Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Overveld, P.G.M. van

8 Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Discussion and future perspectives

Table of contents

The subtelomeric 4q35 locus associated with FSHD has a repetitive character that is in part due to a variable number of tandemly arranged D4Z4 repeat units. The number of repeat units varies between 11 and 100 copies in healthy individuals, while a contracted array with 1-10 copies carrying the 4qA variant distal to D4Z4 is associated with the clinical expression of the disease [5, 26, 66]. As a consequence of its subtelomeric location, this region is more prone to rearrangements than other parts of chromosome 4q. Furthermore, as a result of the repeat array contraction the chromatin conformation of the 4q35 region may change.

Even though the discovery of the D4Z4 repeat contraction in FSHD patients was published in 1992 [66], the exact mechanism that causes the FSHD phenotype is still not elucidated. Therefore, knowledge on both structure and behaviour of the 4q35 region will contribute to the disentangling of the defect that underlies FSHD. With regard to molecular behaviour, the objectives of this thesis were to focus on subtelomeric exchanges between the highly homologous repeat arrays on chromosomes 4 and 10 (***Chapters 2*** and ***3***) and on the occurrence of mosaicism for a contracted D4Z4 repeat with the consequences for FSHD (***Chapters 3*** and ***4***). To gain more knowledge about D4Z4 structure, we concentrated on DNA methylation and histone acetylation of the 4q35 region (***Chapters 5***, ***6*** and ***7***). In this chapter our major findings and data from other research groups will briefly be summarised to introduce a proper framework for discussion and future perspectives.

8.1 Subtelomeric exchanges between 4q35 and 10q26

In order to obtain more insight in the occurrence of translocated repeat arrays between chromosomes 4q and 10q and in repeat array size distributions we investigated these phenomena in a large Dutch control population and concomitantly elucidated some aspects of the dynamic behaviour of the subtelomeres of chromosome 4 and 10.

8.1.1 Exchange frequencies

8.1.1.1 Comparisons between different control and patient populations

The presence of repeat exchanges between chromosomes 4 and 10 has previously been demonstrated in small groups of Dutch FSHD patients and control individuals [4, 27]. In a more extensive follow-up study in a population of blood donors (***Chapter 2***) we observed 21% translocated alleles, either being homogeneous or heterogeneous, which was in agreement with earlier observations [4]. Of these translocated alleles, 9% was present on chromosome 4 and 12% on chromosome 10. Furthermore, chromosome 4-derived repeat arrays on chromosome 10 tended to be more homogeneous than chromosome 10-derived repeat arrays on chromosome 4.

Table 8.1 Subtelomeric exchanges between repeats on chromosomes 4q35 and 10q26 in control populations

Reference	Population	Individuals (*n*)	% Exchange	% 4 on 10	% 10 on 4
Van Deutekom *et al.* 1996	Dutch	50	20[a]	10	10
Van Overveld *et al.* 2000	Dutch	208	21[a]	12	9
Matsumura *et al.* 2002	Japanese	153	28	20	8
	Korean	124	30	26	4
	Chinese	114	19	14	5
Su *et al.* 2003	Chinese	52/107[b]	19	10	9
Wang *et al.* 2003	Chinese	50	16	8	8
Wu *et al.* 2004[c]	Chinese	110	19	9	10
Average %			**22**	**14**	**8**

4 on 10: Presence of chromosome 4-derived repeat units on chromosome 10. 10 on 4: Presence of chromosome 10-derived repeat units on chromosome 4. a: This population also contains individuals that carry hybrid repeat arrays. b: Published percentages could be based on control individuals only (n = 52) or on both controls and kindred with FSHD (n = 107). c: The population study described in this article had previously been published [68].

Table 8.2 Subtelomeric exchanges between repeats on chromosomes 4q35 and 10q26 in FSHD patients

Reference	FSHD patients[a]	Individuals (*n*)	% Exchange	% 4 on 10	% 10 on 4
van der Maarel *et al.* 2000	Dutch	8[b]	38	3	0
Matsumura *et al.* 2002	Japanese	56	32	23	9
Su *et al.* 2003	Chinese	70	19	13	6
Wang *et al.* 2003	Chinese	45	20	15	5
Wu *et al.* 2004	Chinese	55	17	15	2
Goto *et al.* 2004	Japanese	35	14	14	0
Average %			**23**	**14**	**4**

4 on 10: Presence of chromosome 4-derived repeat units on chromosome 10. 10 on 4: Presence of chromosome 10-derived repeat units on chromosome 4. a: Published frequencies are calculated based on results from patients with both *de novo* and inherited mutations. b: All exchanges were observed in mosaic FSHD patients.

After publication of our Dutch translocation data, the frequencies of subtelomeric translocations in several other healthy control populations were studied [39, 42, 57, 64, 67, 68]. All published data are summarised in ***Table 8.1***. The observed exchange frequencies between these repeat arrays of chromosomes 4q and 10q in the Asian populations were roughly similar to that of the Dutch population. Preliminary data recently presented by Olckers and colleagues [42] might indicate a remarkable difference in these frequencies between Eurasian and black South African populations. Individuals from the Eurasian population probably carry an excess of chromosome 4-derived repeats on chromosome 10, while the black South African population harbours an excess of chromosome 10-derived repeats on chromosome 4 [42].

Table 8.3 Subtelomeric exchanges between repeats on chromosomes 4q35 and 10q26 in *de novo* and familial FSHD patients.

Reference	FSHD patients	Individuals (*n*)	*de novo* (*n*)	Familial (*n*)	*de novo* (*n*)		Familial (*n*)	
					4 on 10	10 on 4	4 on 10	10 on 4
van der Maarel *et al.* 2000	Dutch	8[a]	8	0	3	0	0	0
Wang *et al.* 2003	Chinese	45	16	29	7	0	0	2
Wu *et al.* 2004	Chinese	55	16	39	6	0	2	1
Goto *et al.* 2004	Japanese	35	35	0	5	0	0	0
Total		**143**	**75**	**68**	**21**	**0**	**2**	**3**

4 on 10: Presence of chromosome 4-derived repeat units on chromosome 10. 10 on 4: Presence of chromosome 10-derived repeat units on chromosome 4. a: All exchanges were observed in mosaic FSHD patients.

When comparing data from FSHD patients from different ethnic origins [13, 38, 39, 57, 64, 67] (***Table 8.2***), exchanges of chromosome 4 repeat units to chromosome 10 were observed more frequently than for chromosome 10 repeats to chromosome 4, 14 versus 4% respectively. Four research groups also differentiated within their patient population between *de novo* and familial cases [13, 38, 64, 67] (***Chapter 4***). This grouping (***Table 8.3***) revealed that translocations of chromosome 4-derived repeat units to chromosome 10 were only observed in *de novo* patients.

8.1.1.2 Discussion and perspectives for future research

When all observed exchange frequencies of different populations were merged, translocations between the subtelomeres of chromosomes 4q and 10q were observed in 22% (***Table 8.1***). However, when each Asian study was compared separately to the Dutch population, a higher exchange frequency of chromosome 4 repeat units to chromosome 10 was seen in the studies Matsumura *et al.* performed on DNA samples from the Japanese, Korean and Chinese population [39]. With regard to the frequencies observed in the four Chinese studies, it is conceivable that within the large Chinese population specific ethnically differences occur and this may explain the discrepancy between the data from Matsumura *et al.* [39] and the other three studies. Preliminary data presented by Olckers *et al.* [42] from the Eurasian population also points towards an excess of chromosome 4-derived repeats on chromosome 10. Olckers and colleagues [42] furthermore reported preliminary data on the black South African population, where they observed an excess of chromosome 10-derived repeats on chromosome 4. Especially in combination with the observation that there are no *de novo* patients published with a translocated chromosome 10 repeat array on chromosome 4 [13, 38, 64, 67] (listed in ***Table 8.3***), the fact that no FSHD patient has yet been reported in the black South African population [42] might suggest that a chromosome 4 repeat array exchanged to chromosome 10

could predispose to the development of FSHD or even that a chromosome 10 repeat array exchanged to chromosome 4 may protect against disease.

Analysing translocation frequencies in various patient populations, including both sporadic and familial cases, is thus necessary to confirm if the observations made so far in patients with a *de novo* mutation is a consistent finding. It would also be important to separate datasets that actually contain subpopulations, since this provides new clues on the dynamics and evolution of the 4q35 and 10q26 repeat array configurations. For example, the different subpopulations from the Eurasian dataset should be separated and compared with published data to see if indeed Japanese, Korean and some regions in China carry an excess of chromosome 4-derived repeats on chromosome 10.

Most population studies were done using the *Bgl*II/*Bln*I dosage test that cannot detect the presence of hybrid repeat arrays consisting of repeat units derived from both chromosomes [37]. For both Dutch studies ([4]; ***Chapter 2***), we applied pulsed-field gel electrophoresis to identify all four fragments of chromosomes 4q and 10q, which also enabled the detection of hybrid fragments. When analysing these repeat arrays, we observed more homogeneous chromosome 4-derived repeats on chromosome 10 than chromosome 10-derived repeats on chromosome 4, which may indicate a biological difference between both repeat copies that results in a preference for homogeneous chromosome 4-derived repeats on chromosome 10. Repetitive sequences generally undergo concerted evolution, a process that homogenises repetitive DNA sequences, which occurs at a much higher rate within than between arrays [30, 31]. However, we observed many heterogeneous repeat arrays next to homogeneous repeat arrays, regardless of their chromosomal origin. This appears to suggest that D4Z4 repeats of chromosomes 4 and 10 in part escape concerted evolution and evolve via relatively unrestricted intrachromosomal and interchromosomal recombination, which might be initiated by the microsatellites and low copy repeat sequences within the D4Z4 repeat [1, 24, 25, 71].

Since concerted evolution normally ensures the integrity of repetitive sequences coding for genes, this may imply that the putative open reading frame, *DUX4*, present within each D4Z4 repeat unit, has lost its function during the expansion of the repeat array. Combining the observation of exchanged arrays with the findings that no expression of *DUX4* has been observed *in vivo*, haploinsufficiency of the 4q35 region does not cause FSHD [62] and that FSHD is only associated with a contracted D4Z4 on the 4qA telomeric variant [26], makes the model that D4Z4 repeat units harbour part or the whole FSHD gene unlikely. However, other models putting forward a role for the repeat itself are still compatible with the presence of homogeneous and heterogeneous repeat arrays. Since the repeats of chromosomes 4q and 10q are highly homologous, it is still feasible that the sequences of both arrays contain the same

binding sites for the multiprotein D4Z4 recognition complex observed by Gabellini *et al.* [11]. In this manner, only the number of copies would likely be important to generate the suggested suppression of gene expression [11]. It should be noted though that this model does not explain the 4q35- and 4qA-specificity of the disease. Finally, the configuration of the array is also not important for the looping model, proposed by Jiang and colleagues [19].

8.1.2 Size distributions of both repeat arrays

8.1.2.1 Analysis of 4q and 10q size distribution

To establish a normal unbiased allele size distribution, we grouped alleles with the standard pattern (i.e. disomic for D4Z4) in intervals of 20 kb. Because we cannot discriminate between the standard disomic configuration and the presence of double exchanged alleles, this distribution includes a very low frequency of apparent disomics (1% [44]; see also ***Chapter 1***; ***Figure 1.5.A***). It was unexpected to observe six healthy individuals with an FSHD-sized 4q35 repeat array in our control population. For the purpose of our study these repeat sizes, which varied between 25-35 kb, were excluded, because of their association with FSHD. In the resulting population we observed an unequal repeat size distribution on chromosomes 4q and 10q, presented as a multimodal pattern with a similar mixture of three normal distributions with equidistant means at intervals of approximately 65 kb (***Chapter 2***; ***Figure 2.1***).

Recently, it was demonstrated that the homology between chromosomes 4q and 10q also extends towards the telomere and that a polymorphic segment distal to D4Z4 exists in two variants, designated 4qA and 4qB [12, 26]. With this new information it was possible to study the D4Z4 size distribution of 4q and 10q alleles in more detail. We compared D4Z4 allele sizes of the 4qA and 4qB repeat arrays in a set of control individuals partially overlapping with the dataset tested in ***Chapter 2*** and observed that 4qA alleles were significantly longer than 4qB alleles and that the sizes of alleles carrying one of these variants were also different (***Chapter 3***). Here too, 4q alleles were significantly longer than 10q alleles, but this was solely caused by alleles that carried the 4qA variant. 4qB and 10q alleles were not different in size.

8.1.2.2 Discussion and perspectives for future research

Since we observed six individuals that carry FSHD-sized repeat arrays in our population of blood donors, we hybridised their DNA with distal probes 4qA and 4qB [12, 26] and observed an equal and striking distribution among these contracted alleles: the three shortest alleles were of the 4qB variant, while the upper FSHD-sized fragments were of the 4qA variant. Individuals with a contracted allele of the 4qB variant do not suffer from FSHD [28], but this still did not explain the other three healthy individuals carrying contractions indicative of FSHD. However,

4qA fragments such as theirs, which are borderline for FSHD diagnosis, are usually more often observed in non-penetrant gene carriers (approximately 30% of all individuals in large families) [35, 47, 51]. These individuals with short chromosome 4qA fragments might thus simply be asymptomatic individuals, but they could also still be extremely mild cases of FSHD that will, as a result, never visit a neurologist.

These findings also imply that probably more individuals in the Dutch population carry a borderline-sized repeat array on chromosome 4, putting individuals in these pedigrees at risk for FSHD. Since overlap between D4Z4 fragment size in control individuals and FSHD patients seems to exist, it is therefore necessary to determine the penetrance for these 4qA fragments nowadays classified as borderline for FSHD and, when possible, to redefine a proper molecular threshold value for these 4q35 alleles associated with FSHD.

A similar multimodal allele distribution as that of 4q and 10q has also been observed for unstable minisatellite repeats [18, 52, 54]. In Caucasian populations, bimodal size distributions have been described for the insulin minisatellite and D2S44, whereas trimodal size distributions were seen at D19S20 and MS51 (i.e. D11S97) [54]. These low variability repeats often showed discontinuous allele size distributions probably as a result of genetic drift compounded by low mutation rates and coupled with mutation processes, like slippage or interallelic gene conversion, that display a bias towards specific allele lengths or even show a selection for certain sizes [31, 32, 54]. However, it seems questionable whether a similar mechanism operates at the 4q or 10q subtelomere.

Rather than the mechanism controlling the multimodal distribution of minisatellites we hypothesized that the distribution observed for chromosome 4 and 10 repeat arrays could be a reflection of the local chromatin structure. In a study to elucidate the folding of DNA in higher chromatin structures, Filipski *et al.* [9] discovered two periodicities in chromatin that were approximately 50 and 300 kb. The ~50 kb fragment probably reflects the loop organisation of the chromatin into hexameric rosettes, whereas the 300 kb fragment might reflect the total rosette arrangement of these smaller 50 kb loops into a higher-order chromatin structure. For our two regions on 4q and 10q, peaks with intervals of ~65 kb, could also reflect a loop organisation that might impose restrictions on the repeat size. Furthermore, because usually the shortest allele rearranged to a FSHD allele (***Chapter 4***), this suggested that the unequal distribution on chromosome 4 separated the repeat arrays into a premutation domain in the first peak and a normal domain in the larger peaks. As a result, alleles in the premutation domain might be more prone to contraction than longer alleles, thus predisposing to the development of FSHD-sized alleles.

After the detection of the 4qA and 4qB variant [12], it became possible to determine the size distributions of repeat arrays associated with 4qA or 4qB. The size distribution of the 4qB alleles was similar to that of 10q alleles, but both did differ significantly from that of the 4qA allele size distribution. The evolutionary process of both chromosomes 4 and 10 may explain these findings. Van Geel *et al.* [12] evaluated the homology of the two 4q variants and suggested that an ancestral duplication event involving 4p and 4q sequences first created the 4qA allele. Then additional events distributed this copy with proximal flanking sequences to other chromosomes [12]. Not just the distal part of chromosome 4q was copied onto 10q, but also the D4Z4 repeat and part of the proximal sequence. The 4qB allele was suggested to be the result of a second, more recent duplication of 4p sequences that probably arose through recombination of highly homologous subtelomeric regions [12]. This implied that the 4qA allele is older than both the 4qB and 10q allele. The observed differences in size between 4qA and 4qB might thus suggest that 4qB alleles have not yet reached the same size distributions as the 4qA alleles. By the same token the shorter chromosome 10q alleles fit with the same evolutionary sequence of events.

Alternatively, chromosome 10 repeats could be shorter than chromosome 4 repeats, because chromosome size might be of influence on subtelomere length. It has been suggested that a significant amount of non-telomeric sequences contribute to the size of telomeres by a proposed existence of a 2-4 kb subtelomeric DNA region, termed the X-region, which is resistant to enzymatic digestion [56]. The X-region presents a variable domain, whose size changes as a function of telomere length [56]. Suda and colleagues [58] showed that relative telomeric repeat content (RTC) values are significantly correlated to chromosome size: in the human genome chromosome 10 is smaller than chromosome 4 (125 Mb vs. 193 Mb, respectively) and the telomere size of chromosome 10 is approximately 3.7 kb with an RTC of 12.85, while the chromosome 4 telomere has a size of 5 kb with an RTC of 22.36 [58]. The telomeric structure provides protection against recombination, imposes positional effects on adjacent gene expression and releases key proteins for double strand break repair. This may indicate that longer telomeres are important for genome organisation and function and suggests that there is a lower and upper limit for telomere length [58]. Therefore, it is possible that the DNA base modifications that form the X-region are involved in preventing telomere recombination. This prevention subsequently results in the unusual pattern of exchanges seen between telomeres, up to 50 kb into the subtelomeric region [40], supported by high levels of recombination in subtelomeres [53]. The size of the subtelomere might therefore play a role in the integrity of the telomeres and might be correlated with telomere length for proper functioning. This thus provides an alternative explanation for the difference in size between the chromosome 10 subtelomere and the one on chromosome 4. However, this hypothesis does not explain the size difference observed for 4qA and 4qB.

One may assume that the size differences of 4q and 10q in our first dataset (***Chapter 2***) are most probably also the result of the longer 4qA alleles. However, a partially different group of control subjects was tested in the second study (***Chapter 3***), so to obtain absolute certainty whether repeats in the first dataset will show the same size differences of 4qA and 4qB, DNA from these individuals should be (re)tested with the A and B probes.

8.2 Mosaicism for the FSHD-associated region

The consequences of a mutation that gives rise to two or more different cell populations are dependent on the nature and location of the mutation, the cell types, and/or tissues that are involved [15]. Also timing could be crucial in the effect the mutation will have on phenotypic expression, since mutations that occur early in embryonic development involve both the germline and somatic cells [14]. We investigated the percentage of mosaicism for the D4Z4 repeat in FSHD patients and their parents to gain more insight into the mechanism by which mosaic alleles may arise (***Chapters 3*** and ***4***).

8.2.1 Detection of mosaicism and the relevance for FSHD development

After determining the exchange frequencies between the repeats on chromosomes 4q and 10q in a healthy population (***Chapter 2***), our laboratory focussed on repeat array configurations of *de novo* patients and their parents (***Chapters 3*** and ***4***). In 38 Dutch FSHD families we showed that somatic mosaicism for the FSHD-associated region was detectable in 17 individuals, either in the patient or an asymptomatic parent. This mosaicism for a contracted D4Z4 repeat, generating two genetically distinct cell populations, was more often seen in male than in female patients (9 versus 3) and comparing male and female FSHD patients revealed a higher proportion of cells with the contracted allele in female patients. Furthermore, since the repeat array configuration of most healthy individuals in the Dutch population displays a standard disomic pattern and only 20% of individuals carry translocated repeat arrays [4, 44], it was striking to observe an exchanged chromosome 4 repeat array on chromosome 10 in 6 mosaic cases, while not a single mosaic individual had the reverse configuration. Finally, we identified three FSHD patients with even more complex rearrangements resulting in three genetically distinct cell populations. In one patient, the original allele was shortened into two smaller alleles, one of which was FSHD-sized and caused disease. In the other two individuals, the original allele was mutated to a short FSHD-sized allele in some cells, but expanded to an allele larger that the original parental one in others.

8.2.2 Discussion and perspectives for future research

The observation that the proportion of cells carrying the contracted allele was higher in female *de novo* patients when compared to male *de novo* patients might indicate that females have a higher clinical tolerance for mosaic disease alleles than males and may thus explain the differences amongst *de novo* mosaic cases observed in our studies (9 males versus 3 females [28, 38] see ***Chapter 1***; ***Table 1.1***) and are consistent with the findings that males are usually more severely affected than females [34, 45, 46, 69] and that there is, on average, a female excess among unaffected parents with mosaicism for the FSHD-associated region (***Chapter 1***; ***Table 1.1***). This indicates that there might be more apparently healthy women mosaic for a contacted D4Z4 repeat, who will, if this mosaicism is present in the germline, have an increased risk of having a child that develops FSHD.

We determined the repeat array configuration in the observed 13 mosaic individuals, both healthy parents and patients. Next to 7 individuals with a standard pattern (disomic for D4Z4), this also revealed the presence of translocated chromosome 4 repeats on chromosome 10 in 46% of cases (n= 6). In the Dutch control population, this type of repeat array configuration is also present, but only in 10% of individuals [4, 44]. This high percentage is comparable with earlier observations made in non-mosaic *de novo* FSHD patients summarised in ***Table 8.3***, where a translocated chromosome 10 repeat array on chromosome 4 was never identified. This emphasizes again that an extra chromosome 4-dervived repeat array may predispose to the development of FSHD. Since we have only analysed 13 individuals, the frequency of an extra chromosome 4 repeat array should in our view be investigated in a larger sample set, for example in all mosaic cases that have already been reported (summarised in ***Chapter 1***; ***Table 1.1***).

The mechanism that generally leads to mosaicism for D4Z4 is termed gene conversion, a non-reciprocal recombination process between similar sequences that results in an alteration of a particular sequence to a homologous sequence during meiosis or mitosis [2, 30]. When an allele is damaged due to a double strand break (for example in or near the repeats on chromosome 4q or 10q), the sequence of the sister chromatid or of another homologous chromosome may be used as a template to repair the damage. This copying of the chromosome sequence can start on any arbitrary, but homologous point, potentially resulting in expansion or contraction of the involved chromosome sequences without affecting the integrity of the donor sequence. Gene conversion can be explained by a double strand break repair model that results in crossover and non-crossover products and by a synthesis-dependent strand annealing model that also allows for crossovers (reviewed in [50]). As a consequence of unequal recombination during mitosis,

gene conversion without crossover will, in the case of D4Z4 repeat rearrangements, eventually lead to two cell populations with a single rearranged allele for chromosome 4 or 10 in the majority of FSHD cases. The three identified FSHD patients with even more complex rearrangements are indicative of gene conversion with an unequal crossover that results in contraction and expansion.

The co-existence of mosaicism for the FSHD-associated region in lymphocytes, fibroblasts and germ cells indicates that these mitotic rearrangements can occur early in embryogenesis. The proportion of affected cells in any tissue depends on the timing of the rearrangement, on selection processes and on stochastic events related to cells that contribute to the development of the embryo. Since molecular FSHD analysis is usually restricted to the analysis of DNA from blood lymphocytes due to paucity of skin or muscle tissue, (a low level of) mosaicism may go unnoticed. As the degree of mosaicism may differ between tissues or may even be restricted to a specific tissue alone, the estimated frequency of mosaicism will then always be underestimated. Therefore, the examination of several tissues and cell types is necessary to improve the ability to detect mosaicism for the contracted D4Z4 repeat, to provide correct information in genetic counselling and to gain more insight into mosaicism and its consequences for the FSHD phenotype.

As we have identified individuals mosaic for the FSHD-associated region it may now be possible to isolate genetically identical cell lines differing solely in D4Z4 repeat array size, provided that the two (or three) cell populations within a mosaic culture derived from one individual can be separated. Such cell lines will allow researchers to study gene expression patterns that depend specifically on differences in repeat array size and tissue type in more detail. When used for specific biochemical assays, like chromatin immunoprecipitation, such cell lines may eventually clarify the consequences of D4Z4 repeat size for the FSHD phenotype.

8.3 Epigenetic modifications of the 4q35 region

Due to its repetitive nature and location on the subtelomere involved in FSHD, D4Z4 is proposed to have a heterochromatic character and several models suggested that loss of this heterochromatic structure as a result of contraction of the D4Z4 repeat array would cause activation of genes on 4q35. However, until recently, no studies were performed to test these models. Additionally, no information was available on possible changes of chromatin structure

in FSHD patients. While we were setting up a method to detect DNA methylation patterns in D4Z4, a publication of Kondo *et al.* [23] provided a first clue on D4Z4 chromatin structure.

8.3.1 D4Z4 and the link with the ICF syndrome

Kondo and colleagues [23] studied the ICF syndrome (reviewed in [7]). Mutations in the *DNMT3B* gene encoding a *de novo* DNA methyltransferase cause this rare recessive disease characterised by a phenotype with variable immunodeficiency, facial anomalies, psychomotor retardation and, less frequently, mental retardation and digestive disorders. The genome of these patients shows instability of heterochromatic regions located in the vicinity of centromeres (i.e. juxtacentromeric heterochromatin), which results in stretching, breakage and rearrangements of this heterochromatin of chromosomes 1, 9 and 16 including high frequencies of occurrence of multibranched chromosomes. In addition, several specific DNA sequences were reported to be hypomethylated, probably as a consequence of the reduced or even absent DNMT3B activity. These sequences are satellite DNA from the juxtacentromeric heterochromatin of chromosomes 1, 9 and 16, *Alu* repeats, sequences on the X- and Y-chromosome, H19 and α-satellite DNA [7, 23, 65].

Kondo *et al.* [23] analysed DNA methylation patterns of CpG islands in ICF patients and confirmed that aberrant DNA methylation patterns were restricted to a small percentage of the genome. They observed ICF-specific DNA hypomethylation of two unrelated repeats present on several locations in the genome: a 1.4 kb repeat named NBL2, located at juxtacentromeric regions of chromosome 9 as well as the acrocentric chromosomes 13, 14 and 21 [59], and D4Z4 located on 4q35.

8.3.2 DNA methylation patterns of D4Z4

8.3.2.1 FSHD-associated hypomethylation of D4Z4

After detecting DNA hypomethylation of the D4Z4 repeat in ICF patients, this group also investigated the DNA methylation status of D4Z4 in FSHD by methylation-sensitive restriction enzyme analysis [61]. They observed that in healthy individuals D4Z4 repeat arrays were highly methylated in tissues, like brain, lung, liver and spleen, but hypomethylated in sperm. Furthermore, FSHD samples of different tissue origin displayed no DNA methylation changes, indicating a lack of association between FSHD and the methylation status of D4Z4 repeats in the tissues tested [61].

DNA methylation assays, based on digestion of DNA with CpG methylation-sensitive enzymes and hybridisation with probes adjacent to the restriction site of interest, are very suitable for a first investigation of methylation patterns in a specific region. However, owing to the spreading of D4Z4 repeats over the entire genome, it has proven difficult to analyse DNA

methylation of this repeat by Southern blot-based approaches. We modified our dosage test [37], which enabled us to determine the DNA methylation pattern within the proximal D4Z4 repeat unit on chromosome 4 without the interference of chromosome 10 repeat units or any other region in the genome. Using this method we observed a significant DNA hypomethylation of D4Z4 in FSHD patients, but within this group a further internal difference was displayed, as FSHD patients disomic for D4Z4 displayed higher methylation values than monosomic patients (***Chapter 5***). Whereas in disomic individuals only simultaneous analysis of both chromosomes 4 is possible, yielding an average methylation of both alleles, in monosomic individuals the methylation of a single chromosome 4 can be determined. When corrected for the presence of a healthy and a disease allele in disomic patients, the observed hypomethylation could be attributed to the FSHD-sized allele. We subsequently confirmed that the observed DNA methylation within the proximal repeat unit is representative for the whole repeat array. Furthermore, the D4Z4 methylation pattern is FSHD-specific (thus not general for muscular dystrophies), is established early in development and transmitted stably through life (***Chapter 5***). Surprisingly, phenotypic FSHD patients, who are clinically indistinguishable from FSHD patients, but lack a contracted allele, also showed a pronounced DNA hypomethylation of the D4Z4 repeat (***Chapter 5***). In contrast to FSHD patients with a 4q35 contraction where hypomethylation is restricted to the disease allele, in these patients both chromosome 4 alleles are hypomethylated. Non-penetrant gene carriers, i.e. clinically unaffected individuals with a contracted D4Z4 repeat array, showed similarly hypomethylated repeats as FSHD patients with a 4q35 contraction (***Chapter 5***).

Because DNA from monosomic individuals provides a unique opportunity to examine DNA methylation of a single healthy or disease chromosome 4 repeat array, we took advantage of this phenomenon to gather more information on a possible relationship of DNA hypomethylation observed in FSHD patients with repeat array size and severity of disease (***Chapter 6***). Indeed, the methylation status changed in relation to the shortening of the repeat array, but this method did not reveal a linear relationship of DNA methylation with residual repeat length or disease severity. However, our data suggested a marked difference in methylation status of the D4Z4 repeat when we separated the patients into two groups: a 10-20 kb and a 20-38 kb interval.

8.3.2.2 Discussion and perspectives for future research

Our data implied that there are significant changes in DNA methylation associated with contraction of the repeat array on chromosome 4. Therefore, this does not provide evidence for the model that suggests spreading of telomeric heterochromatin in a proximal direction upon contraction of the D4Z4 array. However, our findings were different from those published by

Tsien *et al.* [61]. There are several explanations for this discrepancy. First, there is a difference in the number of samples tested. Tsien and colleagues [61] tested a limited set of FSHD and control samples of different tissue origin. From these samples no repeat array configurations were known. In contrast, we studied as many as 40 blood samples of FSHD patients and 79 blood samples of healthy control individuals, all monosomic or disomic for D4Z4, to be certain that even subtle differences would be detected. Second, an interdependency between two or more identical restriction sites within the same fragment analysed cannot be excluded. Therefore, we used two different methylation-sensitive restriction enzymes that cut once in each D4Z4 repeat unit, in combination with probe p13E-11 that recognises a region proximal to D4Z4 specific for chromosome 4 alleles when combined with the restriction enzyme *Bln*I. In comparison, Tsien *et al.* [61] used four different methylation-sensitive enzymes that all have multiple restriction sites in the D4Z4 sequence in combination with a 700 bp subfragment of D4Z4 as a probe. Therefore, chromosome 4-specific methylation differences may be masked by cross-hybridisation of this probe to the many sequences in the genome homologous to D4Z4. Finally, the assay applied differed between both studies, which resulted in a major discrepancy concerning the number of D4Z4 copies analysed. Where we focussed on the proximal repeat unit of D4Z4 on chromosome 4, Tsien *et al.* [61] analysed the methylation pattern of the entire array on chromosomes 4 and 10, or in some cases only on chromosome 4. In our survey, we noticed that simultaneous quantification of the proximal units of chromosomes 4 and 10, by deliberately omitting *Bln*I from the methylation assay, could lead to an imprecision of the methylation values by at least 10% in either direction (PGM van Overveld, unpublished results). Therefore, DNA hypomethylation of FSHD alleles may have gone unnoticed by Tsien *et al.* due to the relatively small contribution of affected D4Z4 units to the total number of units assayed from chromosomes 4 and 10, and possibly other homologous units in the genome.

That the observed methylation changes truly have an effect on FSHD pathology is most clearly demonstrated by the analysed alleles of phenotypic FSHD patients. These normal-sized chromosome 4 alleles also showed a pronounced DNA hypomethylation of D4Z4, and with levels even below those observed for FSHD patients with a short 4q35 repeat array. Methylation levels of the D4Z4 repeat in these phenotypic FSHD patients are comparable to the very low levels observed in ICF patients [23, 61]. While low DNA methylation values in FSHD patients linked to 4q35 were restricted to the contracted chromosome 4 allele, in phenotypic FSHD patients D4Z4 repeat arrays are hypomethylated on both chromosomes 4.

In our assay all non-penetrant gene carriers (one person with 22 kb, two unrelated individuals with a 23 kb repeat size, two sisters with 27 kb fragments, one family with 3 individuals with a length of 30 kb and finally one family with two family members in separate

generations with an array of 33 kb) showed on average similar hypomethylation values as disomic FSHD patients. Even though we did not have any monosomic gene carriers to compare them to, the hypomethylation of the FSHD-sized allele in these individuals may also have been masked by the methylation of the long allele, similar to the effect observed in disomic FSHD patients. Non-penetrant gene carriers comprise one-third of all individuals over 20 years old in large pedigrees that do not have any complaints of muscle weakness [45, 47-49, 60], but carry a contracted 4q35 allele associated with D4Z4 hypomethylation. This suggests that hypomethylation of the D4Z4 repeat is necessary, but not sufficient for the expression of the FSHD phenotype. These individuals may thus carry a still unknown modifier gene(s) that prevents disease expression.

To study potential relationships of D4Z4 hypomethylation with residual repeat size and clinical severity of FSHD, we compared the clinical severity score (CSS) of monosomic individuals with the methylation data of these independent FSHD alleles. Our finding that methylation values were significantly different from healthy alleles provided further support for hypomethylation of FSHD disease alleles. We also found additional evidence for involvement of chromatin organisation, as was already suggested by the multimodal size distribution of repeat arrays of chromosomes 4 and 10, because we did not observe monosomic individuals carrying a repeat array of 30-60 kb. Although the number of individuals tested here is small, this suggests that there is a borderline repeat size (>60 kb) necessary to remain a healthy chromosome 4 repeat array; i.e. a minimum of one chromatin loop.

Furthermore, we observed a similar correlation between disease severity, using the CSS list formulated by Ricci *et al.* [51], and residual fragment size as was described previously [17, 22, 36, 51, 60, 70]. However, this correlation had a weak predictive validity. In more detail, within the interval of 20-38 kb a large variability in clinical severity of the patient was observed, while patients with a residual fragment size of less than 20 kb all score high on the CSS list and are thus severely affected. Although it is generally accepted that short repeat arrays coincide with a high clinical severity, there was no linear relationship for the FSHD population as a whole.

We did also not observe a linear relationship of D4Z4 methylation with repeat size or severity of the disease by this method, but our data suggested that D4Z4 methylation may in part determine FSHD severity. Although both groups were too small to test statistically, we observed that FSHD can be separated according to two clinical severity classes: patients with residual repeat arrays of 10-20 kb were always severely affected and showed pronounced hypomethylation of the D4Z4 repeat, while patients in the 20-38 kb interval showed both in clinical severity as in degree of hypomethylation a large interindividual variation not related to residual repeat size. If DNA methylation of D4Z4 is indeed contributing to clinical severity,

differential DNA methylation patterns within a single family may explain the observed intrafamilial variation in disease expression. The development of an assay to study the methylation status of all CpG dinucleotides in the D4Z4 sequence will thus be necessary to address this issue properly.

A satisfactory, and already often applied method, is the treatment of genomic DNA with sodium bisulfite (reviewed in [10, 41]), a procedure that converts all unmethylated cytosines in the DNA to uracil by deamination, but leaving the methylated cytosines unchanged. The target sequence is then amplified by PCR using strand-specific primers, which replace the uracil residues with thymines and methylated cytosines with cytosines [29]. This can be done by methylation-specific PCR, in which primers are designed that exclusively anneal with either methylated or unmethylated converted DNA. Methylation patterns can then be derived from the presence or absence of a PCR product [16, 33]. Another option would be to amplify the converted DNA using primers that anneal at regions that lack CpGs in the original sequence. This results in different PCR products that each have the same size, but differ in sequence at potential CpG methylation sites. The relative occurrence of the different methylation statuses within this mixture can subsequently be analysed [6]. After performing a PCR the products can be analysed with any technique capable of detecting (quantitative) sequence differences. Although these methods could inform us on the percentage of methylated cytosines and their exact locations within the D4Z4 sequence, it will be a challenge to design primers specific for the D4Z4 sequence considering the high GC content of this repeat together with the repetitive nature and the genome-wide distribution of homologous sequences. In addition, one could apply pulsed-field gel electrophoresis on DNA digested with the appropriate restriction enzymes. This allows separation of a single healthy or disease chromosome 4 fragment and thus provides an accurate and complete overview on methylation status in the chromosomal area of interest in combination with specific primers.

Although ICF syndrome patients also showed hypomethylation at D4Z4, this is in fact the only similarity between these two diseases. The observed generalised hypomethylation at all D4Z4 sequences on chromosomes 4 and 10 in DNA from ICF patients is not surprising in view of the primary defect these patients have in *DNMT3B,* a gene that encodes a DNA methyltransferase involved in *de novo* methylation [23]. Therefore, we wondered if mutations in any of the *DNMT* genes would similarly cause the observed hypomethylation in phenotypic FSHD patients. We analysed the DNA sequences of *DNMT3A* and *DNMT3B* in patients with a contracted D4Z4 repeat and phenotypic FSHD patients, but did not yet find any mutation (PGM van Overveld, unpublished results). Once cell lines derived from various tissues of these phenotypic FSHD patients become available, the possible role of DNMTs could be further investigated by

functional complementation to seek proteins that correct a defect in the recipient cell line. If adding the DNMT protein of interest to the cell culture would restore D4Z4 methylation patterns to those observed for the D4Z4 repeat in control individuals, this would suggest that the DNMT tested indeed may play a role in establishing the chromatin structure of D4Z4.

The observed hypomethylation in FSHD patients with and without a D4Z4 repeat contraction indicates a prominent role for this epigenetic DNA modification in FSHD. Interestingly, the hypomethylation of FSHD alleles is most pronounced at the restriction site for *Fse*I, located in the putative promoter region of *DUX4*, and to a lesser extent for *Bsa*AI 350 bp proximal to *Fse*I. It is conceivable that hypomethylation of the *DUX4* promoter region may activate ectopic *DUX4* expression. Homeobox proteins act as transcription factors in normal development. These homeobox-containing genes are often expressed during embryogenesis and their transcription can be influenced by DNA methylation [21]. As *DUX4* encodes a putative homeobox-protein, we cannot exclude that, as a consequence of the DNA hypomethylation, expression of *DUX4* in early development could be changed and might contribute to or even initiate FSHD pathology. Subsequent to the developmental stages, homeoboxes cannot be activated anymore and expression of *DUX4* would then be terminated, thus explaining the absence of *DUX4* expression in postnatal tissues.

Recent findings suggest that a local reduction of a specific repressor complex bound to D4Z4 may cause inappropriate activation of 4q35 genes by contraction of the array [11], although this model does not explain the chromosome 4qA allele specificity of FSHD with aberrant transcription. This local reduction could be enhanced by D4Z4 hypomethylation. Besides this repression model of Gabellini *et al.* [11], DNA hypomethylation of the D4Z4 repeats also fits the looping model proposed by Jiang *et al.* [8, 19] in which communication between a short D4Z4 repeat array and a target gene (or genes) occurs in *cis* by intrachromosomal looping. In this model, looping might be promoted by the more open chromatin structure of D4Z4 and/or activate inappropriate gene expression itself by allowing proteins to bind to hypomethylated sequences. In addition, the model suggesting that a heterochromatic chromatin conformation is a requirement for proper functioning of genes in the 4q35 region also still fits with our methylation data. Thus, with the observed DNA hypomethylation in FSHD patients we obtained support for an allele-specific chromatin change and this may provide an explanation unifying several disease models proposed for FSHD pathology.

Our methylation assay may also have diagnostic value for two small groups of patients. In a pilot study we applied our methylation assay to 25 coded samples derived from two large families that

contained an unknown number of FSHD patients with a 4q35 contraction as well as an unknown number of phenotypic FSHD patients. Since we know that the level of methylation on both tested restriction sites in phenotypic FSHD patients should be below the average methylation of FSHD patients with a 4q35 contraction, it was possible to successfully distinguish which individuals were marked as phenotypic FSHD patients (PGM van Overveld, unpublished results). We also tested DNA from patients where the clinical findings do not point with certainty to FSHD. These individuals did not have a contracted 4q35 repeat array and showed methylation values comparable to control individuals (PGM van Overveld, unpublished results). However, the limitation of this technique is that it only provides information on the methylation status of two methylation-sensitive restriction sites. Once it becomes possible to determine the methylation status of every CpG in the D4Z4 sequence, this may provide a promising decisive molecular marker to confirm the status "phenotypic FSHD patient" in the absence of a D4Z4 contraction. The application of a method allowing complete detailed analysis of every methylated cytosine in the D4Z4 sequence, like bisulfite sequencing [10, 33, 41], can hopefully in the future be used for diagnostic and research purposes, especially to confirm the clinical diagnosis in phenotypic FSHD patients more accurately than be done with the modified dosage test that we have applied thus far.

8.3.3 Histone modifications

8.3.3.1 H4 acetylation of D4Z4

Chromatin immunoprecipitation (ChIP) has become an important tool in studying chromatin and DNA-protein interactions in the living cell. The technique identifies DNA sequences that are directly bound to a nucleosome or protein complex [43, 55]. With a general antibody recognising histone 4 (H4) acetylation, which distinguished best and most consistently between constitutive heterochromatin and unexpressed euchromatin [20], we tested the presumed loss of heterochromatin of the 4q35 region (i.e. the loss-of-PEV model) with ChIP on cell fractions of cell lines, blood samples and muscle biopsies derived from a limited set of FSHD patients and healthy controls (***Chapter 7***). None of the regions studied showed any differences between cells from healthy individuals and FSHD patients. H4 acetylation levels of probe region p13E-11 proximal to the D4Z4 repeat array were comparable to levels observed in unexpressed euchromatin both on chromosome 4 and 10. Furthermore, the different cell fractions displayed H4 hyperacetylation at the 5' regions of candidate genes *FRG1* and *ANT1* located proximal to D4Z4. ChIP assays of the promoter region of *DUX4* and the D4Z4 sequence itself did not provide evidence for low levels of H4 acetylation on 4q35. In addition, no evidence for H4 hypoacetylation was observed in the promoter of *FRG2* on either chromosome 4 or 10.

8.3.3.2 Discussion and perspectives for future research

This study was the first attempt to focus on histone modifications in the 4q35 region. When H4 acetylation of chromatin from this chromosome 4 region was compared to that of the heterochromatic and euchromatic standards [20] in a limited set of samples, the 4q35 region appeared more similar to unexpressed euchromatin than to constitutive heterochromatin. This suggests that the D4Z4 repeat and the proximal sequences are unlikely to be heterochromatic in nature. Furthermore, there were no differences observed between samples from healthy individuals and FSHD patients and no relation between the number of D4Z4 copies and histone acetylation was observed. The H4 acetylation data thus do not support the model, which suggests that D4Z4 repeats are important for normal folding of the chromosome in a heterochromatic state and that a contraction may change the (local) chromatin conformation, resulting in a loss of heterochromatin and subsequent extinguishing or altering of gene activity on 4q35. These analyses do suggest that the FSHD-associated region is located in unexpressed euchromatin, but no difference between patients and control individuals was observed. It is possible that histone modifications other than acetylation take place upon contraction of the D4Z4 array and thus may influence the chromatin conformation. To elucidate if other chromatin modifications are involved in FSHD, ChIP analysis could be extended with antibodies detecting other histone modifications or proteins associated with chromatin components. The most interesting to test is the multiprotein complex capable of binding to D4Z4 [11], consisting of proteins that can activate or repress transcription or that are possibly involved in the organisation and maintenance of heterochromatic regions [3, 63].

Although muscle biopsies would be the first choice to test, they could unfortunately not be used for this ChIP procedure. Where possible, FSHD and control myoblasts, both proliferating and differentiating, should be analysed to obtain extra information on H4 acetylation and other chromatin modifications, and to allow comparisons between the tested cell populations. In addition, obtaining information on chromatin structure and binding of protein complexes at potential candidate gene sequences would also broaden our knowledge on a possible effect of D4Z4 repeat array size on the 4q35 region. Therefore, more research needs to be done before the model that assumed the loss of heterochromatin in the 4q35 region as a cause of FSHD development can truly be discarded. With regard to the other proposed models, the repression model of Gabellini *et al.* [11], the looping model proposed by Jiang *et al.* [8, 19] and a role for *DUX4* located within D4Z4 all still fit well with our data. In the future, a combination of ChIP with microarray technology (ChIP-on-chip) could be applied to detect protein binding to numerous loci [55]. This would facilitate the construction of whole-genome (disease-specific) maps of histone modifications, chromatin-modifying proteins and DNA-binding factors. Combined with DNA methylation data, the results of all these analyses would

provide a detailed picture of chromatin features in the 4q35 region and would hopefully result in a viable model, perhaps a combination of the several ones proposed today, that takes into account and explains all observed molecular changes that lead to the development of FSHD.

8.4 References

1. Agresti A, Meneveri R, Siccardi AG, Marozzi A, Corneo G, Gaudi S, and Ginelli E (1989) Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family). J Mol Biol 205(4): 625-631.
2. Batzer MA and Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3(5): 370-379.
3. Coppée F, Mattéotti C, Ansseau E, Sauvage S, Leclercq I, Leroy A, Marcowycz A, Gerbaux C, Figlewicz D, Ding H, *et al.* (2004) The DUX gene family and FSHD. Chapter 9: 117-134. In: Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
4. Deutekom van JCT, Bakker E, Lemmers RJLF, Wielen van der MJ, Bik E, Hofker MH, Padberg GW, and Frants RR (1996) 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(12): 1997-2003.
5. Deutekom van JCT, Wijmenga C, Tienhoven van EA, Gruter AM, Hewitt JE, Padberg GW, Ommen van GJ, Hofker MH, and Frants RR (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet 2(12): 2037-2042.
6. Eads CA and Laird PW (2004) Combined bisulfite restriction analysis (COBRA). Chapter 7: 71-85. In: DNA methylation protocols. Vol. 200. Mills KL, Ramsahoye BH (Eds.) Humana Press Inc., Totowa NJ.
7. Ehrlich M (2003) The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clin Immunol 109(1): 17-28.
8. Ehrlich M (2004) Exploring hypotheses about the molecular aetiology of FSHD: loss of heterochromatin spreading and other long-range interaction models. Chapter 17: 253-276. In: Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
9. Filipski J, Leblanc J, Youdale T, Sikorska M, and Walker PR (1990) Periodicity of DNA folding in higher order chromatin structures. EMBO J 9(4): 1319-1327.
10. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, and Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89(5): 1827-1831.
11. Gabellini D, Green M, and Tupler R (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell 110(3): 339-348.
12. Geel van M, Dickson MC, Beck AF, Bolland DJ, Frants RR, Maarel van der SM, Jong de PJ, and Hewitt JE (2002) Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics 79(2): 210-217.
13. Goto K, Nishino I, and Hayashi YK (2004) Very low penetrance in 85 Japanese families with facioscapulohumeral muscular dystrophy 1A. J Med Genet 41(1): e12-1-e12-5.

14. Gottlieb B, Beitel LK, and Trifiro MA (2001) Somatic mosaicism and variable expressivity. Trends Genet 17(2): 79-82.
15. Hall JG (1988) Review and hypotheses: somatic mosaicism: observations related to clinical genetics. Am J Hum Genet 43(4): 355-363.
16. Herman JG, Graff JR, Myohanen S, Nelkin BD, and Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA 93(18): 9821-9826.
17. Hsu YD, Kao MC, Shyu WC, Lin JC, Huang NE, Sun HF, Yang KD, and Tsao WL (1997) Application of chromosome 4q35-qter marker (pFR-1) for DNA rearrangement of facioscapulohumeral muscular dystrophy patients in Taiwan. J Neurol Sci 149(1): 73-79.
18. Jeffreys AJ, Wilson V, and Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314(6006): 67-73.
19. Jiang G, Yang F, Overveld van PGM, Vedanarayanan V, Maarel van der SM, and Ehrlich M (2003) 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(22): 2909-2921.
20. Jiang G, Yang F, Sanchez C, and Ehrlich M (2004) Histone modification in constitutive heterochromatin versus unexpressed euchromatin in human cells. J Cell Biochem 93(2): 286-300.
21. Kay PH, Harmon D, Fletcher S, Ziman M, Jacobsen PF, and Papadimitriou JM (1997) Variation in the methylation profile and structure of Pax3 and Pax7 among different mouse strains and during expression. Gene 184(1): 45-53.
22. Köhler J, Rohrig D, Bathke KD, and Koch MC (1999) Evaluation of the facioscapulohumeral muscular dystrophy (FSHD1) phenotype in correlation to the concurrence of 4q35 and 10q26 fragments. Clin Genet 55(2): 88-94.
23. Kondo T, Bobek MP, Kuick R, Lamb B, Zhu X, Narayan A, Bourc'his D, Viegas-Pequignot E, Ehrlich M, and Hanash SM (2000) Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum Mol Genet 9(4): 597-604.
24. Lee JH, Goto K, Matsuda C, and Arahata K (1995) Characterization of a tandemly repeated 3.3 kb *Kpn*I unit in the facioscapulohumeral muscular dystrophy (FSHD) gene region on chromosome 4q35. Muscle Nerve 2: S6-S13.
25. Lee JH, Goto K, Sahashi K, Nonaka I, Matsuda C, and Arahata K (1995) Cloning and mapping of a very short (10-kb) EcoRI fragment associated with facioscapulohumeral muscular dystrophy (FSHD). Muscle Nerve 2: S27-S31.
26. Lemmers RJLF, Kievit de P, Sandkuijl LA, Padberg GW, Ommen van GJ, Frants RR, and Maarel van der SM (2002) Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet 32(2): 235-236.
27. Lemmers RJLF, Maarel van der SM, Deutekom van JCT, Wielen van der MJR, Deidda G, Dauwerse HG, Hewitt J, Hofker M, Bakker E, Padberg GW, and Frants RR (1998) Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. Hum Mol Genet 7(8): 1207-1214.
28. Lemmers RJLF, Overveld van PGM, Sandkuijl LA, Vrieling H, Padberg GW, Frants RR, and Maarel van der SM (2004) Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. Am J Hum Genet 75(1): 44-53.
29. Li LC and Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. Bioinformatics 18(11): 1427-1431.
30. Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64(1): 24-30.
31. Liao D, Pavelitz T, Kidd JR, Kidd KK, and Weiner AM (1997) Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid

intrachromosomal homogenization and rare interchromosomal gene conversion. EMBO J 16(3): 588-598.
32. Liao D and Weiner AM (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the (CT)n.(GA)n microsatellite embedded within the U2 repeat unit. Genomics 30(3): 583-593.
33. Liu ZJ and Maekawa M (2003) Polymerase chain reaction-based methods of DNA methylation analysis. Anal Biochem 317(2): 259-265.
34. Lunt PW, Compston DA, and Harper PS (1989) Estimation of age dependent penetrance in facioscapulohumeral muscular dystrophy by minimising ascertainment bias. J Med Genet 26(12): 755-760.
35. Lunt PW and Harper PS (1991) Genetic counselling in facioscapulohumeral muscular dystrophy. J Med Genet 28(10): 655-664.
36. Lunt PW, Jardine PE, Koch MC, Maynard J, Osborn M, Williams M, Harper PS, and Upadhyaya M (1995) 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(5): 951-958.
37. Maarel van der SM, Deidda G, Lemmers RJLF, Bakker E, Wielen van der M, Sandkuijl LA, Hewitt JE, Padberg GW, and Frants RR (1999) A new dosage test for subtelomeric 4;10 translocations improves conventional diagnosis of facioscapulohumeral muscular dystrophy (FSHD). J Med Genet 36(11): 823-828.
38. Maarel van der SM, Deidda G, Lemmers RJLF, Overveld van PGM, Wielen van der M, Hewitt JE, Sandkuijl LA, Bakker B, Ommen van GJ, Padberg GW, *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(1): 26-35.
39. Matsumura T, Goto K, Yamanaka G, Lee J, Zhang C, Hayashi YK, and Arahata K (2002) Chromosome 4q;10q translocations; Comparison with different ethnic populations and FSHD patients. BMC Neurol 2(7).
40. Mefford HC and Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet 3(2): 91-102.
41. Oakeley EJ (1999) DNA methylation analysis: a review of current methodologies. Pharmacol Ther 84(3): 389-400.
42. Olckers A and van der Merwe A (2004) Possible epigenetic protective factor for FSHD identified in the black South African population. Am J Hum Genet 75(4; supplement) #1131.
43. Orlando V (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem Sci 25(3): 99-104.
44. Overveld van PGM, Lemmers RJLF, Deidda G, Sandkuijl LA, Padberg GW, Frants RR, and Maarel van der SM (2000) Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum Mol Genet 9(19): 2879-2884.
45. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
46. Padberg GW, Frants RR, Brouwer OF, Wijmenga C, Bakker E, and Sandkuijl LA (1995) Facioscapulohumeral muscular dystrophy in the Dutch population. Muscle Nerve 2: S81-S84.
47. Padberg GW, Lunt PW, Koch M, and Fardeau M (1997) Facioscapulohumeral muscular dystrophy. Chapter 3: 9-15. In: Diagnostic criteria for neuromuscular disorders. Emery AEH (Ed.) ENMC, Baarn, The Netherlands.
48. Padberg GW, Vogels OJ, and van der Kooi EL (1998) The clinical picture of FSHD. Muscle Nerve Supplement 7: S25.
49. Padberg GW and Wohlgemuth M (2003) Facioscapulohumeral muscular dystrophy (FSHD). University Medical Center St. Radboud, Nijmegen, The Netherlands.

50. Paques F and Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63(2): 349-404.
51. Ricci E, Galluzzi G, Deidda G, Cacurri S, Colantoni L, Merico B, Piazzo N, Servidei S, Vigneti E, Pasceri V, *et al.* (1999) Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of *Kpn*I repeats at the 4q35 locus and clinical phenotype. Ann Neurol 45(6): 751-757.
52. Rockman MV and Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19(11): 1991-2004.
53. Royle NJ (1995) The proterminal regions and telomeres of human chromosomes. Adv Genet 32: 273-315.
54. Stead JD and Jeffreys AJ (2000) Allele diversity and germline mutation at the insulin minisatellite. Hum Mol Genet 9(5): 713-723.
55. Steensel van B and Henikoff S (2003) Epigenomic profiling using microarrays. Biotechniques 35(2): 346-357.
56. Steinert S, Shay JW, and Wright WE (2004) Modification of subtelomeric DNA. Mol Cell Biol 24(10): 4571-4580.
57. Su QX, Zhang C, Zeng Y, Lu XL, Liu XR, Wang ZH, and Zhu YZ (2003) Translocation between chromosomes 4q35 and 10q26 in facioscapulohumeral muscular dystrophy. Zhongguo Yi Xue Ke Xue Yuan Xue Bao 25(5): 581-584.
58. Suda T, Fujiyama A, Takimoto M, Igarashi M, Kuroiwa T, Waguri N, Kawai H, Mita Y, and Aoyagi Y (2002) Interchromosomal telomere length variation. Biochem Biophys Res Commun 291(2): 210-214.
59. Thoraval D, Asakawa J, Wimmer K, Kuick R, Lamb B, Richardson B, Ambros P, Glover T, and Hanash S (1996) Demethylation of repetitive DNA sequences in neuroblastoma. Genes Chromosomes Cancer 17(4): 234-244.
60. Tonini MM, Passos-Bueno MR, Cerqueira A, Matioli SR, Pavanello R, and Zatz M (2004) Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromusc Disord 14(1): 33-38.
61. Tsien F, Sun B, Hopkins NE, Vedanarayanan V, Figlewicz D, Winokur S, and Ehrlich M (2001) Methylation of the FSHD syndrome-linked subtelomeric repeat in normal and FSHD cell cultures and tissues. Mol Genet Metab 74(3): 322-331.
62. Tupler R, Berardinelli A, Barbierato L, Frants RR, Hewitt JE, Lanzi G, Maraschio P, and Tiepolo L (1996) Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. J Med Genet 33(5): 366-370.
63. Tupler R and Gabellini D (2004) Facioscapulohumeral muscular dystrophy (FSHD): a disorder of muscle gene derepression. Chapter 10: 135-168. In: Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
64. Wang N, Wu ZY, Wang CD, Wang ZQ, Lin MT, Fang L, and Murong SX (2003) Mechanism of translocation between chromosomes 4q and 10q in facioscapulohumeral muscular dystrophy. Zhonghua Yi Xue Za Zhi 83(8): 650-653.
65. Wijmenga C, Hansen RS, Gimelli G, Bjorck EJ, Davies EG, Valentine D, Belohradsky BH, Dongen van JJ, Smeets DF, Heuvel van den LP, *et al.* (2000) Genetic variation in ICF syndrome: evidence for genetic heterogeneity. Hum Mutat 16(6): 509-517.
66. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R, *et al.* (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet 2(1): 26-30.
67. Wu ZY, Wang ZQ, Murong SX, and Wang N (2004) FSHD in Chinese population: characteristics of translocation and genotype-phenotype correlation. Neurology 63(3): 581-583.
68. Wu ZY, Wang ZQ, Wang N, Lin MT, and Murong SX (2004) [Study on EcoR I fragment polymorphism of the subtelomeric domains within 4q35 and 10q26 with pulsed field gel

electrophoresis in the Chinese population]. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 21(6): 552-556.
69. Zatz M, Marie SK, Cerqueira A, Vainzof M, Pavanello RC, and Passos-Bueno MR (1998) The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. Am J Med Genet 77(2): 155-161.
70. Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campiotto S, Cerqueira A, Wijmenga C, Padberg GW, and Frants R r(1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. Am J Hum Genet 56(1): 99-105.
71. Zhang XY, Loflin PT, Gehrke CW, Andrews PA, and Ehrlich M (1987) Hypermethylation of human DNA sequences in embryonal carcinoma cells and somatic tissues but not in sperm. Nucleic Acids Res 15(22): 9429-9449.