



Universiteit
Leiden
The Netherlands

Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Overveld, P.G.M. van

Citation

Overveld, P. G. M. van. (2005, April 27). *Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat*. Retrieved from <https://hdl.handle.net/1887/2310>

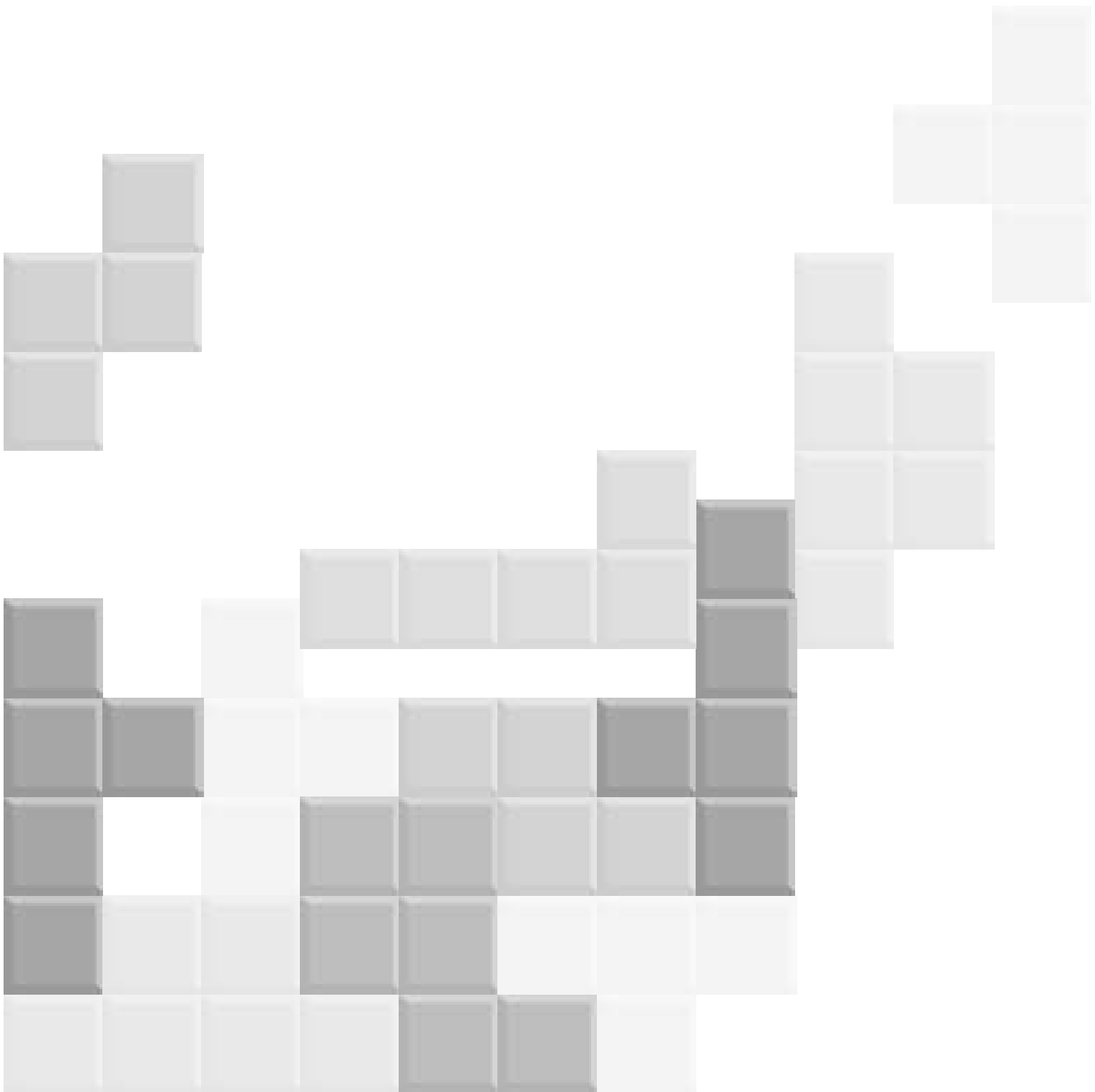
Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/2310>

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

Summary - Samenvatting



Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common hereditary muscle diseases with an estimated frequency of 1 in 20000. The disease has an autosomal dominant inheritance pattern and is characterised by a progressive and often asymmetric muscle weakness with an onset of disease in facial or shoulder girdle muscles.

The major locus for FSHD is linked to 4q35, located in the subtelomere on the long arm of chromosome 4. This region harbours a highly polymorphic *Eco*RI fragment that contains a large polymorphic repeat structure, designated D4Z4, which consists of 3.3 kb tandemly arranged D4Z4 repeat units and is highly susceptible to rearrangement. In the majority of patients this repeat is contracted to an array of 1-10 repeat units. However, 5% of FSHD patients, termed phenotypic FSHD patients, do not manifest a contracted D4Z4 array on chromosome 4, but share all clinical characteristics.

Complicating FSHD diagnosis, the subtelomere on the long arm of chromosome 10 (10q26) is highly similar to 4q35 and also contains a nearly identical polymorphic repeat array. However, size reductions of the chromosome 10 repeat array are non-pathogenic. Furthermore, exchanges between arrays on chromosomes 4 and 10 are frequently observed in patients and control individuals, but will only cause disease when a contracted array resides on chromosome 4. Finally, recent analysis of sequences distal to D4Z4 revealed two variants of the 4qter sequence, designated 4qA and 4qB. While both variants are almost equally present in the population, FSHD is uniquely associated with the 4qA variant.

Although FSHD is associated with a repeat contraction on 4qA, the exact mechanism causing disease is still unknown. The work in this thesis focused on structure and behaviour of the 4q35 subtelomere, aiming at elucidating possible molecular mechanisms that may mediate or cause FSHD pathology.

Subtelomeres are dynamic structures that are more often involved in recombination processes than other parts of the genome. Due to exchanges between subtelomeres of different chromosomes highly homologous DNA sequences can be dispersed throughout pericentromeric and subtelomeric domains in the genome. Since the repeat arrays located on the subtelomeres of the long arms of chromosomes 4 and 10 are highly homologous, we analysed these repeat array configurations in a healthy population, the results of which are presented in *Chapters 2* and *3*. This revealed the existence of translocated repeat arrays in 20% of individuals. Besides the presence of homogeneous chromosome 4 repeat units on chromosome 10, and *vice versa*, we also detected hybrid arrays that contained repeat units derived from both chromosomes.

With regard to repeat array size we observed that on average the arrays on chromosome 4 were longer than those on chromosome 10. This size difference was solely caused by alleles that carried the 4qA variant, as 4qB and 10q alleles did not differ in average length. Both repeat

arrays of chromosomes 4 and 10 displayed a similar multimodal allele size distribution that possibly reflects a higher-order chromatin structure. This multimodal distribution on chromosome 4 furthermore indicated the presence of a premutation domain containing alleles shorter than approximately 100 kb. Alleles in this domain may be more prone to contraction to a disease allele than arrays located in the larger repeat size intervals and may thus be predisposing for FSHD development.

In *Chapters 3* and *4* we examined repeat array configurations in *de novo* FSHD families and observed somatic mosaicism in more than 40% of cases, in either an FSHD patient or an asymptomatic parent of a non-mosaic patient. This mosaicism for a contracted D4Z4 repeat was more often seen in male than in female patients. In addition, affected females showed a higher proportion of cells with the contracted 4q35 repeat array than males, indicating that females have a higher clinical tolerance for mosaic disease alleles. Consistent with this finding is that, on average, there are more females with mosaicism for the FSHD region among unaffected parents.

Besides the observed mosaicism for a contracted FSHD allele generating two genetically distinct cell populations by gene conversion without crossover, we also identified FSHD patients with more complex rearrangements that resulted in three cell populations. This suggests that, alongside gene conversion without crossover, also gene conversion with crossover events that result in contraction and expansion of D4Z4 may contribute to the occurrence of mosaic D4Z4 alleles. Furthermore, whereas we did observe D4Z4 repeat units derived from chromosome 4 on chromosome 10 in mosaic individuals, the reverse configuration was never detected. The presence of such an extra chromosome 4 repeat array may facilitate gene conversion and may thus be a predisposing factor for contraction of the D4Z4 repeat array.

Eukaryotic cells have the capacity to epigenetically modify their genomes with a biochemical mark to alter the phenotype without changing the genotype. Two major epigenetic processes that mark chromatin are DNA methylation and the modifications of histones, such as acetylation. Due to the repetitive nature and the location on the subtelomere of 4q35, D4Z4 has been proposed to have a heterochromatic character. *Chapters 5, 6* and *7* present data on our studies on the chromatin structure of the D4Z4 repeat array, focussing on a possible change in structure of 4q35 in FSHD patients as a consequence of the chromosomal rearrangement. We demonstrated in *Chapters 5* and *6* that the D4Z4 contraction is associated with a significant DNA hypomethylation of D4Z4 at the disease allele in FSHD patients. Furthermore, phenotypic FSHD patients that carry normal-sized 4q35 repeat arrays also showed a pronounced DNA hypomethylation of D4Z4 with levels even below those observed for FSHD patients with a

contracted D4Z4 repeat array. While low DNA methylation values in FSHD patients linked to 4q35 are restricted to the disease allele, in phenotypic FSHD patients D4Z4 repeat arrays on both chromosomes 4 were found to be hypomethylated. These findings support an allele-specific chromatin change in FSHD patients with a D4Z4 contraction and strongly suggest that hypomethylation is a key event in the cascade of events causing the FSHD phenotype.

Analysis of histone 4 acetylation levels in the 4q35 region, as described in *Chapter 7*, indicated that the chromatin structure close to D4Z4 resembled that of unexpressed euchromatin rather than constitutive heterochromatin. This suggests that D4Z4 and proximal sequences are not heterochromatic. Contrary to our data presented on D4Z4 methylation, no histone 4 acetylation differences were observed between control individuals and FSHD patients. However, it remains possible that other histone modifications may influence the chromatin conformation upon contraction of the D4Z4 array.

The data presented in this thesis challenge the model suggesting the spreading of telomeric heterochromatin in a proximal direction upon contraction of the D4Z4 array, but fit in with other models that have been put forward to explain FSHD pathology. In addition to the repression model, suggesting that upon contraction a local reduction of a specific repressor complex bound to D4Z4 will cause inappropriate activation of 4q35 genes, our data also fit with the looping model, in which communication between a short D4Z4 repeat array and a target gene (or genes) occurs in *cis* by intrachromosomal looping. The model proposing a heterochromatic chromatin conformation as requirement for proper functioning of D4Z4 remains valid as well. Furthermore, as *DUX4* located in D4Z4 encodes a putative homeobox-protein, we cannot exclude whether expression of *DUX4* in early development will be altered due to D4Z4 contraction and so contributes to or even initiates FSHD pathology. Probably the actual disease-causing process will be a combination of the proposed mechanisms in which also the unique perinuclear localisation of chromosome 4q also has to be taken into account.

Besides providing new insights in the structure and complex behaviour of the chromosome 4q subtelomere associated with FSHD, this thesis also provides two observations relevant to the clinical practice. First, if mosaicism for D4Z4 is present in the germline, the percentage of these cells carrying the disease allele will determine the risk of having affected offspring, unlike in non-mosaic individuals carrying an FSHD-sized repeat array who have a 50% probability of transmitting the disease. More importantly, since mosaic females can carry a considerably higher percentage of cells with the disease fragment than male mosaics without manifesting FSHD, the disease may more easily go unnoticed in asymptomatic females than previously recognised. This indicates that there might be more apparently healthy women mosaic for a

contracted FSHD allele who will have an increased risk of having a child that develops FSHD provided this mosaicism extends to the germline. It is therefore important to screen for mosaicism in *de novo* FSHD families to provide more accurate information on inheritance risks.

Second, the methylation assay described in this thesis may have diagnostic and prognostic value, especially for phenotypic FSHD patients. Since we now know that the level of D4Z4 methylation in these individuals should be below the average methylation observed in FSHD patients with a 4q35 contraction, this can be used as a predictive molecular marker to confirm the status “phenotypic FSHD patient”. Hopefully, these two findings will become implemented in the molecular diagnosis of FSHD and contribute to improved genetic counselling.

Facioscapulohumerale spierdystrofie (FSHD) is een van de meest voorkomende erfelijke spierziekten, met een incidentie van 1 op 20.000. FSHD wordt klinisch gekarakteriseerd door een progressieve en vaak asymmetrische verzwakking van de skeletspieren, waarbij in een vroeg stadium de aangezichtsspieren en schouderbladspieren worden aangetast. De ernst van de ziekte verschilt per patiënt.

De erfelijke informatie (DNA) wordt in de cellen verpakt in 23 chromosoomparen, waardoor ieder chromosoom twee keer aanwezig is. FSHD is op DNA niveau geassocieerd met een genetische verandering dichtbij het uiteinde van de lange arm van chromosoom 4 (locatie 4q35). In dit gebied ligt de zeer variabele repeat D4Z4, die bestaat uit een herhalend patroon D4Z4 eenheden met elk een lengte van 3.3 kilobasen. De D4Z4 repeat is verder erg gevoelig voor DNA uitwisselingen. In de meeste FSHD patiënten wordt een verkorting van de D4Z4 repeat gevonden op één van de chromosomen 4, waardoor er nog maar 1-10 eenheden overblijven. In 5% van de FSHD patiënten wordt echter geen verkorte D4Z4 sequentie op chromosoom 4 gevonden, ondanks dat deze patiënten wel alle klinische kenmerken van de ziekte hebben. Deze patiënten worden daarom fenotypische FSHD patiënten genoemd. De ziekte heeft een autosomaal dominant overervingspatroon, waardoor een kind een kans van 50% heeft om de verkorte D4Z4 repeat te erven van een aangedane ouder en ziek te worden.

Het stellen van de diagnose FSHD wordt op moleculair niveau bemoeilijkt door de aanwezigheid van een bijna identieke repeat dichtbij het uiteinde van de lange arm van chromosoom 10 (locatie 10q26). Verkorting van deze chromosoom 10 repeat leidt echter niet tot FSHD. Verder worden er uitwisselingen tussen de repeterende eenheden van chromosoom 4 en chromosoom 10 repeats waargenomen. Desalniettemin leidt een repeatverkorting alleen tot ziekte wanneer deze op chromosoom 4 plaatsvindt en kan dus gebeuren onafhankelijk van de samenstelling van deze repeat, die darrdoor ook kan bestaan uit een mix van repeterende eenheden van chromosomen 4 en 10. Analyse van de sequentie tussen de D4Z4 repeat en het uiteinde van chromosoom 4 heeft bovendien twee varianten van deze sequentie aan het licht gebracht, die 4qA en 4qB worden genoemd. Beide varianten komen even vaak voor in de populatie, maar in FSHD patiënten is de repeatverkorting altijd gekoppeld aan de 4qA variant.

Ondanks dat FSHD geassocieerd is met een verkorting van de D4Z4 repeat met de 4qA variant, is het exacte mechanisme, dat de ziekte veroorzaakt, nog niet opgehelderd. Het werk beschreven in dit proefschrift heeft betrekking op de DNA structuur en het instabiele gedrag van de D4Z4 repeat op het uiteinde van de lange arm van chromosoom 4 met als doel de moleculaire mechanismen, die FSHD beïnvloeden of zelfs veroorzaken, te achterhalen.

De gebieden in de nabijheid van het uiteinde van een chromosoom worden subtelomeren genoemd. Dit zijn dynamische gebieden, die vaker dan andere genomische gebieden betrokken

zijn bij uitwisselingen van overeenkomstige DNA sequenties. Door uitwisselingen tussen de subtelomere gebieden van verschillende chromosomen kunnen deze sequenties gemakkelijk verspreid worden over het genoom, met name naar andere subtelomere gebieden en de gebieden rondom de centromeren. Omdat de sequenties van repeterende eenheden op de locaties 4q35 en 10q26 erg homoloog zijn, hebben we in een populatie gezonde mensen de samenstelling van beide repeats geanalyseerd. De resultaten staan beschreven in *Hoofdstuk 2* en *3*. Deze analyse leidde tot het waarnemen van uitwisselingen tussen de repeterende eenheden van de chromosoom 4 en 10 repeats in 20% van de individuen. We zagen op chromosoom 10 repeats opgebouwd uit homogene repeterende eenheden afkomstig van chromosoom 4 en homogene repeterende eenheden afkomstig van chromosoom 10 op chromosoom 4. Ook troffen we gemengde repeats aan, die repeterende eenheden afkomstig van beide chromosomen bevatten.

Naast de verschillende mogelijkheden in de samenstelling van de repeats op beide chromosomen, bleek ook de repeat aanwezig op chromosoom 4 gemiddeld langer te zijn dan de repeat op chromosoom 10. Dit verschil in lengte werd voornamelijk bepaald door de repeats met de 4qA variant, want er werd geen significant lengteverschil gevonden tussen de chromosoom 10 repeats en de chromosoom 4 repeats met de 4qB variant. Verder kunnen alle waargenomen repeatlengten worden opgedeeld in domeinen. Dit geeft een voor chromosoom 4 en 10 vergelijkbare multimodale verdeling, die een bepaalde voorkeursvouwing van de repeatsequentie suggereert. De multimodale verdeling geeft tevens een premutatie domein aan, dat repeatlengten kleiner dan 100 kilobasenparen bevat. Chromosoom 4 repeats die in dit domein vallen zijn mogelijk eerder geneigd te verkorten tot de lengte die geassocieerd is met FSHD dan langere repeats in de andere domeinen en kunnen daardoor mogelijk eerder leiden tot het ontstaan van FSHD.

In *Hoofdstuk 3* en *4* hebben we de samenstelling van de repeats bekeken in families waar de verkorting niet wordt doorgegeven van ouder op kind, maar waar FSHD ontstaan is door een nieuwe verkorting van de D4Z4 repeat. In deze *de novo* families bleek dat meer dan 40% van de onderzochte individuen mozaïek was voor de D4Z4 repeat, wat betekent dat een deel van de cellen in het lichaam de verkorte D4Z4 repeat heeft die geassocieerd is met FSHD, terwijl de andere cellen een lange, gezonde D4Z4 repeat bevatten. De aanwezigheid van twee celpopulaties met ieder een verschillende D4Z4 repeatlengte werd in zowel FSHD patiënten als niet-aangedane ouders waargenomen. In patiënten werd dit mosaïcisme vaker gezien in mannelijke dan in vrouwelijke patiënten. Bovendien bleek in vrouwelijke patiënten de verkorte D4Z4 repeat in een groter percentage cellen aanwezig te zijn dan in mannelijke patiënten, wat suggereert dat vrouwen de aanwezigheid van cellen met een verkorte D4Z4 repeat beter kunnen

verdragen. In overeenstemming met deze tolerantie geldt, dat onder de niet-aangedane ouders, waarin we ook het mosaïcisme voor de D4Z4 repeat verkorting hebben waargenomen, er ook gemiddeld meer vrouwen zijn.

Mosaïcisme voor een verkorte D4Z4 repeat wordt meestal veroorzaakt door het mechanisme “genconversie zonder uitwisseling”, hetgeen resulteert in twee celpopulaties. We hebben echter ook FSHD patiënten geïdentificeerd, waarin een complexe herschikking van de repeterende eenheden van de D4Z4 repeat heeft geleid tot de aanwezigheid van drie celpopulaties, die elk een andere repeatlengte bevatten. Naast cellen met de oorspronkelijke D4Z4 repeatlengte vinden we hier ook een celpopulatie met een D4Z4 repeatverkorting en een celpopulatie met een verlenging van de D4Z4 repeatlengte. Dit geeft aan, dat ook het mechanisme “genconversie met uitwisseling” kan bijdragen aan het ontstaan van mosaïcisme voor de D4Z4 repeat.

We hebben daarnaast ook de aanwezigheid van mogelijke uitwisselingen tussen de repeterende eenheden van chromosoom 4 en 10 repeats bestudeerd. In de mozaïeke individuen waren wel repeterende eenheden van de D4Z4 repeat aanwezig in de repeat op chromosoom 10, maar we hebben nooit repeterende eenheden van de chromosoom 10 repeat op chromosoom 4 gevonden. De aanwezigheid van deze extra repeterende eenheden van chromosoom 4 op chromosoom 10 kunnen mogelijk het “genconversie” mechanisme vergemakkelijken en op die manier predisponeren voor het ontstaan van een D4Z4 repeatverkorting.

DNA wordt samen met DNA-geassocieerde eiwitten, waaronder histonen, geordend tot chromatine. Dit ordenen leidt tot verschillen in vouwing van het DNA met grote gevolgen voor de expressie van genen. Chromatine wordt globaal ingedeeld in twee vormen: heterochromatine en euchromatine. In het algemeen hebben heterochromatische gebieden een compacte vouwing, waardoor genen niet geactiveerd kunnen worden, terwijl in euchromatische gebieden met een meer open structuur genen wel tot expressie kunnen worden gebracht. Eukaryotische cellen hebben de mogelijkheid om chromatine te modifieren met behulp van een biochemische markering, waardoor de expressie van een DNA sequentie kan veranderen zonder dat de genomische sequentie veranderd wordt. Dit worden epigenetische veranderingen genoemd. Twee belangrijke epigenetische processen die chromatine kunnen markeren zijn DNA methylatie en de verschillende modificaties van de histonen, waaronder acetylatie. In gebieden waar het DNA sterk gemethyleerd is en er weinig histonen geacetyleerd zijn staan genen meestal uit. Door het repeterende karakter van de D4Z4 repeat en de locatie in het subtelomeer van chromosoom 4 wordt er aangenomen, dat de D4Z4 sequentie in heterochromatine ligt. In *Hoofdstukken 5-7* worden studies beschreven, waarin we ons gericht hebben op het bepalen van de chromatinestructuur van de D4Z4 repeat, nl. heterochromatine of euchromatine, en of er

mogelijk een verandering in deze structuur optreedt als gevolg van de repeatverkorting. We laten in *Hoofdstuk 5* en *6* zien, dat de verkorte D4Z4 repeat in FSHD patiënten geassocieerd is met een significante afname van DNA methylatie. In fenotypische FSHD patiënten, die geen verkorte D4Z4 repeat op chromosoom 4 hebben, nemen we zelfs een nog sterkere afname van DNA methylatie waar. Bovendien zijn in deze fenotypische FSHD patiënten de D4Z4 repeats op beide chromosomen 4 laag gemethyleerd, terwijl in FSHD patiënten met een verkorting van de D4Z4 repeats op één van de chromosomen 4 alleen de verkorte D4Z4 repeat laag gemethyleerd is. Deze waarnemingen ondersteunen het optreden van een chromatineverandering specifiek voor de verkorte D4Z4 repeat en suggereren sterk, dat een verminderde DNA methylatie een belangrijke factor is in de gebeurtenissen, die uiteindelijk leiden tot ziekte.

De bestudering van de acetylatie van histon nummer 4 in de sequentie van 4q35 wordt beschreven in *Hoofdstuk 7*. Deze analyse geeft aan, dat de chromatinestructuur van het gebied naast de D4Z4 repeat meer op een open, maar inactieve chromatinestructuur lijkt dan op heterochromatine. Dit suggereert, dat de D4Z4 repeat en de sequenties voor de repeat geen echt heterochromatische structuur hebben, maar een intermediair tussen heterochromatine en euchromatine. In tegenstelling tot onze bevindingen voor DNA methylatie vonden we hier geen verschil tussen FSHD patiënten en controle individuen. Het is echter mogelijk, dat de andere histon modificaties door het verkorten van de D4Z4 repeat de chromatinestructuur wel beïnvloeden.

Er zijn een aantal modellen voorgesteld om het ontstaansmechanisme van FSHD te verklaren. De in dit proefschrift gepresenteerde data leveren geen bewijs voor het model, dat de verspreiding van heterochromatine vanaf het uiteinde van chromosoom 4 bij een verkorting van de D4Z4 repeat voorspelt. Onze data ondersteunen wel het “repressor”-model, dat als gevolg van de repeatverkorting een lokale afname van een aan de D4Z4 repeat gebonden specifiek repressorcomplex oppert, wat vervolgens kan leiden tot een toename van genexpressie. Ook passen onze data in het “looping”-model, waarin communicatie tussen een verkorte D4Z4 repeat en een specifiek gen (of meerdere genen) ontstaat door de vorming van een intrachromosomale lus. Het model, dat een heterochromatische structuur als eis stelt voor het goed functioneren van de D4Z4 repeat is ook nog steeds toepasbaar. Tenslotte is er in de D4Z4 repeat een gensequentie aanwezig, die mogelijk codeert voor een eiwit, dat betrokken kan zijn bij de regulatie van genexpressie in de embryogenese. We kunnen niet uitsluiten, dat de expressie van dit gen verandert door de D4Z4 repeat verkorting en op die wijze kan bijdragen aan het ontstaan van FSHD. Waarschijnlijk zal het uiteindelijke mechanisme dat FSHD

veroorzaakt een combinatie zijn van alle voorgestelde modellen, waarin ook de unieke plaats van chromosoom 4 in de kern meegenomen moet worden.

Naast het verkrijgen van nieuwe inzichten in de structuur en het complexe gedrag van het chromosoom 4 subtelomeer heeft dit proefschrift ook twee observaties opgeleverd die van belang zijn voor klinisch onderzoek. Wanneer mosaïcisme voor de D4Z4 repeat ook aanwezig is in de geslachtscellen, is het van belang om het percentage geslachtscellen te bepalen dat daadwerkelijk de repeatverkorting bevat. Dit percentage bepaalt namelijk de kans om een kind met FSHD te krijgen en ligt anders dan bij niet-mozaïeke individuen, waar de kinderen 50% kans hebben om de verkorte D4Z4 repeat te erven en ziek te worden. En wat misschien nog wel belangrijker is, juist omdat vrouwen een hoger percentage cellen met een verkorte D4Z4 repeat kunnen hebben zonder ziek te worden, kan de aanwezigheid van een ziekte vaker onopgemerkt blijven dan eerder werd aangenomen. Dit geeft meteen aan, dat er waarschijnlijk meer vrouwen drager zijn van mosaïcisme voor een verkorte D4Z4 repeat dan mannen en dus een verhoogde kans hebben een kind te krijgen dat FSHD ontwikkelt, mits dit mosaïcisme inderdaad in de eicellen aanwezig is. Het is daarom van belang om op mosaïcisme te screenen in *de novo* families om accurate informatie te kunnen geven met betrekking tot de risico's van ziekte-overerving.

Als tweede kan de in dit proefschrift beschreven analyse voor DNA methylering mogelijk gebruikt worden als diagnostische test, met name voor fenotypische FSHD patiënten. Omdat we nu weten dat de DNA methylering van de D4Z4 repeat in deze groep patiënten lager ligt dan de gemiddelde waarden gevonden in de FSHD patiënten met een verkorte D4Z4 repeat, kan dit gebruikt worden als marker om de status "fenotypische FSHD patiënt" te bevestigen. Hopelijk worden deze twee bevindingen geïmplementeerd in het moleculaire FSHD onderzoek en kunnen ze een bijdrage leveren aan het verbeteren van het erfelijkheidsadvies.

