosaic FSHD patient was compared in PBL and muscle tissue. The two different cell types displayed the FSHD allele in 16–35% of their cells, confirming a different contribution of *de novo* FSHD cells in different tissues, but otherwise cannot challenge the fact that D4Z4 contractions do indeed occur in early embryogenesis (Tonini *et al.*, manuscript in preparation).

1.5.3 Subtelomeric plasticity and the definition of the FSHD allele

As described for other subtelomeric loci, also the 4q35 region has been subjected to multiple telomeric duplications. All FSHD candidate genes in this region, *FRG1*[120], *TUBB4Q*[124], *FRG2*[94] and *DUX4*[30], underwent several duplication events and as a consequence they all belong to different gene families. Probably more recently, the most distal 4q35 region (D4S2463–4qter) was completely duplicated to the telomere of chromosome 10q.[121] Because of this, an almost identical D4Z4 repeat is found on 10q26, but contractions of this repeat have never been associated with FSHD.

According to these observations, the FSHD allele was defined as a 4-type D4Z4 repeat of 1–10 D4Z4 units located on chromosome 4. However, further research resulted in the identification of FSHD alleles with translocated homogeneous 10-type repeats and hybrid repeats.[15,116] Additional analyses have shown that all pathogenic alleles harbor a D4Z4 repeat of 1-10 units of which at least one unit is 4-type (*Xap*I sensitive and *Bln*I resistant), while pathogenic homogeneous 10-type repeats on chromosome 10 have not been identified in FSHD.[58]

Region distal to D4Z4

The region distal to D4Z4 appears to be important in the pathogenesis of FSHD, because only D4Z4 contractions on 4qA-type alleles, and not on 4qB-type alleles, are associated with the disease. Up to now we have no explanation for this observation, but telomeric sequencing and protein binding studies of this region are in progress (Section 1.6). About 15 kb of the region distal to D4Z4 has now been sequenced, and the main differences identified thus far between 4qA- and 4qB-type alleles are the 260 bp pLAM sequence directly distal to D4Z4 and the 6.2 kb-sized β-satellite repeat that are both present only on 4qA and 10q. There is some debate about the length of the region distal to D4Z4. Based upon digestions with *Not*I, we estimate this to be about 60 kb on both 4qA- and 4qB-type alleles, leaving 45 kb still to be analyzed.[59,60]

In the Dutch population many hybrid repeats have been detected on chromosome 4 (manuscript in preparation). In general, half of these hybrid repeats start with a 4- and end with a 10-type

repeat array in distal direction, while the other half starts with a 10- and have a 4-type repeat array at their distal end. In the latter group, distal to the 4-type D4Z4 repeats, either an A- or a B-type telomeric variation was detected. Most probably, the A-type hybrid alleles carry a 4qA, and not a 10qA, telomeric region distal to D4Z4 but until now we do not have experimental evidence for this. To date, only a few hybrid FSHD alleles have been reported.[15,116] It is important to further analyze the composition of these pathogenic hybrid alleles. If hybrid FSHD alleles with both 4qA and 10qA distal regions can be detected then this region is most probably identical on both chromosomes or not essential in FSHD pathogenesis. However, if only hybrid D4Z4 repeat alleles with a 4qA polymorphism are associated with FSHD, then a specific sequence in the 4qA distal region, different from 4qB and 10q, possibly must underly the 4qA-specific pathogenicity.

1.6 Future perspectives

Genetic research in the last decade has helped to increase our knowledge of the etiology of FSHD, the rearrangement mechanism, and has especially improved the molecular diagnostics for FSHD.
Analysis of mosaic individuals for the D4Z4 methylation of ancestral and contracted repeats, and the distribution of FSHD cells in different tissues, might support the observation that mitotic D4Z4 contractions occur during first few zygotic divisions after fertilization.
The PEV model as potential molecular mechanism was called into question based on the lack of evidence for a 4q35-specific transcriptional upregulation, and the lack of evidence for a gradient of hypoacetylated histone H4.[48,144] These observations promoted a whole-genome search for FSHD candidate genes, focusing on gene transcription but also on translation and metabolism, between well-defined FSHD patients, healthy controls and other myopathies.
The 4qA sequence distal to D4Z4 might play a role in the pathogenesis of FSHD. Consequently, it is important to completely elucidate the 4qA, 4qB and 10qA telomeric sequences to determine the differences between these sequences. Furthermore, the binding of specific proteins to the (sub)telomeric region of FSHD alleles has to be studied to elucidate their role in the localization to the nuclear periphery and the pathogenesis of FSHD.

1.6.1 Timing D4Z4 rearrangement

Analysis of mosaicism in gonadal cells

Analysis of 11 mosaic FSHD patients suggested that mitotic D4Z4 contractions often occur during the first zygotic divisions after fertilization. If the early timing of mitotic D4Z4 rearrangements applies for all mosaic persons, then all different tissues in the mosaic person will have approximately the same dosage of FSHD cells, when not considering the potential contribution of stochastic events and selection (Section 1.5.2, Chapter 7). Previously, mosaicism for the D4Z4 contraction was observed in PBL, muscle tissue and fibroblasts of a mosaic FSHD patient (Deidda *et al.*, unpublished results; Tonini *et al.*, manuscript in preparation). When an equal dosage of FSHD cells is observed in gonadal tissue and blood, the risk of affected offspring can be roughly estimated based on the proportion of the FSHD allele in the patient's blood. To compare D4Z4 mosaicism in gonadal tissue and blood, PFGE analyses have to be performed on high quality DNA from both sperm and blood of mosaic FSHD patients and carriers.
In about 60% of the new FSHD cases, the D4Z4 mutation is thought to occur meiotically, because both parents do not display contracted D4Z4 alleles in their PBL. These parents are proposed to have a low recurrence risk because it is assumed that only one of their gonadal cells carries the FSHD mutation. On the other hand, three cases have been described in which one of

the parents seems to display FSHD mosaicism restricted to the germline.[38,133] These carriers have more than one affected child with an identical FSHD allele suggesting a substantial proportion of FSHD gonadal cells. Apparently, in these individuals a mitotic D4Z4 rearrangement occurred in the germline leading to a significant recurrence risk. From two of these cases, Southern analyses were performed on the parents PBL-DNA. However, both the quality of DNA and blotting/hybridization are to some extent hampering the accurate analysis of these cases. Mitotic rearrangements that give rise to germline mosaicism have occurred after the blastocyst stage, while all D4Z4 rearrangements described in chapter 7 most likely took place before the blastocyst stage.
Possibly, germline mosaicism for the D4Z4 contraction occurs, but this phenomenon deserves further attention. Nowadays, families with more than three children are unusual in the Netherlands, making the recognition of this phenomenon rather difficult. For further elucidation of germline mosaicism, sperm DNA of fathers of non-mosaic *de novo* FSHD children needs to be analyzed and, if possible, previous cases that were suspected of germline mosaicism need to be reanalyzed.

D4Z4 methylation analysis

The CpG-methylation of the D4Z4 repeat has been studied using a Southern-based methylation assay that was focused on two sites in the most proximal unit of chromosome 4 D4Z4 repeats by using methylation-sensitive enzymes.[127] It was shown that FSHD-sized D4Z4 repeats display a significant lower methylation compared to normal sized D4Z4 repeats. In addition, a strong hypomethylation was observed in normal sized alleles of phenotypic FSHD patients, suggesting a possible role for D4Z4 methylation in the etiology of FSHD (Chapter 2).[127]
Strong hypomethylation of D4Z4 was first reported in ICF syndrome, an autosomal recessive disorder characterized by immunodeficiency, facial dysmorphologies and subtle developmental delay.[54] ICF is a very rare disorder and is caused by mutations in the *DNMT3B* gene, reducing its DNA methyltransferase activity.[149] ICF patients were shown to have even lower D4Z4 methylation than individuals with FSHD. Furthermore, hypomethylation was not restricted to the FSHD allele but was detected on D4Z4 repeats of both chromosome 4 and 10 (Chapter 2).[127]
In contrast to FSHD, strong D4Z4 hypomethylation in ICF does not cause a myopathic phenotype even in relatively old patients with an unusually mild ICF phenotype.
Active DNA methylation (methylation reprogramming) has been described to occur in the blastocyst stage (Figure 7a).[64,96] Mitotic D4Z4 rearrangement in mosaic FSHD patients and carriers are supposed to occur during the first few zygotic divisions after fertilization (Section 1.3.5). Consequently, the timing of the mitotic D4Z4 rearrangement can be confirmed by analyzing the methylation state of alleles that have been involved in mitotic rearrangements; we expect to observe a clear methylation difference between the mitotically contracted FSHD allele and the ancestral (normal sized) or expanded allele. Furthermore, D4Z4-hypomethylation of a

mitotic contracted FSHD allele is supported by the fact that individuals that carry these alleles in more than 40% of their cells usually display an FSHD phenotype (Chapter 8).[61]
The CpG-methylation of D4Z4 alleles that were involved in mitotic rearrangements can only be analyzed after the alleles have been separated by PFGE. Unfortunately, the isolation of DNA from individual D4Z4 alleles does not permit isolation of the large amounts of DNA needed for the Southern-based methylation analysis. Alternatively, CpG-methylation of single D4Z4 alleles can be analyzed using bisulfite sequencing. Bisulfite sequencing is a PCR-based method that enables the identification of methylated cytosines in genomic DNA after sodium bisulfite treatment.[40] Sodium bisulfite treatment converts all nonmethylated cytosines into thymines, while methylated cytosines remain unchanged. This method requires small amounts of genomic DNA and because there is no need to digest the DNA, every nucleotide of the genome is inspected. Therefore, bisulfite sequencing allows the comparative methylation analysis of mosaic parental and contracted alleles. Furthermore, it allows the examination of individual CpG-sites in hypomethylated ICF and FSHD alleles, which might explain the lack of a myopathic phenotype in ICF patients.

1.6.2 Interacting proteins

It was shown that 4qter localizes to the nuclear periphery and that lamin A/C is required for this localization (Section 1.1.4).[72] The nuclear envelope proteins lamin A/C, LAP2β and BAF have been shown to anchor chromatin to the inner nuclear envelope.[85] Furthermore, mutations in nuclear envelope proteins lamin A/C and emerin have been shown to disrupt the peripheral chromatin organization and cause other neuromuscular diseases, such as Emery-Dreifuss muscular dystrophy, Limb Girdle muscular dystrophy 1B and dilated cardiomyopathy diseases.[14,85] It was suggested that FSHD arises from improper interactions with transcription factors or chromatin modifiers at the nuclear envelope.[72] Together this makes the (sub)telomeric region an interesting target for binding studies with nuclear binding proteins. The D4S139 region (about 200 kb proximal to D4Z4) was shown to localize closer to the nuclear envelope than D4Z4, and is possibly involved in the association to the nuclear periphery.[72]
Another interesting target for protein binding studies is the β-satellite repeat, which is one of the most prominent differences between the telomeres of 4qA and 4qB, and is often associated with heterochromatin and telomeres. Intriguingly, the highly conserved nucleotide sequence GATCAGTGC within the β-satellite repeat unit that may serve as a protein-binding site is conserved on 4qA and 10q.[1] Binding of specific proteins to this sequence might explain the 4qA-specificity of the disease.
It has been suggested that a D4Z4 recognition complex (DRC), consisting of YY1, HMGB2 and nucleolin, binds to D4Z4 and represses 4q35-gene expression in healthy individuals.[29] Presumably, the DRC binds to D4Z4 repeat units at a 27 bp sequence, the D4Z4 binding element (DBE). The DBE contains a putative YY1 recognition sequence and YY1 (transcription

factor, involved in repressing and activating a number of promoters) is most probably the DRC component that directly binds to the DBE (Section 1.1.4). However, these binding experiments cannot explain the association of FSHD with D4Z4 contractions on only 4qA-type alleles and not on 4qB alleles, because the DBE is identical in 4qA- and 4qB-derived D4Z4 repeats (unpublished results).
Specific protein binding can be modulated by CpG-methylation of the binding site and consequently the significant hypomethylation observed for FSHD-causing D4Z4 repeats might alter the binding potential of putative protein-binding sites (Chapter 2).[127] As discussed in section 1.6.1, the analyses of specific protein-binding sites require the use of the PCR-based methylation analysis. The putative YY1-binding site in D4Z4 contains a single CpG dinucleotide. *In vitro* experiments for the imprinted *Peg3* mouse gene have shown that YY1 cannot bind to the YY1 binding site when a CpG in its bindingsite is methylated.[50] Possibly, CpG-methylation of the putative YY1 binding site in D4Z4 could interrupt the binding of DRC and influence the pathogenicy of a D4Z4 repeat.[29]
Another interesting sequence for binding and methylation analysis of D4Z4 is the putative CTCF-binding site that is located about 2 kb upstream of the *DUX4* sequence (Winokur, personal communication). The CCCTC-binding factor CTCF can mediate many different processes like gene activation, gene repression, gene silencing and chromatin insulation.[80] Possibly, CTCF binding to D4Z4 could play a role in the disease mechanism and make the CTCF binding site an interesting target in protein-binding studies. The binding of CTCF might be methylation sensitive as has been shown for the CTCF binding site in the imprinted locus of the Igf2 and H19 genes.[49] Consequently, specific hypomethylation of D4Z4 in FSHD alleles might influence the binding of CTCF to D4Z4, which could influence the expression of nearby genes or alter the chromatin structure by the recruiting of histone deacetylases.[70]

1.6.3 Transcription analysis D4Z4

The difference in homogeneity of translocated repeats on 4 and 10 (Section 1.2.5) can be explained by 1) for unknown reasons, 10-type D4Z4 repeats on chromosome 4 alleles are more susceptible to interchromosomal rearrangements with the 4-type D4Z4 repeat on the homologous chromosome 4, than the opposite allele constitution, i.e. translocated 4-type repeat on chromosome 10 with the 10-type repeat on the homologous chromosome 10. 2) Translocations of 10-type repeats to chromosome 4 have occurred much earlier in the evolution of D4Z4 than 4 to 10 translocations, allowing more time for interchromosomal rearrangements on the 4;10 translocated chromosome. 3) Possibly, all 4;10 and 10;4 translocated alleles originate from a few translocation founder alleles and the founder translocation breakpoint on chromosome 4 occurred within the D4Z4 repeat while on chromosome 10 it occurred within the homologous region proximal to the D4Z4 repeat. In this model, subsequent rearrangements of translocated D4Z4 repeats, like the non-translocated repeats, generally occur between sister-

chromatids and not between homologous chromosomes (Chapter 7).[62] 4) Finally, the difference in homogeneity can be explained if the absence of 4-type D4Z4 units on chromosome 4 is for unknown reasons not compatible with life.
As long as transcriptional activity from the D4Z4 repeat has not been excluded it remains essential to identify specific sequence differences between D4Z4 units derived from 4qA, 4qB and 10q repeats. Possibly specific polymorphisms in D4Z4 can explain that FSHD-sized homogeneous 4qB and 10-type repeat arrays on chromosome 10 are non-pathogenic.
Individual *Kpn*I units of D4Z4 repeats from chromosomes 4 and 10 have been sequenced from different individuals. Some of these sequences can be found in the Genbank-database and have been compared using the multiple sequence alignment program CLUSTALW. This alignment confirms that the 3.3 kb D4Z4 sequences from chromosomes 4 and 10 are 98% identical. Some typical differences can be found between D4Z4 sequences from chromosomes 4 and 10, e.g. *Xap*I on 4 and *Bln*I on 10. As described in chapter 7, we have identified a *Pvu*II polymorphism in D4Z4. Surprisingly, this polymorphic *Pvu*II-site was detected in approximately 30% of the most proximal unit of D4Z4 repeats on 4qB alleles, while it was almost completely absent in the proximal D4Z4 unit of 4qA alleles. Conversely, internal D4Z4 units of D4Z4 repeat arrays on 4qA and 4qB chromosomes were shown to be either *Pvu*II-sensitive or *Pvu*II-resistant independent of their chromosomal origin, which is suggestive that internal units are mixtures of 4qA- and 4qB-type D4Z4 units.[62] This finding indicates that the most proximal D4Z4 unit is the only region of the D4Z4 repeat array in which other allele specific differences between 4qA and 4qB chromosomes can be found.
As described in section 1.1.4, a putative gene was identified in D4Z4 comprising a 424 amino acid ORF for the double homeodomain protein *DUX4*. Sequence differences between chromosomes 4 and 10, were largely detected in the putative promotor of *DUX4*; the ORF of *DUX4* on both chromosomes 4- and 10-derived D4Z4 repeats remains intact. The *DUX4* sequence is not present on commercially available micro-arrays.[144] In addition, the detection of *DUX4*-transcription by RT-PCR is complicated due to a high background. This background is mainly caused by the fact that *DUX4* has no introns and is tandemly repeated. Consequently, discrimination cannot be made between processed *DUX4*-RNA and the abundantly present *DUX4*-DNA in the D4Z4 repeats. Also the homology to other members of the *DUX*-gene family complicates transcriptional analysis. Possibly due to these difficulties, *in vivo DUX4* transcription has not been demonstrated.[30,41,71,144] However, the fact that the *DUX4* protein was possibly detected in a myoblast extract from an FSHD patient and not in that of a control individual [18], is suggestive of an FSHD-specific *DUX4* transcription in myoblasts.
Initial transcriptional analysis of *DUX4* can be focused on monochromosomal cell hybrids containing single human chromosomes 4 or 10 or transgenic human D4Z4 mice, to minimize interference with homologous sequences. In these experiments, *DUX4* transcription from human chromosomes with different D4Z4 length, type and origin can be tested. Since pathogenic D4Z4

repeats display a significantly lower CpG-methylation than control alleles, *in vitro* demethylation of D4Z4 might induce *DUX4*-expression.[127] It has been shown *in vitro* that certain hypermethylated cancer genes can be transcriptionally reactivated by the DNA demethylating agent 5-aza-2′deozycytidine. Furthermore, a synergic upregulation of transcription was shown by DNA demethylation in combination with the histone deacetylation inhibition agent, trichostatin A.[16] Possibly, DNA demethylation combined with histone acetylation might induce *DUX4* transcription. In this perspective it is interesting that hypomethylation of the *DUX4* promotor has been detected in cultured human cancer cells in which two major DNA methyltransferases, *DNMT1* and *DNMT3b*, were mutated.[91]

1.6.4 Completing sequence 4qter and 10qter

As discussed in section 1.5.3, the 4qA region distal to D4Z4 might encompass a sequence that is important in the etiology of FSHD. Possibly this sequence is not present or mutated in the distal 4qB region and the 10qA region. Sequencing of all these telomeric regions might explain why FSHD is only associated with 4qA-type alleles. Therefore, it is important that the complete sequences of the telomeric regions of chromosomes 4qA, 4qB and 10q are elucidated. The region distal to D4Z4 on chromosome 4 has been estimated to be approximately 60 kb (unpublished results) of which only 15 kb has been sequenced.[121]
Additional sequencing projects will be focused on alleles that have a deletion of the region proximal to D4Z4 (Chapter 6)[59] and 4;10 and 10;4 translocated alleles (Section 1.2.5). These studies should reveal the deletion and translocation breakpoints, and possibly the mechanism by which these rearrangements occurred. Furthermore, these analyses will exactly determine the size of proximal deletions and might confirm that equally composed translocated alleles with different D4Z4 repeat numbers are derived from a single translocation founder allele, when they display identical breakpoints.

1.6.5 Phenotype-genotype study

The inter- and intrafamilial phenotypic variability observed in FSHD is an interesting phenomenon. Some patients are almost asymptomatic while others develop a more severe phenotype despite the fact that they carry equal-sized, or even identical FSHD-alleles.[112] Several explanations can be given for this phenomenon. The inter- and intrafamilial variability might just be a common difference in strength, also observed in the control population. In addition, some reports have suggested a role for anticipation in FSHD, which is defined as an earlier onset of the disease with increase in severity in subsequent generations.[105,152] Possibly, some cases that were initially suspected of anticipation can now be explained by mosaicism for the FSHD allele (Chapter 8).[61] Other cases of anticipation might be explained by ascertainment bias, the overrepresentation of relatively mildly affected mothers with affected offspring (Padberg, personal communication).

Further elucidation of the clinical variability requires large phenotype-genotype studies including different aspects, like the role of D4Z4 methylation (Chapter 2 and Section 1.6.2)[127] and the homologous chromosome. We have shown that FSHD patients with two FSHD-sized alleles display a more severe phenotype than individuals with only one of these alleles (Chapter 5).[145] Also, RNA, protein and metabolic profiling need to be included to explain the phenotypic differences (Section 1.6.6). Most interesting are large FSHD families that harbor asymptomatic as well as severely affected patients.

1.6.6 Genome-wide high throughput studies

For a long time the identification of the causative FSHD gene(s) was focused on genes in the 4q35 region, close to D4Z4. Recent genome-wide transcription analysis on muscle biopsies from FSHD patients and controls, showed a more prominent differential expression of *trans*-genes than on chromosome 4q35.[144] These experiments support a genome-wide candidate-gene approach. Up to now, global differential gene expression analyses were performed on RNA using microarrays and Northern blotting. Alternatively, FSHD may be associated with alterations in protein and metabolite levels.

Genome-wide analyses of protein and metabolite levels in tissue, blood and urine samples can be performed using for example liquid chromatography-tandem mass spectrometry (LC-TMS).[129] Using this method, many samples can rapidly be analyzed for numerous different proteins and metabolites. LC-TMS generates proteomic and metabolic patterns from different samples and enables the comparison of these patterns from, for example, healthy and affected individuals.[129] To perform LC-TMS for FSHD, blood plasma, urine and possibly muscle biopsies need to be collected from FSHD patients, healthy controls and patients suffering from other muscular dystrophies.

Apart from the FSHD candidate gene approach, these analyses might identify FSHD-specific RNA, protein or metabolite profiles. The use of these profiles would further automate, simplify and speed-up FSHD diagnostics. Possibly, these analyses would enable the identification of phenotypic FSHD patients without an FSHD-sized D4Z4 allele, if the RNA, protein or metabolite profiles of these patients display sufficient resemblance to the profiles of regular FSHD patients.

In general, analyses of differential expression require the use of a matched patient group, in order to reduce the background of individual variability. Unfortunately, even in a well-matched patient group, inter-individual non-FSHD-related expression differences might still disturb the analyses. Furthermore, a significant inter-individual variability can also be expected in the control groups. To tackle this problem, the use of affected and non-affected cell populations might be considered from mosaic individuals (Chapter 7).[62] These cell-populations are isogenic; FSHD and control cells only differ in the length of the D4Z4 repeat, making them perfect for expression studies. The different cell populations can be separated either by *in vitro* clonal

expansion or by cell sorting experiments. Clonal expansion requires the use of cells that can undergo multiple duplication steps. Therefore, probably only immortalized PBLs can be used for these experiments. For the cell sorting experiments we can use fluorescence-activated cell-sorting (FACS), which can be used on various single cell suspensions. However, this method requires an FSHD-specific marker.

Most suitable for these cloning experiments are PBL from mosaic individuals that display almost an equal proportion of affected and non-affected cells. Moreover, some D4Z4 mosaic individuals even display three cell populations enabling the study of expression differences caused by three different D4Z4 repeat lengths (Chapter 7).[62]

1.7 Summary of major findings

-FSHD alleles are associated with D4Z4 hypomethylation, which is suggestive of a (local) chromatin alteration on 4q35 in FSHD patients

-The subtelomeric variants of chromosome 4, 4qA and 4qB-type, are almost equally present in control individuals

-D4Z4 contractions on 4qA chromosomes are associated with FSHD, those on 4qB chromosomes are not

-Compound heterozygosity for two FSHD-sized D4Z4 repeats on 4qA alleles is not lethal, but may cause a phenotypic dosage effect

-D4Z4 contractions leading to FSHD occur almost equally frequently during meiosis and mitosis

-Mitotic D4Z4 rearrangements generally occur intrachromosomally via a gene conversion mechanism that is associated with crossovers in about 25% of the cases

-Mitotic D4Z4 rearrangements most likely occur in the first few zygotic divisions after fertilization, causing gonasomal mosaicism

-A marked linkage disequilibrium is present between the distal 4qA/4qB polymorphism and the most proximal D4Z4 unit of the D4Z4 repeat

-LGE is not appropriate for the detection of mosaicism for the D4Z4 contraction in FSHD patients, which comprises about 25% of all *de novo* FSHD cases

-The 55 kb region proximal to D4Z4 on chromosome 4q35, including p13E-11 and the genes *FRG2* and *DUX4c*, probably does not play a role in FSHD because it is deleted in several independent FSHD patients with a normal disease spectrum.

-At least 3% of the FSHD alleles carry a deletion of probe p13E-11and will not be detected in the regular DNA diagnostic for FSHD

-The triple DNA analysis with *Eco*RI, *Eco*RI/*Bln*I and *Xap*I is essential for the recognition of hybrid D4Z4 alleles, which can cause false positive or negative FSHD testing

1.8 References

1. Agresti,A. *et al.* Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family). *J. Mol. Biol.* **205**, 625-631 (1989).
2. Armour,J.A., Patel,I., Thein,S.L., Fey,M.F. & Jeffreys,A.J. Analysis of somatic mutations at human minisatellite loci in tumors and cell lines. *Genomics* **4**, 328-334 (1989).
3. Auer-Grumbach,M. *et al.* Clinical predominance of proximal upper limb weakness in CMT1A syndrome. *Muscle & Nerve* **23**, 1243-1249 (2000).
4. Bailey,J.A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003-1007 (2002).
5. Bakker,E. *et al.* The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* **2**, 39-44 (1995).
6. Boan,F. *et al.* Recombination analysis of the human minisatellite MsH42 suggests the existence of two distinct pathways for initiation and resolution of recombination at MsH42 in rat testes nuclear extracts. *Biochemistry* **41**, 2166-2176 (2002).
7. Bois,P., Collick,A., Brown,J. & Jeffreys,A.J. Human minisatellite MS32 (D1S8) displays somatic but not germline instability in transgenic mice. *Hum. Mol. Genet.* **6**, 1565-1571 (1997).
8. Boissonneault,G. Chromatin remodeling during spermiogenesis: a possible role for the transition proteins in DNA strand break repair. *FEBS letters* **514**, 111-114 (2002).
9. Brouwer,O.F., Padberg,G.W., Bakker,E., Wijmenga,C. & Frants,R.R. Early onset facioscapulohumeral muscular dystrophy. *Muscle Nerve* **Supplement**, S67-S72. (1995).
10. Brouwer,O.F. *et al.* Hearing loss in facioscapulohumeral muscular dystrophy. *Neurology* **41**, 1878-1881 (1991).
11. Buard,J., Collick,A., Brown,J. & Jeffreys,A.J. Somatic versus germline mutation processes at minisatellite CEB1 (D2S90) in humans and transgenic mice. *Genomics* **65**, 95-103 (2000).
12. Buck,L. & Axel,R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175-187 (1991).
13. Buck,L.B. The olfactory multigene family. *Curr. Opin. Genet. Dev.* **2**, 467-473 (1992).
14. Burke,B. & Stewart,C.L. Life at the edge: the nuclear envelope and human disease. *Nat. Rev. Mol. Cell Biol.* **3**, 575-585 (2002).
15. Buzhov,B.T. *et al.* Genetic confirmation of facioscapulohumeral muscular dystrophy in a case with complex D4Z4 rearrangments. *Hum Genet.* **116,** 262-266 (2005).
16. Cameron,E.E., Bachman,K.E., Myohanen,S., Herman,J.G. & Baylin,S.B. Synergy of demethylation and histone deacetylase inhibition in the re- expression of genes silenced in cancer. *Nat. Genet.* **21**, 103-107 (1999).
17. 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**, 180-189 (1996).

18. Coppee,F. *et al.* Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology. Upadhyaya,M. & Cooper,D.N. (eds.), pp. 117-134 (Garland Science/BIOS Scientific Publishers, Oxon, UK, 2004).
19. Cornforth,M.N. & Eberle,R.L. Termini of human chromosomes display elevated rates of mitotic recombination. *Mutagenesis* **16**, 85-89 (2001).
20. Cremer,T. & Cremer,C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301 (2001).
21. Croft,J.A. *et al.* Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell. Biol.* **145**, 1119-1131 (1999).
22. Deidda,G., Cacurri,S., Piazzo,N. & Felicetti,L. Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J. Med. Genet.* **33**, 361-365 (1996).
23. Elder,J.F., Jr. & Turner,B.J. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70**, 297-320 (1995).
24. Felice,K.J. & Moore,S.A. Unusual clinical presentations in patients harboring the facioscapulohumeral dystrophy 4q35 deletion. *Muscle Nerve* **24**, 352-356 (2001).
25. Felice,K.J., North,W.A., Moore,S.A. & Mathews,K.D. FSH dystrophy 4q35 deletion in patients presenting with facial-sparing scapular myopathy. *Neurology* **54**, 1927-1931 (2000).
26. Flint,J. *et al.* Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6**, 1305-1314 (1997).
27. Flint,J. *et al.* The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat. Genet.* **9**, 132-140 (1995).
28. Funakoshi,M., Goto,K. & Arahata,K. Epilepsy and mental retardation in a subset of early onset 4q35- facioscapulohumeral muscular dystrophy. *Neurology* **50**, 1791-1794 (1998).
29. Gabellini,D., Green,M. & Tupler,R. Inappropriate Gene Activation in FSHD. A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell* **110**, 339-348 (2002).
30. Gabriels,J. *et al.* 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 (1999).
31. Galluzzi,G. *et al.* Molecular analysis of 4q35 rearrangements in fascioscapulohumeral muscular dystrophy (FSHD): application to family studies for a correct genetic advice and a reliable prenatal diagnosis of the disease. *Neuromuscul. Disord.* **9**, 190-198 (1999).
32. Gerlich,D. *et al.* Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* **112**, 751-764 (2003).
33. Giacalone,J., Friedes,J. & Francke,U. A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat. Genet.* **1**, 137-143 (1992).
34. Gondo,Y. *et al.* Human megasatellite DNA RS447: copy-number polymorphisms and interspecies conservation. *Genomics* **54**, 39-49 (1998).

35. Grewal,P.K., Carim Todd,L., van der Maarel,S., Frants,R.R. & Hewitt,J.E. FRG1, a gene in the FSH muscular dystrophy region on human chromosome 4q35, is highly conserved in vertebrates and invertebrates. *Gene* **216**, 13-19 (1998).
36. Grewal,P.K., van Geel,M., Frants,R.R., de Jong,P. & Hewitt,J.E. Recent amplification of the human FRG1 gene during primate evolution. *Gene* **227**, 79-88 (1999).
37. Grewal,P.K. *et al.* The mouse homolog of FRG1, a candidate gene for FSHD, maps proximal to the myodystrophy mutation on Chromosome 8. *Mamm. Gen.* **8**, 394-398 (1997).
38. Griggs,R.C., Tawil,R., Storvick,D., Mendell,J.R. & Altherr,M.R. Genetics of facioscapulohumeral muscular dystrophy: new mutations in sporadic cases. *Neurology* **43**, 2369-2372 (1993).
39. Haaf,T. The battle of the sexes after fertilization: behaviour of paternal and maternal chromosomes in the early mammalian embryo. *Chromosome Res.* **9**, 263-271 (2001).
40. Herman,J.G., Graff,J.R., Myohanen,S., Nelkin,B.D. & Baylin,S.B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9821-9826 (1996).
41. Hewitt,J.E. *et al.* Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **3**, 1287-1295 (1994).
42. Hoglund,M., Mitelman,F. & Mandahl,N. A human 12p-derived cosmid hybridizing to subsets of human and chimpanzee telomeres. *Cytogenet. Cell. Genet.* **70**, 88-91 (1995).
43. Holliday,R. Mechanism for Gene Conversion in Fungi. *Genetical Research* **5**, 282-304 (1964).
44. Jeffreys,A.J. *et al.* Human minisatellites, repeat DNA instability and meiotic recombination. *Electrophoresis* **20**, 1665-1675 (1999).
45. Jeffreys,A.J. *et al.* Spontaneous and induced minisatellite instability. *Electrophoresis* **18**, 1501-1511 (1997).
46. Jeffreys,A.J., Neil,D.L. & Neumann,R. Repeat instability at human minisatellites arising from meiotic recombination. *EMBO Journal* **17**, 4147-4157 (1998).
47. Jeffreys,A.J. & Neumann,R. Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.* **6**, 129-132 (1997).
48. Jiang,G. *et al.* Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum. Mol. Genet.* **12**, 2909-2921 (2003).
49. Kanduri,C. *et al.* Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* **10**, 853-856 (2000).
50. Kim,J., Kollhoff,A., Bergmann,A. & Stubbs,L. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Hum. Mol. Genet.* **12**, 233-245 (2003).
51. Kirschner,M.W. Microtubule assembly and nucleation. *Int. Rev. Cytol.* **54**, 1-71 (1978).

52. Klaus,A.V., McCarrey,J.R., Farkas,A. & Ward,W.S. Changes in DNA loop domain structure during spermatogenesis and embryogenesis in the Syrian golden hamster. *Biol. Reprod.* **64**, 1297-1306 (2001).
53. Kohler,J., Rupilius,B., Otto,M., Bathke,K. & Koch,M.C. Germline mosaicism in 4q35 facioscapulohumeral muscular dystrophy (FSHD1A) occurring predominantly in oogenesis. *Hum. Genet.* **98**, 485-490 (1996).
54. Kondo,T. *et al.* Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. *Hum. Mol. Genet.* **9**, 597-604 (2000).
55. Lander,E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
56. Landouzy,L. & Dejerine,J. De la myopathy atrophique progressive. *Rev. Med.* **5**, 81-117 (1885).
57. Lemmers,R.J.L.F. *et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* **32**, 235-236 (2002).
58. Lemmers,R.J.L.F. *et al.* Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. *Ann. Neurol.* **50**, 816-9 (2001).
59. Lemmers,R.J.L.F. *et al.* D4F104S1 deletion in facioscapulohumeral muscular dystrophy: Phenotype, size, and detection. *Neurology* **61**, 178-183 (2003).
60. Lemmers,R.J.L.F. *et al.* Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum. Mol. Genet.* **7**, 1207-1214 (1998).
61. Lemmers,R.J.L.F. *et al.* Somatic mosaicism in FSHD often goes undetected. *Ann. Neurol.* **55**, 845-850 (2004).
62. Lemmers,R.J.L.F. *et al.* Mechanism and Timing of Mitotic Rearrangements in the Subtelomeric D4Z4 Repeat Involved in Facioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* **75**, 44-53 (2004).
63. Lemmers,R.J.L.F. *et al.* Contractions of D4Z4 on 4qB Subtelomeres Do Not Cause Facioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* **75**, 1124-1130 (2004).
64. Li,E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**, 662-673 (2002).
65. Liao,D. Concerted evolution: molecular mechanism and biological implications. *Am J Hum. Genet.* **64**, 24-30 (1999).
66. Linardopoulou,E. *et al.* Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum. Mol. Genet.* **10**, 2373-2383 (2001).
67. Louis,E.J., Naumova,E.S., Lee,A., Naumov,G. & Haber,J.E. The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**, 789-802 (1994).

68. Lunt,P.W. *et al.* 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). *Hum. Mol. Genet.* **4**, 951-958 (1995).
69. Lupski,J.R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**, 417-422 (1998).
70. Lutz,M. *et al.* Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* **28**, 1707-1713 (2000).
71. 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**, 389-397 (1995).
72. Masny,P.S. *et al.* Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum. Mol. Genet.* **13**, 1857-1871 (2004).
73. Matsumura,T. *et al.* Chromosome 4q;10q translocations; Comparison with different ethnic populations and FSHD patients. *BMC Neurol.* **2**, 7. (2002).
74. Mclay,D.W. & Clarke,H.J. Remodelling the paternal chromatin at fertilization in mammals. *Reproduction* **125**, 625-633 (2003).
75. Mefford,H.C. & Trask,B.J. The complex structure and dynamic evolution of human subtelomeres. *Nat. Rev. Genet.* **3**, 91-102 (2002).
76. Mills,K.A. *et al.* Genetic and physical mapping on chromosome 4 narrows the localization of the gene for facioscapulohumeral muscular dystrophy (FSHD). *Am. J. Hum. Genet.* **51**, 432-439 (1992).
77. Miura,K. *et al.* Two cases of chromosome 4q35-linked early onset facioscapulohumeral muscular dystrophy with mental retardation and epilepsy. *Neuropediatrics* **29**, 239-241 (1998).
78. Monfouilloux,S. *et al.* Recent human-specific spreading of a subtelomeric domain. *Genomics* **51**, 165-176 (1998).
79. Nassif,N., Penney,J., Pal,S., Engels,W.R. & Gloor,G.B. Efficient Copying of Nonhomologous Sequences from Ectopic Sites Via P-Element-Induced Gap Repair. *Molecular and Cellular Biology* **14**, 1613-1625 (1994).
80. Ohlsson,R., Renkawitz,R. & Lobanenkov,V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520-527 (2001).
81. Okada,T. *et al.* Unstable transmission of the RS447 human megasatellite tandem repetitive sequence that contains the USP17 deubiquitinating enzyme gene. *Human Genetics* **110**, 302-313 (2002).
82. Olckers,A. & Van der Merwe,A. Possible epigenetic protective factor for FSHD identified in the Black South African population. Am. J. Hum. Genet. (54th Annual Meeting) 217 (2004).
83. Orlando,V. Polycomb, epigenomes, and control of cell identity. Cell 112, 599-606 (2003).
84. Orr-Weaver,T.L. & Szostak,J.W. Yeast recombination: the association between double-strand gap repair and crossing-over. Proc. Natl. Acad. Sci. U.S.A. 80, 4417-4421 (1983).

85. Ostlund,C. & Worman,H.J. Nuclear envelope proteins and neuromuscular diseases. Muscle & Nerve 27, 393-406 (2003).
86. Padberg,G.W. Facioscapulohumeral disease. Thesis, Leiden University, Leiden The Netherlands (1982).
87. Padberg,G.W., Lunt,P.W., Koch,M. & Fardeau,M. Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul. Disord. 1, 231-234 (1991).
88. Paques,F. & Haber,J.E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews* **63**, 349-404 (1999).
89. Paques,F., Leung,W.Y. & Haber,J.E. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Molecular and Cellular Biology* **18**, 2045-2054 (1998).
90. Pavelitz,T., Rusche,L., Matera,A.G., Scharf,J.M. & Weiner,A.M. Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. *EMBO Journal* **14**, 169-177 (1995).
91. Paz,M.F. *et al.* Genetic unmasking of epigenetically silenced tumor suppressor genes in colon cancer cells deficient in DNA methyltransferases. *Hum. Mol. Genet.* **12**, 2209-2219 (2003).
92. Pirrotta,V. Transvection and chromosomal trans-interaction effects. *Biochim. Biophys. Acta.* **1424**, 1-8 (1999).
93. Richard,G.F. & Paques,F. Mini- and microsatellite expansions: the recombination connection. *EMBO Rep.* **1**, 122-126 (2000).
94. Rijkers,T. *et al.* FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J. Med. Genet.* **41**, 826-836 (2004).
95. Samonte,R.V. & Eichler,E.E. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **3**, 65-72 (2002).
96. Santos,F. & Dean,W. Epigenetic reprogramming during early development in mammals. *Reproduction.* **127**, 643-651 (2004).
97. Sarfarazi,M. *et al.* Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: combined analysis of an international consortium. *Am. J. Hum. Genet.* **51**, 396-403 (1992).
98. Sinden,R.R. Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am. J. Hum. Genet.* **64**, 346-353 (1999).
99. Spector,D.L., Fu,X.D. & Maniatis,T. Associations between distinct pre-mRNA splicing components and the cell nucleus. *EMBO Journal* **10**, 3467-3481 (1991).
100. Stankiewicz,P. & Lupski,J.R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74-82 (2002).
101. Stout,K. *et al.* Somatic pairing between subtelomeric regions: implications for human genetic disease? *Chrom. Res.* **7**, 323-329 (1999).
102. Szostak,J.W., Orrweaver,T.L. & Rothstein,R.J. The Double-Strand-Break Repair Model for Recombination. *Cell* **33**, 25-35 (1983).

103. Talbert,P.B. & Henikoff,S. A reexamination of spreading of position-effect variegation in the white-roughest region of Drosophila melanogaster. *Genetics* **154**, 259-272 (2000).
104. 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. *J. Cell. Biol.* **167**, 269-279 (2004).
105. Tawil,R. *et al.* Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann. Neurol.* **39**, 744-748 (1996).
106. Thomas,M.J. & Seto,E. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* **236**, 197-208 (1999).
107. Tonini,M.M.O. *et al.* Homozygosity for autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) does not result in a more severe phenotype. *Journal of Medical Genetics* **41**, (2004).
108. Trask,B.J. *et al.* Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7**, 13-26 (1998).
109. Tsien,F. *et al.* Methylation of the FSHD Syndrome-Linked Subtelomeric Repeat in Normal and FSHD Cell Cultures and Tissues. *Mol. Genet. Metab.* **74**, 322-331 (2001).
110. Tupler,R. *et al.* Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J. Med. Genet.* **33**, 366-370 (1996).
111. Tupler,R., Perini,G., Pellegrino,M.A. & Green,M.R. Profound misregulation of muscle-specific gene expression in facioscapulohumeral muscular dystrophy. *Proc. Natl. Acad. Sci. U S A* **96**, 12650-12654 (1999).
112. Upadhyaya,M., MacDonald,M. & Ravine,D. Prenatal diagnosis for facioscapulohumeral muscular dystrophy (FSHD). *Prenat. Diagn.* **19**, 959-965 (1999).
113. van der Kooi,A.J. *et al.* Extension of the clinical range of facioscapulohumeral dystrophy: report of six cases. *J Neurol Neurosurg Psychiatry* **69**, 114-116 (2000).
114. van der Maarel,S.M. *et al.* A new dosage test for subtelomeric 4;10 translocations improves conventional diagnosis of facioscapulohumeral muscular dystrophy (FSHD). *J. Med. Genet.* **36**, 823-828 (1999).
115. van der Maarel,S.M. *et al.* 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 (2000).
116. van Deutekom,J.C. *et al.* Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. *Hum. Mol. Genet.* **5**, 1997-2003 (1996).

117. van Deutekom,J.C. *et al.* 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 (1993).
118. van Deutekom,J.C.T. Towards the molecular mechanism of facioscapulohumeral muscular dystrophy. Thesis, Leiden University, Leiden The Netherlands (1996).
119. van Deutekom,J.C.T. *et al.* Search for the FSHD gene using cDNA selection in a region spanning 100 kb on chromosome 4q35. *Muscle Nerve* **Supplement**, S19-S26 (1995).
120. van Deutekom,J.C.T. *et al.* Identification of the first gene (*FRG1*) from the FSHD region on human chromosome 4q35. *Hum. Mol. Genet.* **5**, 581-590 (1996).
121. van Geel,M. *et al.* Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics* **79**, 210-217 (2002).
122. van Geel,M. *et al.* A Cascade of Complex Subtelomeric Duplications during the Evolution of the Hominoid and Old World Monkey Genomes. *Am. J. Hum. Genet.* **70**, 269-278 (2002).
123. van Geel,M. *et al.* The FSHD region on human chromosome 4q35 contains potential coding regions among pseudogenes and a high density of repeat elements. *Genomics* **61**, 55-65 (1999).
124. van Geel,M. *et al.* 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 (2000).
125. van Koningsbruggen,S. *et al.* FRG1P is localised in the nucleolus, Cajal bodies, and speckles. *Journal of Medical Genetics* **41**, (2004).
126. van Overveld,P.G. *et al.* Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Hum. Mol. Genet.* **9**, 2879-2884 (2000).
127. van Overveld,P.G. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* **35**, 315-317 (2003).
128. Vergnaud,G. *et al.* The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. *Genomics* **11**, 135-144 (1991).
129. Vogeser,M. Liquid chromatography-tandem mass spectrometry--application in the clinical laboratory. *Clin. Chem. Lab. Med.* **41**, 117-126 (2003).
130. Wakimoto,B.T. Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. *Cell* **93**, 321-324 (1998).
131. Weber,J.L. & Wong,C. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**, 1123-1128 (1993).
132. Weiffenbach,B. *et al.* Linkage analyses of five chromosome 4 markers localizes the facioscapulohumeral muscular dystrophy (FSHD) gene to distal 4q35. *Am. J. Hum. Genet.* **51**, 416-423 (1992).
133. Weiffenbach,B. *et al.* Mapping the facioscapulohumeral muscular dystrophy gene is complicated by chromsome 4q35 recombination events. *Nat. Genet.* **4**, 165-169 (1993).
134. Weiler,K.S. & Wakimoto,B.T. Heterochromatin and gene expression in Drosophila. *Annu. Rev. Genet.* **29**, 577-605 (1995).

135. Wijmenga,C., Brouwer,O.F., Padberg,G.W. & Frants,R.R. Transmission of de-novo mutation associated with facioscapulohumeral muscular dystrophy. *Lancet* **340**, 985-986 (1992).
136. Wijmenga,C. *et al.* Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* **336**, 651-653 (1990).
137. Wijmenga,C. *et al.* Molecular genetics of facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **3**, 487-491 (1993).
138. Wijmenga,C. *et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* **2**, 26-30 (1992).
139. Wijmenga,C. *et al.* Genetic linkage map of facioscapulohumeral muscular dystrophy and five polymorphic loci on chromosome 4q35-qter. *Am. J. Hum. Genet.* **51**, 411-415 (1992).
140. Wilkie,A.O. *et al.* Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64**, 595-606 (1991).
141. Winokur,S.T. *et al.* Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul. Disord.* **13**, 322-333 (2003).
142. Winokur,S.T. *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 Res.* **2**, 225-234 (1994).
143. Winokur,S.T., Bengtsson,U., Vargas,J.C., Wasmuth,J.J. & Altherr,M.R. 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**, 1567-1575 (1996).
144. Winokur,S.T. *et al.* Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum. Mol. Genet.* **12**, 2895-2907 (2003).
145. Wohlgemuth,M. *et al.* Possible phenotypic dosage effect in patients compound heterozygous for FSHD-sized 4q35 alleles. *Neurology* **61**, 909-913 (2003).
146. Wooster,R. *et al.* Instability of short tandem repeats (microsatellites) in human cancers. *Nat. Genet.* **6**, 152-156 (1994).
147. Wright,T.J. *et al.* Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. *Hum. Mol. Genet.* **2**, 1673-1678 (1993).
148. Wu,Z.Y., Wang,Z.Q., Murong,S.X. & Wang,N. FSHD in Chinese population: characteristics of translocation and genotype-phenotype correlation. *Neurology* **63**, 581-583 (2004).
149. Xu,G.L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187-191 (1999).
150. Young,J.M. & Trask,B.J. The sense of smell: genomics of vertebrate odorant receptors. *Hum. Mol. Genet.* **11**, 1153-1160 (2002).
151. Zatz,M. *et al.* The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. *Am. J. Med. Genet.* **77**, 155-161 (1998).
152. Zatz,M. *et al.* High proportion of new mutations and possible anticipation in Brazilian Facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **56**, 99-105 (1995).

153. Zhang,Y., Forner,J., Fournet,S. & Jeanpierre,M. Improved characterization of FSHD mutations. *Ann. Genet.* **44**, 105-110 (2001).

154. Zlotogora,J. Germ line mosaicism. *Hum. Genet.* **102**, 381-386 (1998).