



Universiteit
Leiden
The Netherlands

Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Overveld, P.G.M. van

Citation

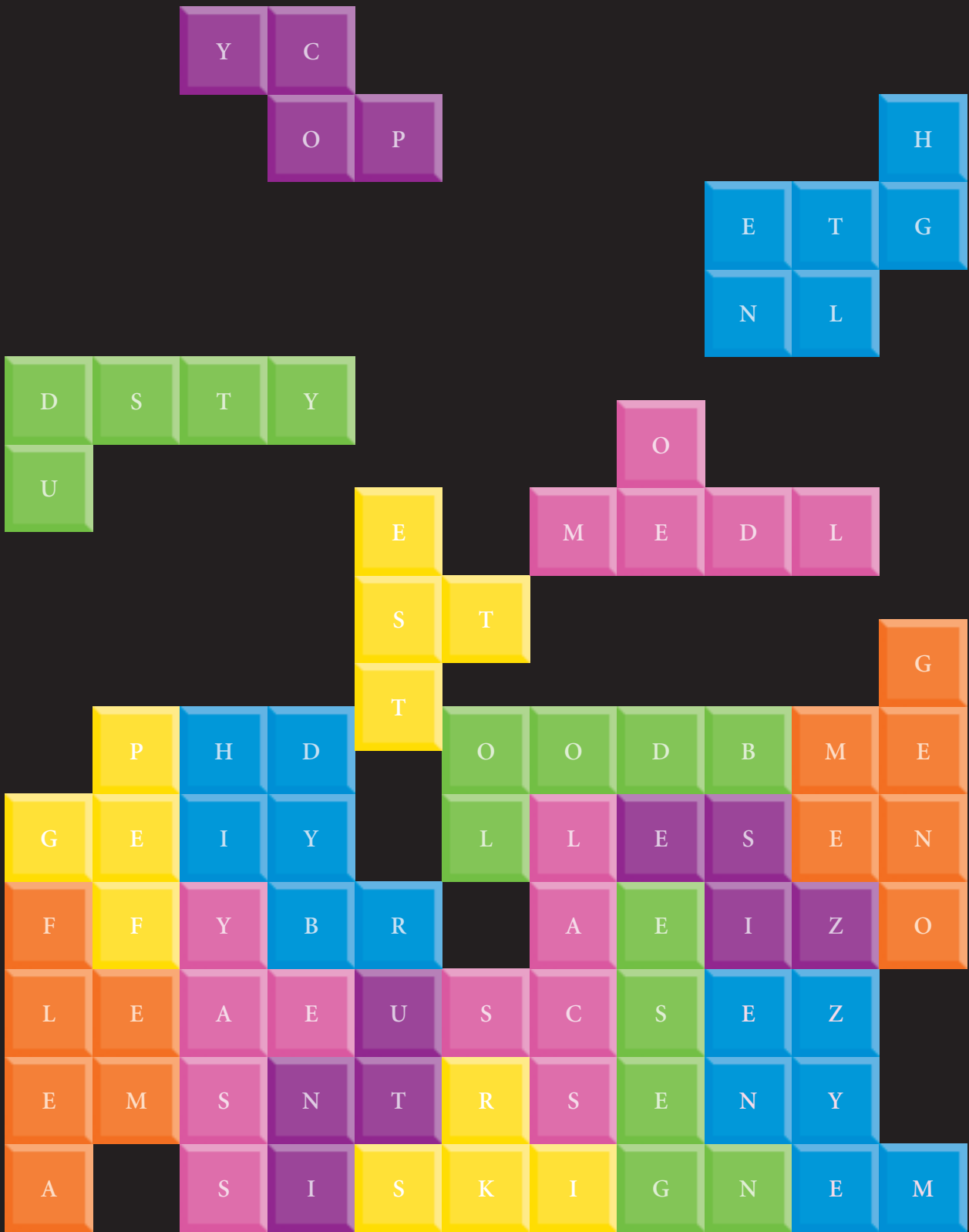
Overveld, P. G. M. van. (2005, April 27). *Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat*. Retrieved from <https://hdl.handle.net/1887/2310>

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).



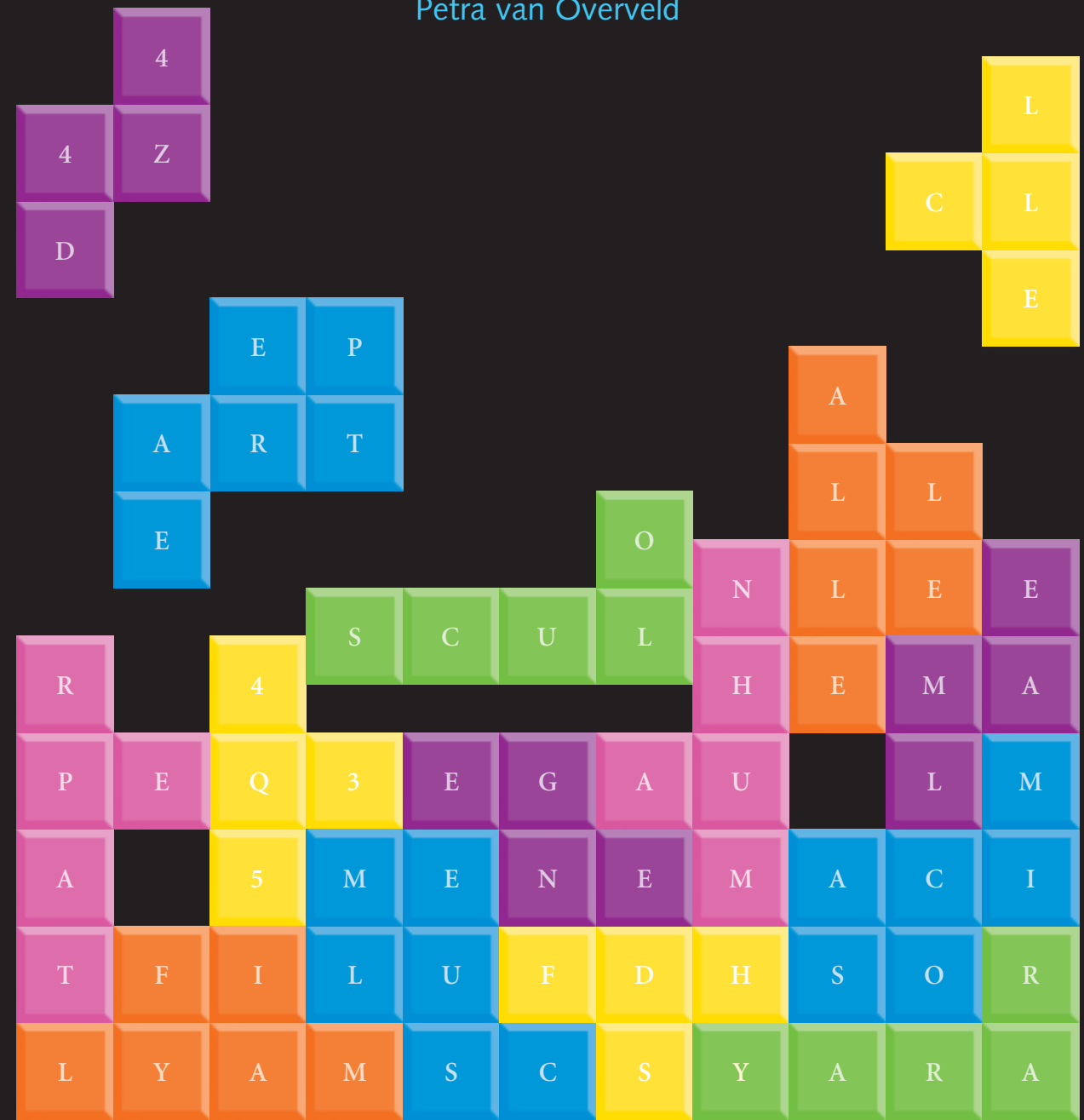
Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Petra van Overveld

2005

Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Petra van Overveld



Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Petra van Overveld

Petra Grada Maria van Overveld

Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Thesis, Leiden University

April 27, 2005

ISBN 90-9019370-7

No part of this thesis may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system without written permission of the copyright owner.

© 2005, P.G.M. van Overveld except (parts of) the following chapters

Chapter 1: Garland Science/BIOS Scientific Publishers Limited, 2004

Chapter 2: Oxford University Press, 2000

Chapter 3: The University of Chicago Press, 2004

Chapter 4: The University of Chicago Press, 2000

Chapter 5: Nature Publishing Group, 2003

Chapter 7: Oxford University Press, 2003

Cover: RotsLeeuw Design, Leiden and X-ingredient, 's Hertogenbosch

Print: Graphic Support, Dreumel

Genetic and epigenetic studies of the F5HD-associated D4Z4 repeat

Proefschrift

Ter verkrijging van
de graad van Doctor aan de Universiteit Leiden
op gezag van de Rector Magnificus Dr. D.D. Breimer
hoogleraar in de faculteit der Wiskunde en Natuurwetenschappen
en die der Geneeskunde
volgens besluit van het College voor Promoties
te verdedigen op woensdag 27 april 2005
klokke 11:15 uur

door

Petra Grada Maria van Overveld

geboren te Roosendaal en Nispen in 1976

Promotiecommissie

Promotoren:	Prof. Dr. RR Frants	
	Prof. Dr. GWAM Padberg	UMC St. Radboud, Nijmegen
Co-promotor:	Dr. Ir. SM van der Maarel	
Referent:	Prof. Dr. C Wijmenga	UMC Utrecht, Utrecht
Overige leden:	Prof. Dr. JC van Houwelingen	
	Prof. Dr. RGJ Westendorp	
	Dr. CMR Weemaes	UMC St. Radboud, Nijmegen

The studies described in this thesis have been performed at the Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands and were financially supported by grants from The Netherlands Organisation for Scientific Research (NWO), Prinses Beatrix Fonds, FSHD Stichting, Stichting Spieren voor Spieren, the Muscular Dystrophy Association USA, the FSH Society, the Shaw Family and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIH).

Publication of this thesis was financially supported by:
FSH Society, Inc., USA.
FSHD Stichting, The Netherlands.

The most exciting phrase to hear in science, the one that heralds new discoveries, is not “Eureka!” (I found it!), but rather “hmm... that’s funny...”

Isaac Asimov

Voor Leo

Table of contents

Preface	Aim and outline of this thesis	9
Chapter 1	Literature overview	15
	Clinical characteristics of FSHD	21
	Molecular characteristics of the FSHD-associated chromosomal region	28
	Organisation of repeats and subtelomeric regions in the genome	40
	Mosaicism and its consequences for human disease	50
	Epigenetic modifications	57
	Postulated disease models for FSHD pathology	71
Chapter 2	Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity Human Molecular Genetics (2000) 9: 2879-2884	99
Chapter 3	Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy American Journal of Human Genetics (2004) 75: 44-53	115
Chapter 4	<i>De novo</i> facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10 American Journal of Human Genetics (2000) 66: 26-35	135
Chapter 5	Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulo- humeral muscular dystrophy Nature Genetics (2003) 35: 315-317	153
Chapter 6	Residual D4Z4 repeat size and D4Z4 methylation levels separate FSHD into two clinical severity classes Submitted	167
Chapter 7	Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q Human Molecular Genetics (2003) 12: 2909-2921	181
Chapter 8	Discussion and future perspectives	209
	Subtelomeric exchanges between 4q35 and 10q26	213
	Mosaicism for the FSHD-associated region	220
	Epigenetic modifications of the 4q35 region	222
Summary - Samenvatting		237
List of abbreviations		251
List of publications		253
Curriculum Vitae		255
Dankwoord		257

Preface

Aim and outline of this thesis



Facioscapulohumeral muscular dystrophy (FSHD) is a myopathy with an autosomal dominant pattern of inheritance. After Duchenne muscular dystrophy and myotonic dystrophy, this disease is the third most common hereditary muscular dystrophy with a prevalence of approximately 1 in 20000 worldwide. FSHD is characterised by progressive muscle weakness of the facial and shoulder girdle muscles, which may then progress to pelvic girdle weakness or foot-extensor weakness with highly variable expression. The muscle weakness is often asymmetrical. Also the rate and extent of disease progression may differ greatly per patient. In most cases FSHD is associated with a contraction of an *Eco*RI fragment that contains a repeat array, D4Z4, consisting of 3.3 kb repeat units, located within the subtelomeric region 4q35 on the long arm of chromosome 4. The majority of affected individuals has a parent with clinical characteristics and a contraction of this repeat array on 4q35 and are thus described as familial FSHD patients. Approximately 10-30% of individuals will develop FSHD as a result of a new mutation and are therefore called *de novo* or sporadic patients. A small percentage of patients (5%; so-called phenotypic FSHD patients) has a phenotype characteristic for FSHD, but lack the 4q35 contraction.

Since the linkage of FSHD to chromosome 4 in 1990, important observations have been made with respect to the molecular characteristics and pathogenesis of the disease. Unfortunately, no true candidate gene or genes responsible for the development and progression of FSHD have thus far been identified. Unravelling the molecular structure of the 4q35 region and gaining more knowledge of the behaviour of the D4Z4 repeat are therefore important to elucidate the disease mechanism, as both features can give more insight in complex genetic events and possible molecular mechanisms triggering or modifying FSHD pathology. The aim of this thesis was therefore to focus on the structure and behaviour of D4Z4, which would add to our understanding of the molecular mechanism underlying the disease. The research described here focuses on three topics: (1) interactions of the subtelomeric region 4q35, in which D4Z4 resides, with other regions in the genome; (2) the consequences of mosaicism for FSHD pathology; and (3) epigenetic modifications of the 4q35 region, including DNA methylation and histone acetylation.

A literature overview on the FSHD phenotype and molecular characteristics of the chromosomal region 4q35 associated with FSHD is provided in *Chapter 1*. Furthermore, this chapter contains a broad overview on repeat characteristics, subtelomeric regions, the possible consequences of mosaicism, and the diverse functions of DNA methylation and histone modifications in the human genome in general. The observed interactions of the subtelomeric region 4q35 with chromosomal region 10q26 and the size distributions of chromosomes 4 and 10 are described in

Chapters 2 and 3. The studies on mosaicism for the FSHD-associated region presented in *Chapters 3 and 4* mainly concentrate on the occurrence of mosaicism and the determination of frequency and possible consequences for FSHD. The epigenetic modifications of the 4q35 region, D4Z4 and the proximal sequences, and their possible effects on chromatin conformation are addressed in *Chapters 5, 6 and 7.* *Chapters 5 and 6* pertain to DNA methylation of D4Z4, with special attention to phenotypic FSHD patients and the effect of repeat array length on DNA methylation, whereas *Chapter 7* outlines observations on histone acetylation. *Chapter 8* summarises the results described in this thesis and discusses the consequences of structure and behaviour of D4Z4 for FSHD pathology and perspectives for future research.

1

Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Literature overview

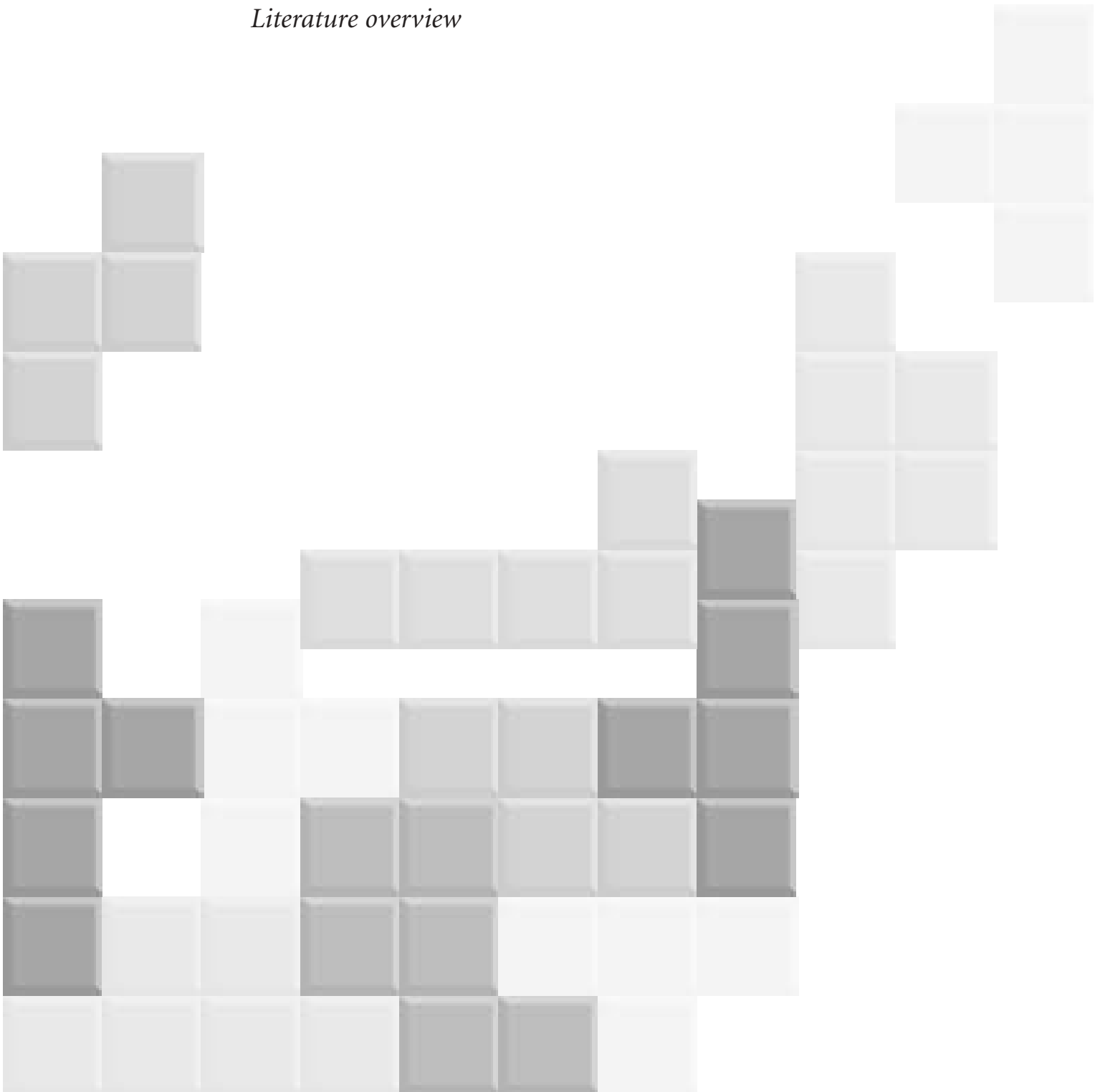


Table of contents

1.1 Clinical characteristics of FSHD	21
1.1.1 Historical perspective	21
1.1.2 Clinical characteristics.....	21
1.1.2.1 Diagnostic criteria	21
1.1.2.2 Distribution of muscle weakness.....	22
1.1.2.3 Muscle biopsy features.....	23
1.1.2.4 Onset and progression of the disease.....	23
1.1.3 Fitness.....	24
1.1.4 Differences between males and females	24
1.1.5 Extramuscular phenotypic features	25
1.1.5.1 Pain.....	25
1.1.5.2 Retinal abnormalities and sensorineural hearing loss	25
1.1.5.3 Respiratory problems.....	26
1.1.5.4 Unusual clinical observations	26
1.1.6 Phenotype-genotype correlations	26
1.2 Molecular characteristics of the FSHD-associated chromosomal region	28
1.2.1 Identification of the FSHD-associated region	28
1.2.1.1 Linkage analysis.....	28
1.2.1.2 Repeat array contractions	28
1.2.2 Analysis of the FSHD-associated D4Z4 repeat.....	29
1.2.2.1 D4Z4 repeat unit	30
1.2.2.2 D4Z4 proximal and distal sequences	31
1.2.2.3 Analysis of D4Z4 in the genome	32
1.2.2.4 D4Z4 in other species	32
1.2.3 Complications due to D4Z4 sequence homology.....	33
1.2.3.1 Restriction sites discriminate between chromosomes 4 and 10	33
1.2.3.2 A second FSHD locus?	34
1.2.4 Identification of 4q35 genes	34
1.2.4.1 ANT1; adenine nucleotide translocator gene 1	35
1.2.4.2 ALP; actinin-associated LIM protein gene.....	35
1.2.4.3 FRG1; FSHD region gene 1	36
1.2.4.4 TUBB4Q; tubulin beta polypeptide 4 member Q gene	36
1.2.4.5 FRG2; FSHD region gene 2	37
1.2.4.6 Detection of gene expression	38
1.2.5 <i>In vitro</i> cell culture systems	39

1.3 Organisation of repeats and subtelomeric regions in the genome	40
1.3.1 Repetitive elements in the genome	40
1.3.2 Mechanisms of repeat evolution	41
1.3.3 Repeat sequences associated with disease	42
1.3.3.1 Low Copy Repeats	42
1.3.3.2 Trinucleotide repeat expansions	43
1.3.4 Subtelomeric regions	45
1.3.4.1 Function of subtelomeres	45
1.3.4.2 Organisation of subtelomeres	46
1.3.5 Subtelomeric region 4q35 and its association with FSHD	47
1.3.5.1 Identification of complex exchange patterns	48
1.3.5.2 Somatic pairing of 4qter and 10qter	49
1.4 Mosaicism and its consequences for human disease	50
1.4.1 Inheritance of mosaicism	50
1.4.2 Germline and somatic mosaicism	51
1.4.3 Mosaicism is a common feature in cancer	53
1.4.4 Detection of mosaicism in FSHD	54
1.5 Epigenetic modifications	57
1.5.1 Chromatin architecture	57
1.5.1.1 Heterochromatin and euchromatin	57
1.5.1.2 Nucleosome conformation	58
1.5.1.3 Epigenetics	58
1.5.2 DNA modification by methylation	59
1.5.2.1 DNA methyltransferases	59
1.5.2.2 CpG islands	60
1.5.2.3 DNA methylation and development	60
1.5.2.4 Influence of DNA methylation on silencing	61
1.5.2.5 Suppression and defence	61
1.5.3 DNA modifications via histones	62
1.5.3.1 Histone acetylation	63
1.5.3.2 Histone methylation	63
1.5.3.3 Histone phosphorylation	63
1.5.3.4 Histone ubiquitination	64
1.5.3.5 Histone sumoylation	64
1.5.3.6 Histone ADP-ribosylation	65
1.5.3.7 Histone interactions	65
1.5.3.8 Histone variants	66

1.5.3.9 Function of histone modifications	66
1.5.4 Imprinting and X-inactivation	66
1.5.5 Chromatin remodelling and human disease.....	67
1.5.6 Epigenetic modifications and FSHD	69
1.6 Postulated disease models for FSHD pathology	71
1.6.1 Direct effect of D4Z4	71
1.6.2 Changes in chromatin structure of the 4q35 region	72
1.6.3 Nuclear localisation of the 4q35 region	73
1.7 References	75

1.1 Clinical characteristics of FSHD

1.1.1 Historical perspective

Physicians presented many clinical descriptions of patients with a total of 18 different muscle diseases in the nineteenth century. Most of the progressive muscle diseases were thought to have neurogenic causes [339]. In 1868, Guillaume Duchenne de Boulogne described a form of muscular dystrophy, nowadays called Duchenne muscular dystrophy, a progressive proximal muscular dystrophy with characteristic pseudohypertrophy of the calves, which is linked to the X-chromosome. In that same period, Louis Landouzy published the characteristics of two boys with progressive muscular wasting [218]. Together with his colleague Joseph Dejerine he observed this family for more than ten years and they described the key features of facioscapulohumeral muscular dystrophy (FSHD) in the title of their article “De la myopathie atrophique progressive: myopathie héréditaire débutant, dans l’enfance, par la face, sans altération du système nerveux” (i.e. a progressive muscle atrophy, which is hereditary and starts in childhood with the face, without involvement of the nervous system). They emphasized the onset in facial and shoulder muscles to differentiate FSHD from Duchenne muscular dystrophy [219-221]. FSHD is therefore also known as Landouzy-Dejerine disease.

1.1.2 Clinical characteristics

1.1.2.1 Diagnostic criteria

There are four key criteria, formulated by the international FSHD consortium, that specify FSHD at a clinical level [306]:

1. Onset of the disease in facial or shoulder girdle muscles; sparing of the extraocular and pharyngeal muscles and the myocardium.
2. Facial weakness is observed in more than 50% of the affected family members.
3. Evidence of a myopathy seen in electromyography and muscle biopsy without biopsy features specific for alternative diagnoses in at least one affected member.
4. An autosomal dominant inheritance pattern in familial cases.

On a molecular level FSHD is in most patients (95%) associated with a contraction of the D4Z4 repeat located in the subtelomere of chromosome 4 (4q35) [419], which can be inherited from an affected parent or be the result of a new mutation. The molecular characteristics will be discussed in more detail in *Paragraph 1.2*.

1.1.2.2 Distribution of muscle weakness

FSHD is characterised by weakness of specific muscles. Most patients note shoulder girdle weakness as a first symptom. Weakness of the facial muscles (with particular involvement of M. orbicularis oculi and M. orbicularis oris) is rarely an early complaint and is only in some cases (~5%) by patients recognised as onset of the disease [301]. Facial muscle involvement is often subtle and sometimes only noticeable by asymmetry of facial expression [306, 308]. On first clinical examination facial weakness is present in more than 90% of patients [206, 251, 298, 308].

Most patients seek medical attention, because of weakness of the shoulder girdle [298, 301, 339]. The scapula fixators are most prominently involved and also the pectoralis major muscles will in most cases become affected [306]. Most patients have at rest a sloped-shouldered posture with anterior rotation of the shoulders and elevation of the scapula from the rib cage [192]. The initially spared deltoid muscles become affected at a later stage [306]. Approximately 30% of patients never worsen beyond shoulder weakness [251, 301], in the remaining patients the disease progresses with the involvement of abdominal muscles and foot-extensor muscles. Pelvic girdle weakness and weakness of the upper arms and upper legs may occur at a later stage after onset of shoulder girdle weakness [214, 306, 308]. The severe weakness of the abdominal, pelvic girdle and back extensor muscles can all contribute to a lumbar hyperlordosis [187, 308]. Two FSHD patients with prominent weakness of facial and shoulder girdle muscles are depicted in *Figure 1.1*.

The muscle weakness is often asymmetrical and the degree of muscle weakness varies from person to person. Eventually, 10-20% of all patients will require a wheelchair due to

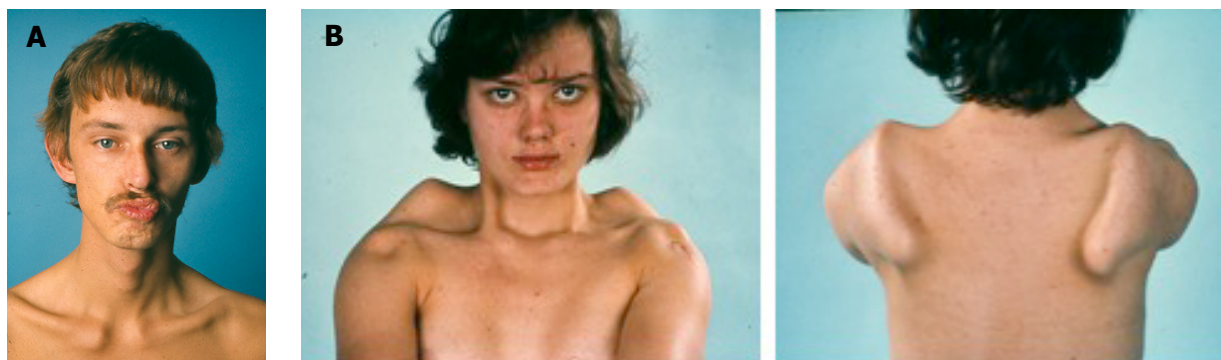


Figure 1.1

(A) Photograph of a male FSHD patient showing asymmetrical facial weakness, which is observed in more than 50% of patients [301].

(B) Two photographs illustrating shoulder girdle weakness with elevation of the scapula on attempted ante flexion of the area in a female FSHD patient.

Photographs were kindly provided by Prof. Dr. GW Padberg, Department of Neurology, UMC St. Radboud, Nijmegen, The Netherlands.

proximal lower limb involvement [255, 308]. This is often by the fifth decade of life, but sporadic patients (i.e. patients with a new mutation) may even require a wheelchair before the age of 20 years [255, 298, 306].

1.1.2.3 Muscle biopsy features

Muscle biopsies of FSHD patients do not show any disease-specific morphological characteristics and are to some extent variable, depending on disease progression and the site of biopsy [192, 340]. These biopsies display general dystrophic features such as increased variation in fibre type and size [340], fibre necrosis and fibrosis and an increased number of internal nuclei. Also moth-eaten fibres are frequently seen as well as scattered small angular fibres, indicative for regeneration as they contain fetal myosin [9, 299, 306, 340]. Mononuclear inflammatory cells with an increase in necrotic fibres have been detected in up to 40% of FSHD patients. The mechanism that causes these infiltrates and their significance is still unknown [9, 105, 192, 298].

1.1.2.4 Onset and progression of the disease

The clinical presentation of FSHD exhibits a wide range of clinical severity and variable age at onset, even within one family where all patients carry an identical mutation in their DNA. Symptoms can vary from severe progressive proximal and distal involvement of upper and lower limbs together with an expressionless face in early life to minimal signs of asymptomatic scapular girdle or facial weakness and thus barely detectible even at old age [17, 112, 255]. Usually patients become symptomatic in their adolescence, when the individual notices symptoms that reveal shoulder girdle weakness or signs of muscle wasting in this region [173, 251, 255, 292, 298, 444]. The disease manifests almost complete penetrance (95%) in all patients during the second decade of life and penetrance is probably close to 100% at 30 years [250, 298, 339].

Since facial weakness is almost never recognised as a first complaint, it is therefore very difficult to indicate the precise disease onset [298]. Because the development is usually determined in retrospect and depends on the recall of the patient, age at onset may be a doubtful marker for development [253]. Onset has now been defined as the moment a patient becomes aware of having FSHD or has difficulties caused by muscle weakness [308]. With this definition, one-third of all patients over 20 years old in large pedigrees do not have any complaints of muscle weakness. These individuals are also called non-penetrant gene carriers [298, 306-308, 380].

The disease is progressive in the majority of patients with a highly variable rate and usually spreads first to the abdominal and foot-extensor muscles and subsequently to pelvic

and upper arm muscles [214, 308]. Within families, approximately 50% of patients will develop lower limb involvement, and 20% of those become to some extent wheelchair-dependent. As a result of this intrafamilial variation, a precise prediction about disease progression and severity is still not possible [255].

Disease onset before the age of 5 years is seen in more severe patients, who usually have a new mutation, where facial weakness is the earliest and most prominent sign [250, 255, 301, 306]. In The Netherlands, less than 5% of patients have an onset before the age of 10 years [49], while in contrast Japanese researchers reported a frequency of 13-17% of early childhood cases [280, 437]. This early onset FSHD is part of the wide clinical spectrum of FSHD [51, 254] and has been somewhat arbitrarily defined by the following two criteria: signs or symptoms of facial weakness before the age of five and signs or symptoms of shoulder girdle weakness before the age of ten years [51, 300]. These patients often become wheelchair-dependent [301].

1.1.3 Fitness

Most affected individuals (i.e. 70-90% of patients) have a parent with clinical characteristics and a contraction of the D4Z4 repeat on chromosome 4, and are thus described as familial FSHD patients. Approximately 10-30% of individuals will develop FSHD as a result of a new mutation and are therefore called *de novo* or sporadic patients [8, 14, 199, 253, 305, 377, 392, 415, 419, 445]. Given this high mutation frequency, it is unlikely that the fitness, i.e. the ability to transmit one's genes to the next generation and have them survive in that generation to be passed on to the next, is normal [298, 299]. Although the relative fitness has been calculated at 1.0 in the Dutch kindreds studied, this does not correspond with the (probably conservative) estimation of 10% of living patients having *de novo* mutations [298]. The results collected from a Brazilian population suggest that new mutations may account for at least one-third of FSHD cases, and a reduced biological fitness of 0.6-0.82 by different estimates based on both familial and sporadic patients appears to be more correct [445].

1.1.4 Differences between males and females

A consistent gender difference is observed when asymptomatic and symptomatic individuals of the same age are compared. There are significantly more asymptomatic females and when the symptomatic patients are considered, a later onset of the disease and a slower disease progression are noted in females [298], implicating that females show less severe features of FSHD than males of the same age [251, 307, 330]. These findings are confirmed by a study in which Zatz *et al.* [444] studied a large Brazilian patient population and suggested, based on molecular analyses, that the average penetrance is 83% at the age of 30 years, in which males from that population showed a significantly higher penetrance than females (95 versus 69%).

In patients mosaic for the D4Z4 contraction, the proportion of mutated cells also influences expression of the disease. A smaller percentage of cells carrying the disease allele is required for males to manifest signs of FSHD compared to the detected percentages in females and these males are more frequently symptomatic [257]. In addition, DNA analysis of parents of FSHD patients shows that there is a female predominance of mosaic asymptomatic carriers [199, 257, 444] (see also *Paragraph 1.4* for more details on mosaicism). The underlying mechanism for the observed gender differences is not yet known. Maybe differences in hormone levels between males and females are of influence, but this needs further investigation.

1.1.5 Extramuscular phenotypic features

1.1.5.1 Pain

Approximately 75% of all patients complain about multifocal muscle and tendon pains, mostly located to the shoulder, lower back and arm regions [196, 197, 299, 329, 395], but pain is also reported around the thighs [56]. This pain is often described as serious and disabling and presents a major problem in daily life. Some types of pain may be related to strenuous activities or postural problems [56, 329]. To find a possible cause, metabolic factors and components indicative for inflammation were examined in blood and muscle biopsies with histological, histochemical and immunocytochemical techniques, but no clues were found that indicate a cause for the pain or an association with the muscle dystrophy in FSHD patients [56].

1.1.5.2 Retinal abnormalities and sensorineural hearing loss

The presence of retinal abnormalities, which include telangiectasis, microaneurysm formation, vessel occlusions, small exudates, haemorrhages, capillary closure and leakage in the macular area as well as in the peripheral retina, is observed in 50-75% of patients [113, 303, 305] with a high frequency in early onset cases [51]. These abnormalities can be visualised with fluorescein angiography, a method to diagnose and evaluate a variety of ocular diseases. Observed changes are often subtle and very focal and will usually not lead to any severe visual complications [112, 113].

A significant difference in hearing levels with frequencies between 4000 and 6000 Hz is reported in patients when compared to healthy family members. The hearing deficiency often displays as high-tone hearing loss observed in 64% of familial autosomal dominant cases and 20-30% of infantile cases [50, 173, 195, 198, 302] and may progress to involve all frequencies [306]. This hearing loss can be related to cochlear dysfunction in early onset FSHD [402], but hearing loss with a maximum of 8000 Hz in adult FSHD is also very difficult to distinguish from impairment in the normal population, which is age-related or caused by acoustic trauma [341]. Severe clinical deafness occurs rarely and has only been noted in severe childhood cases

[51, 112]. Comparing both sporadic and familial FSHD patients, no correlation between hearing impairment and disease severity was observed [302]. Häfner *et al.* [151] reported a locus linked to hearing loss at 4q35-qter (DFNA24) that could indicate a potential candidate gene for deafness in this region, but if this finding is of any significance for hearing deficits in FSHD remains to be investigated.

1.1.5.3 Respiratory problems

The respiratory capacity of FSHD patients diminishes with the progression of the disease, but usually does not lead to severe respiratory failure [187, 301]. Kilmer *et al.* [187, 188] reported a vital capacity evidence of restrictive lung disease measured in almost 50% of patients, but only 13% showed severe involvement and only 22% had a history of pulmonary complications. In a Japanese study, 8% of FSHD patients had progressive respiratory failure [281], while in a Dutch survey about 1% of all FSHD patients suffer from respiratory insufficiency and require ventilation support [429]. These Dutch patients all have a disease onset before the age of 24 years, a wheelchair-dependency and an abnormal curve of the vertebral column (scoliosis and lumbar lordosis) [429, 430]. A mild scoliosis is present in one third of all FSHD patients, mostly in early onset cases [301]. The presence of a pectus excavatum has been reported in 5% of all patients, which exceeds the incidence in the normal population [298, 308]. Both features, scoliosis and pectus excavatum, can impair respiratory function [429].

1.1.5.4 Unusual clinical observations

Several phenotypic findings have been published that are observed in combination with a FSHD diagnosis: involvement of the lingual muscle [202, 437, 438], cardiac involvement [103, 215, 283, 356, 366], schizophrenia [355], mental retardation and epilepsy [3, 49, 120, 271, 355, 389, 437]. These findings may be disease-specific, but as most of them were only detected in a few patients it is likely that they are not part of the clinical core phenotype.

1.1.6 Phenotype-genotype correlations

FSHD is associated with a D4Z4 repeat contraction at 4q35 that in general causes disease when the *EcoRI* fragment in which the array is located becomes smaller than 38 kb (reviewed in [391]; see also *Paragraph 1.2* for a detailed description on the molecular characteristics). Since the repeat array may vary in size between patients of different families, several research groups have studied the effect of residual repeat size on the severity of the phenotype. Comparing repeat size with clinical data revealed that age at onset and the severity in probands and *de novo* cases correlates roughly inversely with the size of the residual repeat array [8, 10, 167, 198, 253, 330, 377, 445]. Very severely affected patients (usually children) with lower limb involvement

most often carry fragments of 10-17 kb, whereas patients with onset in their teens and large recognised families mostly carry fragments of 18-30 kb [112, 120, 251-253, 255, 380]. Thus individuals with a milder phenotype on average have larger fragments, whereas those more severely affected have the smaller ones [112, 255, 308, 380]. Early onset patients usually are severe sporadic cases with a *de novo* mutation resulting in a short residual fragment [49, 120, 173, 253, 254, 301, 308]. Nevertheless, sporadic patients can also present a short fragment in combination with a mild phenotype [380, 435], short fragments are detected as well in asymptomatic gene carriers [380] and patients with repeat sizes above 30 kb can also show a severe presentation of the disease [57, 435], indicating that residual repeat length is not the only factor modulating the phenotype.

1.2 Molecular characteristics of the FSHD-associated chromosomal region

1.2.1 Identification of the FSHD-associated region

1.2.1.1 Linkage analysis

When polymorphic markers became available, it was possible to search for genes responsible for or associated with disease by association analysis and linkage analysis. In order to identify the genetic defect responsible for FSHD, linkage analysis was started in the early eighties with all available genetic markers, mainly consisting of blood group markers and enzyme and protein polymorphisms. Unfortunately, none of the 35 markers present at that time showed linkage to FSHD [304]. Next, the search for the FSHD locus was continued by use of restriction fragment length polymorphisms (RFLP) [250], but also these analyses did not yield any evidence for linkage.

By the early nineties, almost 95% of the human genome was excluded [347], but the FSHD locus had not been found. Then marker technology shifted from RFLPs to (CA)_n type microsatellite markers [133, 244, 411]. These markers were applied to DNA from two large Dutch FSHD families and gave promising results. One of these markers, Mfd22 [410], displayed positive linkage and located the putative FSHD locus to chromosome 4 [416]. A number of chromosome 4 markers were tested next to identify markers in closer linkage or flanking marker Mfd22. A variable number tandem repeat marker pH30 [269], showed high linkage corresponding to locus D4S139 and located the region associated with FSHD to 4q35 [420].

1.2.1.2 Repeat array contractions

After the linkage to chromosome 4q, clone 13E, which mapped distal to locus D4S139, was isolated from a chromosome 4 cosmid library. From this clone, which contained multiple copies of a 3.3 kb repeat sequence, an 800 bp probe termed p13E-11 (D4F104S1) flanking the repeats was isolated that could be used for hybridisation. On Southern blots human DNA digested with restriction enzyme *Eco*RI and probed with p13E-11 displayed a short fragment in FSHD families that co-segregated stably with the disease through generations [88, 415, 419] (*Figure 1.2.A*). This probe also displayed short fragments in sporadic FSHD patients, while this fragment was absent in either clinically unaffected parent, revealing a *de novo* DNA rearrangement [419] (*Figure 1.2.B*). Once this rearranged fragment had arisen, it could stably be transmitted to the next generation in association with the disease [415]. Additional genetic studies on FSHD families in other laboratories confirmed these findings [58, 137, 148, 261, 348, 392, 412, 445].

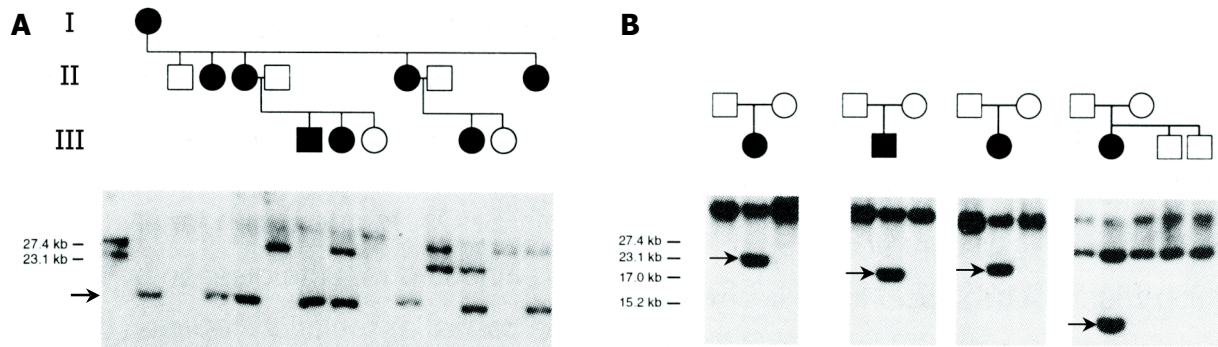


Figure 1.2

(A) Molecular analysis of an FSHD family by linear gel electrophoresis of DNA digested with *Eco*RI followed by Southern blot analysis and hybridisation with probe p13E-11. In this FSHD family, transmission of an 18 kb *Eco*RI fragment is observed through three generations. This fragment, which is highlighted by an arrow, is transmitted from the affected grandmother (generation I) to four affected children (generation II) and three affected grandchildren (generation III). All affected individuals are indicated by black symbols (Adapted from [419]).

(B) Molecular analysis of four families by linear gel electrophoresis of DNA digested with *Eco*RI followed by Southern blot analysis and hybridisation with probe p13E-11. In each patient (black symbol), a new short fragment is present that is absent in both unaffected parents. This *de novo* fragment is indicated with an arrow (Adapted from [419]).

This *Eco*RI fragment observed with probe p13E-11 contained a polymorphic repeat array and accurate sizing indicated size differences in the number of repeat units between unrelated FSHD families. Healthy individuals carried an *Eco*RI fragment ranging between 40 to over 300 kb, while the observed repeat array in FSHD patients was reduced to an *Eco*RI fragment of 10-28 kb [419]. Nowadays, almost all patients (95%) have been linked to this locus on chromosome 4. The residual *Eco*RI fragment size observed in these patients is smaller than 38 kb [231]. A minority of FSHD patients (5%), called phenotypic FSHD patients, who display all phenotypic characteristics of the disease, do not show a contracted 4q35 fragment after DNA analysis nor are linked to the identified chromosome 4 locus [49, 56, 138, 171, 205, 253, 255, 379, 390, 394, 413, 420, 444]. Despite all efforts, a second locus for all these patients has not yet been identified [15, 136, 171, 364], although recently some evidence was presented for linkage to chromosome 15 in one family only [322].

1.2.2 Analysis of the FSHD-associated D4Z4 repeat

The observed *Eco*RI fragment contains a repeat structure, termed D4Z4, consisting of *Kpn*I units ordered in a head-to-tail fashion [88, 159, 419]. In healthy individuals this repeat array can contain from 11 up to more than 100 repeat units. In FSHD patients this repeat is contracted, which results in an array of 1-10 repeat units [88, 159, 419]. Therefore, the assumption was made that contraction of D4Z4 would lead to a partial or complete deletion of a gene. The most logical step then was to sequence the D4Z4 repeat units [123, 159, 226, 426]. A complete

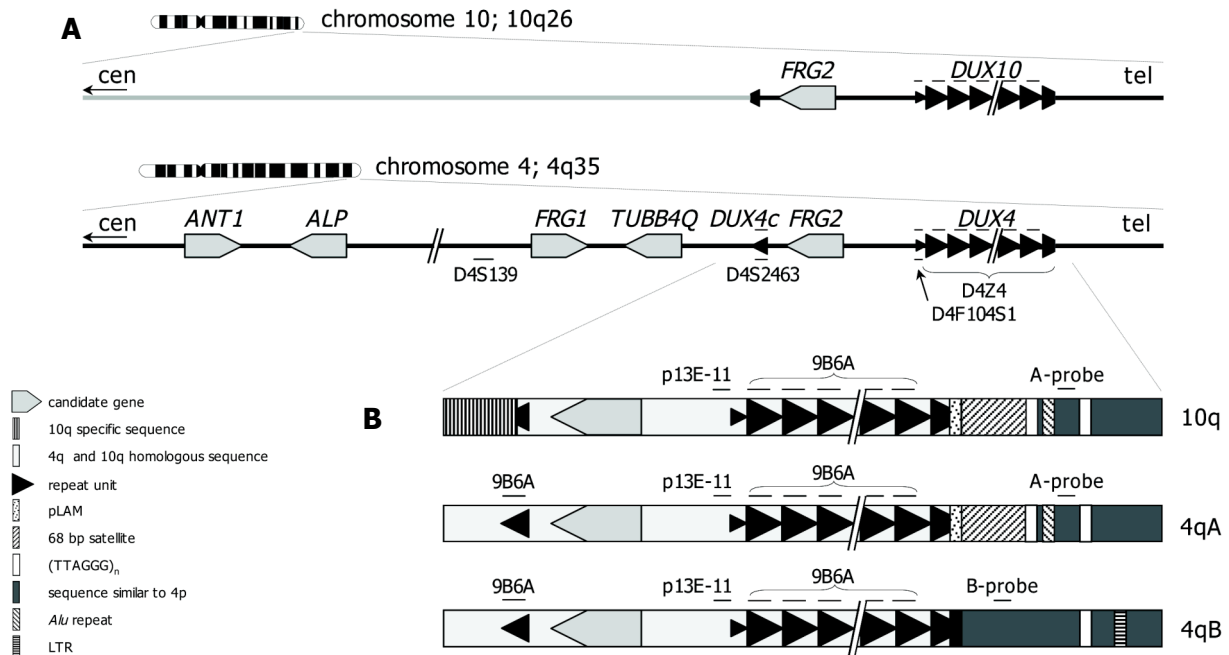


Figure 1.3

Schematic representation of the FSHD-associated region at 4q35 and its homologous region on 10q26 (not drawn to scale). The relative position of three markers in the 4q35 region (i.e. D4S139, D4S2463 and D4F104S1) and the repeats on both chromosomes are visualised in (A). This part also shows the relative positions of the candidate genes on 4q35 and 10q26. (B) shows an enlarged section of (A) containing the sequence elements and regions of sequence similarity between the subtelomeres of 10q, 4qA and 4qB. Also indicated are the regions recognised by the various probes used for hybridisation after gel electrophoresis of DNA digested with proper restriction enzymes for 4qter and 10qter and Southern blot analysis. cen: centromere, tel: telomere. See keys for further details.

overview of the FSHD region with all currently available information, including surrounding sequences, candidate genes and probes, is schematically depicted in *Figure 1.3*.

1.2.2.1 D4Z4 repeat unit

D4Z4 is composed of tandem arranged repeat units that are 3.3 kb in size and consists of both highly conserved and variable regions [159, 226, 419]. Because the sequence is very GC-rich, it has characteristics of a CpG island [159, 433] (see also *Paragraph 1.5.2.2*). Several regions within the sequence show homology with other sequences in the human genome, like a 316 bp region with similarity to the GC-rich repeat *LSau* [2] and an extreme GC-rich DNA region of 461 bp that shows homology with the human low copy repeat *hhspm3* [446]. Throughout the repeat unit several microsatellites were identified as well [225, 226]. Distal to these GC-rich sequences, an open reading frame (ORF) of 405 bp was identified, now designated *DUX4*, that did not span the entire repeat unit and encodes a putative protein with two homeodomains [123, 159, 225, 226, 256, 426]. Homeodomains are DNA binding domains that regulate gene expression during embryonic development [292] and usually encode regulatory transcription factors [1, 68, 78]. The observed homeodomains in D4Z4 show similarity to other

homeodomain sequences, like the human paired box gene family that includes *PAX3*, *PAX6*, the homeobox family gene *OTX1* and the muscle-specific homeogene *PMX1*, *Xenopus HmixX* and *Drosophila paired* and *Hmpr D*. [225, 226].

EST databases and human genomic phage DNA and cDNA libraries were screened for *DUX4* transcripts, but none of the transcripts detected mapped to 4q35 suggesting *DUX4* is transcriptionally inactive [159, 225, 226, 426]. Sequences homologous to *DUX4* were identified and some were named *DUX1*, *DUX2*, *DUX3*, *DUX5* and *DUX10*. Gene expression has been detected, but all these sequences originate from regions other than chromosome 4 [25, 73, 90, 123, 159, 256]. Recently, some evidence has been reported for protein expression of a *DUX* gene exclusively observed in FSHD myoblasts and needs further investigation [73].

1.2.2.2 *D4Z4 proximal and distal sequences*

The 161 kb region proximal to the D4Z4 repeat was sequenced completely. Apart from a large number of repeat sequences, like L1 repeats, long terminal repeat transposons, *Alu* repeats and long interspersed nuclear elements (see also *Paragraph 1.3.1*), this region also contains sequences with (partial) homology to genes located on other chromosomes. Various computer software programs for sequence annotation were used to detect potential coding regions in the 4q35 region and identified, in a region spanning 5 Mb proximal to D4Z4, five candidate genes, *ANT1*, *ALP*, *FRG1*, *TUBB4Q* and *FRG2*, which will be discussed in more detail in *Paragraph 1.2.4*.

The region distal to the repeat array is difficult to clone and therefore has only been partially sequenced. Sequence data derived from a clone that spans 11 kb distal to D4Z4 uncovered only highly repetitive elements and pseudogenes [128] and no transcribed regions have been identified yet [129]. Recent analysis of the distal sequences revealed two variants of the 4qter sequence, designated 4qA and 4qB. Although both variants are almost equally present in the population, FSHD is uniquely associated with the 4qA variant [128, 228]. Variant 4qA contains a 6.2 kb region of β -satellite DNA [128] that consists of 68 bp *Sau3A* monomers [88] and a 1 kb divergent (TTAGGG)_n array, both features that are not present on the 4qB variant [128]. Furthermore, the terminal D4Z4 in the 4qB variant contains only 570 bp of a complete repeat, while variant 4qA carries a divergent repeat called pLAM that shows high homology with *LSau* and *hhspm3* sequences, but does not contain homeobox sequences [128]. Consequently, even though this area was initially discarded for further analysis, these findings now emphasise the importance of this distal part for future research.

1.2.2.3 Analysis of D4Z4 in the genome

Fluorescence *in situ* hybridisation (FISH) analysis located the repeat array next to the telomere of 4q, less than 215 kb from the telomere sequence [426]. FISH-probes containing D4Z4 sequences revealed (weak) hybridising signals throughout the entire human genome: 4q35, 10q26, the Y-chromosome and all short arms of acrocentric chromosomes 13, 14, 15, 21 and 22 as well as on chromosomes 3 and 8, chromosomal locations 1q12, 1p12, 2p11, 9q12, 16p11, centromeres of chromosomes 9, 10 and 20 and the centromere plus region q11 of chromosome 10 [6, 16, 159, 163, 256, 268, 317, 419, 426]. Furthermore, probes containing the homeodomain sequences, such as 9B6A, hybridised to each copy of D4Z4 and also to a 2.5 kb truncated and inverted copy 40 kb proximal to the D4Z4 array that contains homeobox sequences similar to those of D4Z4 using Southern analysis [433].

Due to spreading of homologous D4Z4 sequences through the genome, it was suggested that D4Z4-related sequences are part of a 3.3 kb repeat family with two different structures: tandem arranged repeats, like D4Z4, or repeat clusters interspaced by β -satellite repeats. Family members are organised into subfamilies and located at heterochromatic regions in the genome, mainly on the acrocentric chromosomes and partially interspersed with ribosomal RNA gene clusters [256, 427]. These regions seem to have a different organisation per locus as a consequence of inter- and intrachromosomal recombination events. Because no transcript of D4Z4 has yet been isolated from muscle or any other tissue, this sequence may have been amplified along with other repetitive sequences and scattered through the genome [256].

1.2.2.4 D4Z4 in other species

The evolutionary conservation of D4Z4 was studied in a variety of species by the use of zoo-blots. *DUX4*-derived probes cross-hybridised to a few sequences of DNA from baboon, chicken, cow, goat and pig, but no probes from D4Z4 hybridised with rodent DNA [159]. A Southern blot containing DNA from several Old World and New World monkeys did show signal with probe 9B6A recognising the homeoboxes of *DUX4* [159], which suggests that D4Z4 is primate-specific.

4q35-like sequences are localised at multiple loci in Old World simians (i.e. Catarrhini) [16, 70, 126, 159, 427], which include apes and humans (Hominoidea) and Old World monkeys (Ceropithecoidae) [259]. In macaques (Ceropithecoidae) and marmoset (Ceboidea/New World Monkeys) a reduced number of loci is detected [70]. This suggests that duplications of the 4q35 region have occurred frequently in these primates [16, 126, 427]. FISH analyses showed conservation of the subtelomeric 4q localisation in the analysed primates and this region therefore contains the ancestral copy of D4Z4 [16, 256, 427]. Probably the ancestral

4q region has been duplicated and transposed, followed by genome-specific deletion and expansion [16].

Recently, it was possible to identify potential homologues of D4Z4 in mouse and rat by applying computational analysis of draft genome sequences [69]. Preliminary data revealed that the sequence and organisation of the mouse D4Z4 shows a 4.9 kb repeat and contains an ORF of 2 kb encoding two homeodomains with 55% homology on amino acid levels when compared to those encoded in the human D4Z4. Outside the coding region there is no sequence conservation observed between human and mouse. Furthermore, the mouse repeats are arranged in a large tandem array and are concentrated at one single locus in the mouse genome, the location of which still needs to be determined. The identified rat homologue of D4Z4 also encodes a potential homeodomain protein with 66% similarity on amino acid levels to the mouse protein [69]. These new findings weaken the D4Z4 primate-specific idea and may provide possibilities for the development of an animal model for FSHD.

1.2.3 Complications due to D4Z4 sequence homology

The disease-associated fragment can be visualised with probe p13E-11. Unfortunately, this is not a single-copy probe and detects other loci as well. In the human genome, it recognises a 9.4 kb fragment only observed in males, which corresponds to a sequence on the Y-chromosome [392], and two additional highly polymorphic loci. One of these loci is the repeat array on chromosome 4q35 [419], while the other is a repeat array located on the subtelomere of chromosome 10q (10q26) with a similar organisation as the array on 4q35 [15, 82], which complicates DNA diagnosis. The organisation of the 10q26 region is also depicted in *Figure 1.3*. Thus, probe p13E-11 identifies a total of four polymorphic *EcoRI* fragments in females and five fragments in males, of which the Y fragment is not polymorphic. Usually linear gel electrophoresis is used to separate the FSHD allele from the larger alleles. Visualisation of all D4Z4 fragments of chromosomes 4 and 10 can only be achieved by pulsed-field gel electrophoresis (PFGE) or field inversion gel electrophoresis (FIGE) [232, 354].

1.2.3.1 Restriction sites discriminate between chromosomes 4 and 10

Detailed sequence analysis demonstrated the high homology of the repeats on 4q35 and 10q26 [82]. Comparing the 4q35 and 10q26 sequences also yielded a unique *BlnI* restriction site that was only present in repeat units originating from chromosome 10 [83]. This finding allowed discrimination of these two repeat arrays and subsequently resulted in more accurate diagnoses, because a contracted fragment on 10q26 could now easily be discriminated from a contracted fragment on chromosome 4 [15, 198]. Six years later, a second sequence difference was observed. A restriction site for *XapI* located in the chromosome 4 repeat sequence resulted in

complementary fragments to those detected with *BlnI* [227]. With these two enzymes it is now possible to obtain complete allele information and to deduce the chromosomal origin of the observed *EcoRI* fragment in most cases. The complete procedure to separate both sequences by the use of different restriction enzymes and probes was recently described in detail [231].

As a consequence of the high sequence homology and the subtelomeric location of the repeat arrays, translocations of repeat units between 4q35 and 10q26 are observed in 20% of individuals [86, 295]. This issue will be further addressed in *Paragraph 1.3*.

1.2.3.2 A second FSHD locus?

The identification of a chromosome 10 sequence hybridising with probe p13E-11 raised the question whether this sequence perhaps represented the second FSHD locus, since no linkage to chromosome 4 has been observed in 5% of FSHD patients [15, 136, 171, 322, 364]. Linkage analysis was applied to test the involvement of this locus in two large non-4q35 linked families, but this did not reveal evidence for linkage to chromosome 10qter [15, 364].

In some of the FSHD families without linkage to 4q35, markers on chromosome 12 were also tested flanking the regions of two diseases with clinical features similar to FSHD, i.e. scapuloperoneal muscular dystrophy and scapuloperoneal muscular atrophy [379]. Both disease regions were excluded by extensive linkage analysis [379]. In addition, mutations in the gene myotilin, which has been identified for one of the autosomal dominant forms of limb girdle muscle dystrophy, were also eliminated as a cause of non-4q linked FSHD [156].

Further genomic screening has identified a region on chromosome 15 consistent with linkage (peak lod score 3.20) in one non-4q linked FSHD family [322]. Sequence analysis ruled out a possible candidate gene, *POLG* on 15q25 [22, 322], in which mutations are responsible for progressive external ophthalmoplegia [285]. Other possible candidates in this region were also evaluated [22], but no mutations were observed in desmuslin, an intermediate filament protein that may play an important role in maintaining muscle integrity [272, 285] nor in chromodomain helicase DNA binding protein 2 that may be involved in chromatin structure regulation and gene transcription [285, 432]. The sequences of three proteins recently identified to bind D4Z4 (YY1, nucleolin and HMGB2) [121], were also excluded as candidate genes [22]. Currently other potential candidate regions are under investigation [21, 22].

1.2.4 Identification of 4q35 genes

Extensive analysis of the D4Z4 repeat together with the findings that *DUX4* seems to be transcriptionally inactive and that haploinsufficiency of the 4q region including D4Z4 does not cause FSHD [386], made the assumption that one or more repeat units would comprise the FSHD gene unlikely. With the identification of several 4q35-derived genomic clones it was

possible to extend sequence analyses in the regions adjacent to D4Z4, but the detection of 4q35 transcripts remained very difficult due to the presence of many repeat sequences and the spreading of (pseudogene-) sequences [84, 129, 158]. Despite sequencing difficulties, several genes were identified and characterised in the sequence proximal to D4Z4. Whether or not (some of) these genes on 4q35 indeed play a role in FSHD pathology is still not clear. In addition to *DUX4*, five genes have been studied extensively and will be discussed here in more detail.

1.2.4.1 *ANT1*; adenine nucleotide translocator gene 1

ANT is an integral protein of the inner mitochondrial membrane, organised in homodimers, with a single binding site for adenosine triphosphate (ATP) and adenosine diphosphate (ADP). ANT is responsible for the exchange of mitochondrial ATP for cytosolic ADP across the inner mitochondrial membrane. Since ANT is the only mitochondrial translocase for nucleotides, it is an important link between energy-producing and energy-consuming processes [93, 161]. ANT can also be converted into a pro-apoptotic pore and plays a significant role within the regulation of mitochondrial membrane permeability during apoptosis under the control of multiple apoptosis modulators [26]. In humans at least three different full length cDNAs for ANT have been detected so far: ANT1 of which the gene is located on chromosome 4, ANT2 derived from the X-chromosome and ANT3 that is transcribed from both the Y-chromosome and the inactive and active X-chromosomes [351, 363].

ANT1 maps approximately 5 Mb proximal to D4Z4 [102, 155, 237, 421] and the highly abundant protein is expressed in post-mitotic cells, like differentiated tissues such as skeletal muscle, heart and brain [26, 93, 237]. Mutations in *ANT1* are associated with autosomal dominant progressive external ophthalmoplegia in 11% of patients [161, 285]. This gene was an interesting candidate for FSHD, because of its function and expression in skeletal muscle. However, sequence analysis did not reveal any differences between patients and control individuals [155] and Wijmenga *et al.* [421] discarded this gene as a true candidate gene, because of its distance from the FSHD region. Recently, using radioactive PCR and immunoblotting *ANT1* upregulation was demonstrated in FSHD muscle [121, 222], although this was not confirmed by quantitative real-time PCR [177] or expression profiling [428]. With these contradictory results, possibly due to applying different methods, *ANT1* involvement in FSHD remains elusive.

1.2.4.2 *ALP*; actinin-associated LIM protein gene

By screening a human muscle cDNA array a novel muscle-specific transcript, ALP, was identified that mapped to the FSHD region [44, 436]. Thereafter, a detailed study was

performed to evaluate *ALP* as being the FSHD gene [44]. *ALP* is expressed at high levels in differentiated skeletal muscle and an alternatively spliced form of the protein is also detected at low levels in heart. The PDZ domain of *ALP* interacts with the spectrin-like repeats of α -actinin-2 at the Z-lines of myofibres [436]. In two studies, no significant differences between FSHD patients and control biopsies were observed in protein sizes, expression levels and subcellular localisation [44, 436]. Also microarray data did not reveal an altered gene expression [428], which excludes *ALP* as a possible candidate gene for FSHD.

1.2.4.3 *FRG1*; FSHD region gene 1

FRG1 is located 120 kb proximal of the D4Z4 repeat array and has been identified through its association with a CpG island [87, 422, 433]. The gene has multiple related copies in the human genome with a minimum of seven locations that are in part expressed pseudogenes [147]. The protein is highly conserved in vertebrates and invertebrates, is ubiquitously expressed and localises to nucleoli, Cajal bodies and speckles after transient and stable transfection [87, 201]. Nucleoli are mainly involved in RNA ribogenesis, nuclear export of a subset of mRNAs and maturation of small nuclear ribonuclear proteins [42, 267, 314, 350, 361], while Cajal bodies are thought to be involved in the post-transcriptional modification of small nucleolar RNAs and small nuclear RNAs, and in shuttling small nuclear ribonuclear proteins from the nucleoplasm to the nucleus [40, 41, 59, 79, 286, 361, 362, 400]. Speckles may coordinate transcription and RNA processing, and may also be a storage site for protein splicing factors and small nuclear ribonuclear proteins [91]. The localisation of *FRG1* could therefore imply a functional role in RNA processing [201]. Attempts to detect differences in allele-specific transcription of *FRG1* between patients and controls using reverse transcribed RNA isolated from muscle and lymphocytes failed [87], but recently upregulation of gene expression was observed in FSHD muscle biopsies ([121] and T Rijkers, unpublished results). In contrast, Jiang *et al.* [177] observed a modest, but significant decrease of transcription levels in FSHD muscle samples by using real-time PCR and recently generated microarray data does not show expression changes [428]. However, in the latter it is not possible to discriminate between expression from different copies of *FRG1* in the genome [428]. Even though a reason for the discrepancy in all data is still unclear, this gene remains a candidate.

1.2.4.4 *TUBB4Q*; tubulin beta polypeptide 4 member Q gene

Exon-trapping experiments revealed an exon-trap clone 80 kb proximal to D4Z4 [127]. Sequence analysis identified a transcript that contains four exons encoding an ORF of 434 amino acids. This putative gene sequence, designated *TUBB4Q*, is highly homologous to β 2-tubulin, a member of a large tubulin family with copies on many chromosomes and maps to

various regions in the genome (see also *Paragraph 1.3.4.1*) [129]. The tubulin proteins form microtubuli in eukaryotic cells and are associated with chromosome division, cellular movement, cell polarity, cytoskeleton integrity and intracellular vesicle transport [150, 191]. The putative protein sequence indicates a truncated protein and displays amino acid substitutions in functional protein domains that are highly conserved in the β -tubulin family. Furthermore, allelic sequence variation with the most relevant polymorphism at the initiation codon is detected, which is remarkable for a highly conserved β -tubulin protein. These observations suggest that this gene might be inactive [126, 127]. Despite efforts to detect *TUBB4Q* expression no transcripts have been observed in fetal and adult tissues, suggesting that this copy of the β -tubulin homologues is indeed an inactive pseudogene [85, 127].

1.2.4.5 *FRG2*; *FSHD region gene 2*

Computer prediction algorithms identified a novel transcript 37 kb proximal to the D4Z4 repeat array, called *FRG2* [334, 335]. This gene has a putative muscle-specific promoter and generates a transcript of 2 kb, encoding a putative protein of 278 amino acids. Related copies of this gene are detected on multiple chromosomes, mostly in subtelomeric or pericentromeric regions. One of them is located on chromosome 10, which is highly homologous to the chromosome 4 copy. Furthermore, luciferase reporter assays indicated that increasing numbers of D4Z4 repeats inhibit *FRG2* promoter activity. Transient transfection experiments revealed a nuclear localisation of the encoded protein, in which overexpression of *FRG2* apparently causes morphological changes in all cell lines tested. Apart from constitutive active expression in the rodent monochromosomal chromosome 4 cell line GM11687, *FRG2* expression from chromosomes 4 and 10 is only observed in primary myoblasts of FSHD patients upon differentiation and in fibroblasts of FSHD patients and control individuals that are forced into myogenesis with an adenovirus expressing MyoD. *FRG2* transcripts detected in these cell lines were mainly derived from chromosome 10 and to a lesser extent from chromosome 4. The expression detected in myoblasts upon differentiation from non-FSHD myopathies is derived from *FRG2*-related copies on chromosome 3 or 22 [334, 335]. Recently, expression was detected in FSHD muscle biopsies [121], but Rijkers *et al.* [334, 335] could not confirm *FRG2* expression in muscle biopsies, neither in proliferating myoblasts, fibroblasts, peripheral blood lymphocytes or brain tissue.

The *FRG2* genes on chromosomes 4 and 10 are attractive candidate genes for FSHD, because of their location, the D4Z4-sensitive *FRG2* promoter and the observed expression in differentiating FSHD myoblasts and fibroblasts that are forced into myogenesis [334, 335]. Recently, it was suggested that as a consequence of D4Z4 repeat contraction a disturbed interaction of 4qter with the nuclear periphery may result in aberrant gene expression [260].

Furthermore, this contraction can result in a local chromatin change [296]. Both features could lead to *FRG2* expression in *cis* on chromosome 4 [334, 335]. Activation of *FRG2* may also occur in *trans* for the chromosome 10 copy [334, 335], possibly via transvection, i.e. the ability of a locus to influence the activity of another allele in *trans* (reviewed in [318]). This process is possible during chromosomal pairing in interphase, which is also observed for chromosomes 4 and 10 [367].

Lemmers and colleagues [229] described a healthy father with a normal D4Z4 repeat array, but lacking one of the *FRG2* copies due to a deletion in the 4q35 region. His affected son inherited this chromosome lacking *FRG2* in combination with a contraction of D4Z4 to three repeat units. Recently, also two other families with p13E-11-associated deletions were observed [230]. Although these results challenge a possible role for *FRG2* in FSHD pathology, expression studies in these individuals may provide support for the proposed transvection mechanism for *FRG2* [334].

1.2.4.6 Detection of gene expression

Since the disease-causing mechanism of FSHD has still not been elucidated, several research groups decided to shift their focus to the detection of possible molecular changes as a consequence of D4Z4 contraction. A PCR-based subtractive hybridisation was used to study global differences in mRNA expression patterns in muscle from FSHD patients and healthy control individuals to identify genes involved in FSHD muscle [387]. With this method a global alteration of gene expression specific for FSHD muscle was observed in a significant number of muscle-specific genes and genes encoding transcription regulators. The hypothesis that the contraction of the D4Z4 repeat array initiates transcriptional deregulation by a positional effect (see *Paragraph 1.6.2*) was then further investigated using human muscle samples and revealed an FSHD muscle-specific overexpression of *ANT1*, *FRG1* and *FRG2* [121]. Furthermore, this study discovered an element in D4Z4 that binds a multiprotein complex, termed D4Z4 recognition complex, which includes the transcription activator and/or repressor protein YY1, DNA helicase-containing nucleolin and HMGB2, a protein involved in chromatin architecture. This complex can bind to D4Z4 sequences *in vitro* and *in vivo* in HeLa cells and in a rodent-human monochromosomal cell hybrid containing a single human chromosome 4 [121]. However, this has not yet been confirmed in myoblasts or muscle extracts. Furthermore, reducing the levels of the proteins in this D4Z4 recognition complex in cell culture resulted in *FRG2* transcriptional upregulation. Gabellini *et al.* [121] therefore suggested that genes located at 4q35 would be inappropriately expressed in FSHD patients as a consequence of a reduced number of D4Z4 recognition complexes.

A global gene expression profile of mature FSHD muscle tissue was generated using microarrays to gain more insight in disease-specific muscle changes [428]. The profile contained altered genes that are involved in cell cycle control and cellular differentiation and proliferation. Also many observed deregulated genes are associated with myogenesis and direct targets of the transcription factor MyoD. This may suggest a defect in FSHD muscle cell differentiation, such as an inefficient completion of the myogenic program. Genes that confer a reduced capacity to buffer oxidative stress are deregulated as well. However, none of the 4q35 genes show an altered expression [428], which was also observed by other groups [44, 87, 97], but contrasts with the findings of Gabellini *et al.* [121]. The reason for this discrepancy remains to be elucidated.

1.2.5 *In vitro* cell culture systems

Because culturing of muscle cells *in vitro* may provide useful information, several groups explored cell culture techniques to clarify aberrant growth patterns and possible altered pathways [109, 346, 397, 428, 440].

FSHD myoblasts studied in culture present a necrotic phenotype displaying large vacuoles and swelling of the nucleus and cytoplasm, which could be due to a degradative process [108, 428]. This morphology is likely mediated by altered expression of extracellular matrix components. These myoblasts display a decreased replicative capacity, possibly because the normal cell cycle is disrupted or cells approach senescence [108]. Furthermore, FSHD myoblasts show an increased susceptibility to oxidative stress (induced with paraquat) before they differentiate and fuse into myotubes [428]. In addition, FSHD myoblasts fuse much faster than control myoblasts when exposed to differentiation media due to an upregulation of MyoD [108, 109]. These growth processes may contribute to the observed cellular morphology and eventually to a muscle disease. Unfortunately, none of the features observed in culture are detected in FSHD muscle biopsies (see *Paragraph 1.1.2.3* for a description on muscle biopsy characteristics), but do suggest that a deviant gene expression may occur in the first phases of skeletal muscle development, i.e. in undifferentiated myocytes [108, 428]. This altered myogenic differentiation process was supported by the results obtained from GeneChip expression profiling of mature FSHD muscle mentioned above (*Paragraph 1.2.4.6*) [425].

1.3 Organisation of repeats and subtelomeric regions in the genome

As FSHD is associated with the D4Z4 repeat located in the subtelomeric region of chromosome 4q, a description on several repeat characteristics and the consequences when repeats are involved in human disease is given below, followed by a characterisation of subtelomeric regions. The observed D4Z4 behaviour in relation to FSHD will then be discussed in more detail.

1.3.1 Repetitive elements in the genome

A substantial fraction of the eukaryotic genome, at least 50% [217], consists of repetitive sequences of various sizes and composition that can occur in tandem, inverted and dispersed organisations [166], organised as clusters in specific chromosomal regions or randomly spread throughout the genome [240, 373]. Repetitive DNA components consist of families of sequences that are related to each other, but not identical [234]. There are various classes of repetitive DNA in the human genome [53, 94, 217, 233, 235, 368, 373]: (1) telomere repeats (TTAGGG)_n located at the ends of chromosomes that have a size of 5-12 kb; (2) subtelomeric repeats, classes of repetitive sequences that are interspersed in the last 500 kb of non-repetitive DNA located adjacent to the telomere. Some of these repeat sequences are chromosome-specific, while others can be observed near the ends of all chromosomes; (3) microsatellite repeats (also called simple tandem repeats or short tandem repeats), which are common in the human genome in a variety of simple mono-, di-, tri-, tetra-, and penta-nucleotide tandem repeats that are dispersed in the euchromatic arms of most chromosomes. Their whole repetitive region spans less than 150 bp. The dinucleotide repeat is the most common with a size of 20-60 bp and a total copy number of 100000, occurring approximately every 30 kb in the genome. Many of these sequences are polymorphic in copy number; (4) minisatellite repeats, a class of dispersed tandem repeats in which the repeating unit is on average 25-35 bp that has a variable sequence, but does contain a 10-15 bp core sequence. These satellites range in size from 200 bp up to several kb and have a lower copy number than the microsatellites. These repeats tend to occur in greater numbers towards the ends of chromosomes. One type of minisatellite repeat is the variable number of tandem repeats (VNTR); (5) Alpha satellite DNA (or alphoid DNA), a family of related repeats that occur as long tandem arrays consisting of two dimers of each approximately 170 bp long at both sides of the centromeres of all human chromosomes and extend over a 1-5 kb region; (6) Satellite I, II and III repeats, of which type I is AT-rich and composed of alternating arrays of 17-25 bp repeat units, while satellite II and III repeats are both derived from the repeat unit ATTCC. The size may vary from 100 kb to over 1 Mb. All types occur as long tandem arrays in heterochromatic regions of chromosomes 1, 9, 16,

17 and Y and in satellite regions on chromosomes 13p, 14p, 15p, 21p and 22p; (7) mega- or macrosatellite DNA, which is characterised by array sizes that can be modest compared to some satellite DNA arrays. The prefix “mega-“ or “macro-“ has been used to emphasize the large size of the repeating unit that can be several kb; (8) Cot1 DNA is a fraction of repetitive DNA that contains sequences that have copy numbers of 10000 or more; (9) A large category of transposon-derived interspersed repeats that can be divided in four different sub-groups: (a) *Alu* repeats, which are the most common interspersed repeats in the human genome, are part of the non-viral superfamily consisting short interspersed nuclear elements (SINEs). This 100-400 bp sequence occurs on average once every 3300 bp with a total copy number of one million copies. Another major human SINE family is the mammalian-wide interspersed repeat family that is present in approximately 2% of the genome; (b) L1 repeats are long interspersed nuclear elements (LINEs) originating from a viral super family that are 1-7 kb in length. These repeats occur on average every 28 kb in the genome with a total copy number of about 500000; (c) Long terminal repeat retroposons (LTRs), which are autonomous interspersed elements. Although a variety of LTRs exist, only the vertebrate-specific endogenous retroviruses appear to have been active in the mammalian genome; and (d) DNA transposons, which are also interspersed repeats. This group resembles bacterial transposons, of which the human genome contains at least seven different classes.

About 45% of the human repeat sequence belongs to the last category (category 9), with SINEs and LINEs for 34%, LTRs for 8% and DNA transposons for 3% of the sequence [53, 94, 217, 235, 368, 373]. Two percent of the total repetitive sequences encode multigene families, usually arranged in tandem arrays [241, 242, 311]. This extensive repetition of a gene may occur when large amounts of the gene product are required [233], like those for ribosomal RNAs and U1 and U2 small nuclear RNAs [241, 242, 311]. However, the majority of these repetitive sequences seems to be non-coding [217].

1.3.2 Mechanisms of repeat evolution

Repetitive sequences in all organisms can evolve in a concerted fashion to maintain sequence homogeneity within a repeat array sequence. This is a universal biological phenomenon, described as a complex process that leads to homogenisation of DNA sequences that belong to a given repetitive family containing both homologous and non-homologous sequences [240]. Homogenisation occurs by unequal crossover, gene conversion or sister chromatid exchange of the repeat array and implies that individual repeats of a tandem array are very similar or even identical within a species, but deviate significantly from the orthologous repeats of closely related species ([241] and all references therein). To account for the sequence homogeneity, two distinct complex processes operating at different rates are involved: intrachromosomal and

interchromosomal exchanges [240, 241]. Intrachromosomal events homogenise individual repeat arrays within a single lineage completely, while interchromosomal exchanges will homogenise all repeat arrays within the population. Homogenisation within an array occurs at a much higher rate than between arrays [241]. A good example of a repeat array that evolves in this way is the *RNU2* locus encoding human U2 small nuclear RNA [241, 242]. Though not all repetitive multigene families undergo concerted evolution, as is demonstrated for the major histocompatibility complex and immunoglobulin multigene families [240, 289]. There is also evidence that concerted evolution of repetitive non-coding DNA sequences is possible and may work via similar mechanisms [99, 240]. However, repetitive non-coding DNA can also evolve via recent amplification and transposition of the repetitive sequence [240].

1.3.3 Repeat sequences associated with disease

Recombination is a key event in the genome that occurs during meiosis. This mechanism, which is most likely to occur between homologous DNA sequences, takes place approximately 1 or 2 times per chromosome during a meiotic division and is essential for both survival and evolution of a species as it generates genetic diversity [132]. However, when chromosomes are misaligned to homologous sequences, these recombination processes can, for example, result in rearranged gene clusters, but also in damaging mutations and disease [132, 320]. The existence of repeated sequences make such unequal crossover events possible.

1.3.3.1 Low Copy Repeats

Low copy repeats, which are present in approximately 5% of the human genome, influence the stability of the genome. By non-allelic homologous recombination between these repeats structural rearrangements may occur, resulting in deletions, duplications, inversions and translocations and may eventually result in genomic disorders (reviewed in [365]). For example, Charcot-Marie-Tooth disease type 1A and hereditary neuropathy with liability to pressure palsies are both caused by non-allelic homologous recombination between flanking low copy repeats in which a 1.4 Mb duplication on 17p12 causes Charcot-Marie-Tooth disease type 1A, while deletion of this fragment results in hereditary neuropathy with liability to pressure palsies. Some other examples are Williams syndrome in which most patients have a *de novo* deletion of 1.6 Mb of chromosome 7q11.23, hemophilia A in which 45% of males carry a 500 kb inversion of Xq, and two syndromes linked to imprinting genes, Prader-Willie syndrome and Angelman syndrome, that are both (in most cases) due to a 4 Mb deletion of 15q11-13 on the inherited paternal chromosome in the case of the Prader-Willie syndrome or on the maternal chromosome in the Angelman syndrome [71, 288, 365].

1.3.3.2 Trinucleotide repeat expansions

Trinucleotide repeats are simple stretches of DNA (like for example CCG or CAG) that, when expanded, sometimes cause disease [77, 125, 174, 233, 331, 409]. These repeat sequences vary in copy number and become unstable when they increase in size beyond a certain threshold. Several shared characteristics have been identified for diseases associated with these unstable repeat expansions: (1) the mutation manifests as a change in repeat copy number and the mutation rate is related to the initial copy number of the repeat; (2) rare “founder” events will result in alleles that have an increased likelihood of undergoing changes in the number of repeats; (3) a disease that results from a repeat expansion displays a relationship between repeat copies and the severity and/or age at onset of the disease, which means that an earlier age at onset and increasing severity of the disease are correlated with larger repeat size; (4) the origin of the disease allele (i.e. maternal or paternal) will often influence repeat expansion in subsequent generations as paternal transmission carries a greater risk of expansion; and (5) mutated repeats are usually instable both in the germline as in somatic cells. For each disease, the repeat sequence involved, the specific location of the mutation with respect to a gene (or genes) (i.e. 3’ or 5’ untranslated region, coding region or within intron sequence), the threshold for the amount of repeats allowed and the consequences of the loss and/or gain of function of the proteins involved will vary. As a result, this will cause differences in gene expression, cellular specificity and function(s) that may lead to various disease phenotypes [77, 125, 174, 233, 331, 409]. Surprisingly, all published disease phenotypes are primarily related to neurological and neuromuscular dysfunction [125].

The first trinucleotide repeat expansion disease identified is the fragile X syndrome with a repeat expansion located in non-coding sequence [77, 130, 208]. Friedreich ataxia, myotonic dystrophy 1 and 2, two forms of spinocerebellar ataxia (type 8 and type 12) and fragile XE syndrome are also examples of diseases associated with non-coding repeats characterised by large variable expansions, which will cause mental retardation, behavioural abnormalities and eventually multiple tissue dysfunction or degeneration [77, 125, 323].

The CAG repeat expansions, which are located within exons and coding for polyglutamine tracts, are observed in neurodegenerative disorders, like Huntington’s disease, dentatorubral pallidoluysian atrophy, spinobulbar muscular atrophy, and several types of spinocerebellar ataxia that are characterised by progressive neuronal dysfunction in particular brain regions that will begin in mid-life and eventually will result in severe neurodegeneration [77, 239, 331, 409]. These expansions observed are much smaller and show less variation than those observed for non-coding repeat expansions. The role of these expansions appears to be a

complex process with many different components involved, but due to similar phenotypic characteristics common mechanisms will most likely be the cause of these diseases.

The expansions of alanine tracts (GCG, GCA, GCC) (reviewed in [52]) usually do not exceed more than 20 repeats and are all approximately of the same size. They often occur in a gene encoding for a transcription factor. The effect can be a loss-of-function or a gain of an abnormal function and will result in a disturbance of body plan development, probably through an altered expression of downstream genes. Synpolydactyly type II and Cleidocranial dysplasia are examples of these alanine expansions. In contrast, oculopharyngeal muscular dystrophy is the result of an expansion in a gene that does not code for a transcription factor, but for a polyadenine-binding protein. This disease is similar to the polyglutamine expansion diseases on the clinical and molecular level.

The exact cause of these repeat expansions is not known, although it is thought that the DNA secondary structure plays an important role [125]. One proposed mechanism is the DNA slippage model, which predicts that repeat size variability arises during DNA replication in a cell division-dependent manner [332]. The repetitive character of trinucleotide sequences allows them to fold into hairpins, slipped-stranded DNA or more complex configurations [125, 174, 310], which can cause delay and slippage of DNA polymerase [310]. The DNA mismatch repair pathway should repair these structures, when they are formed during DNA replication, to avoid expansions and deletions. However, with large expansions these structures may occur too frequent, and may, consequently, not all be repaired and thus eventually result in repeat size changes [141, 312]. However, since most cell types have recombination and repair mechanisms, the instability of trinucleotide repeats can also arise from gene conversion as a consequence of an unequal recombination or by error-prone DNA repair [216].

Recently, Gomes-Pereira *et al.* [141] proposed a new hypothetical model in which repeats usually replicate in the form of standard duplex DNA, but at some point in the cell cycle this duplex is melted. The repeat sequences then have the opportunity to misalign and form loop-outs, like hairpins or slipped-stranded DNA, on both strands. These structures are targets for repair by mismatch repair components, but in the absence of a strand-specific signal for the correct sequence the mismatch repair machinery incorporates these looped structures and introduces a single stranded nick on the other strand, which could act as a initiation point for nuclease-mediated gap extension. DNA polymerase then fills the gaps and the free ends are joined by DNA ligase to generate a product expanded by the length of the loop-outs. When mismatch repair components are absent, small loop-outs may accumulate and form large loop-outs. These large loops may be detected by another DNA repair mechanism and result in the removal and deletion of the looped repeats [141].

The examples mentioned above show how repeat arrays may behave and can cause disease when mutated. These repetitive DNA sequences can also be beneficial for an organism, like the repeated sequences that allow the rapid production of ribosomal RNA when needed [166, 233]. An excess of repetitive DNA will eventually cost an organism energy when replicating, interferes in chromosomal crossovers, and recombinates and duplicates resulting in a disturbed chromosomal integrity. These sequences are usually located in constitutive heterochromatin [166] and have been mapped to specific chromosomal regions, like centromeres and (sub)telomeres [343].

1.3.4 Subtelomeric regions

Chromosomes of organisms as diverse as yeast, *Drosophila* and human, contain subtelomeric regions, located immediately adjacent to the simple telomeric repeat sequence $(TTAGGG)_n$ [344, 443]. These regions, which are often GC-rich as a consequence of a diversity of GC-rich minisatellites [344], are defined by multichromosomal blocks of sequence that contain a wide variety of repetitive DNA, ranging from low copy interspersed repeats found on a few chromosome ends, to highly repetitive repeat sequences present in many subtelomeric regions [114, 266, 344]. These subtelomeres show comparable features in structure, function and composition of various repeated elements, but the sequences of different elements and sizes of these regions can vary between organisms [54, 266]. Subtelomeric regions are more often involved in recombination processes at chromosome ends than at other parts of the genome [34, 64, 115]. Due to exchanges between subtelomeres of different chromosomes as a consequence of sequence homology, copies can be dispersed throughout the genome [344]. As a result of this dynamic behaviour, the composition of a chromosome may vary distinctly among individuals of the same species [243, 266] and the size of a subtelomere sequence can vary from a few hundred bp, like the human XpYpter [13], to more than 100 kb, like the regions on 4qter [87] and 16pter [423], before unique DNA sequences specific for each chromosome are detected.

1.3.4.1 Function of subtelomeres

Several hypotheses have been proposed for a function of subtelomeric regions. It has been suggested that they could act as a buffer between functional telomeric repeats and genes proximal to subtelomeric regions to protect these genes from heterochromatinisation and silencing [16, 266, 317]. Subtelomeres may also act as a storage area, where genomic sequences that are no longer used can be discarded. Many repetitive sequences, pseudogenes and partial regions of homology to genes that function elsewhere in the genome are located in the

subtelomeres and support this hypothesis [243]. In contrast, subtelomeres can also be seen as functionally active regions and, because they are more prone to recombination, could then serve as a region where new genes are generated and modified [243, 265]. In favour of this hypothesis are genes that are actually transcribed from the subtelomeric sequences. A good example is the olfactory receptor (*OR*) gene family observed at various chromosomal locations at more than 40 sites in the human genome and preferably located subtelomerically, comprising at least 30 sequences [243, 265]. Their behaviour may contribute to the amplification and generation of variation in these *OR* genes, which encode various receptors that allow the recognition of ~10000 odorants [243]. Individuals can carry up to 56 copies of subtelomeric *OR* genes [383], but so far only one gene copy (*OR-A*) shows expression [243]. This gene copy varies per individual in number, sequence and location due to recent duplications of subtelomeric sequences and is actively transcribed from several chromosomal regions. Individuals may express different combinations of the multiple *OR-A* gene copies and therefore display a prominent aspect of gene diversity and phenotypic variation that will eventually allow humans to recognize numerous new odorants and result in differences between individuals [243].

The β -tubulins are also a good example to underline this hypothesis. In humans, β -tubulins are part of a multigene family that consists of several isotypes with many associated pseudogenes. Seven functionally expressing genes have been reported [247]. Van Geel *et al.* [127] described the identification and characterisation a novel β -tubulin gene family. Besides a copy in the subtelomeric location on 4q35, called *TUBB4Q* (see **Paragraph 1.2.4.4** for a detailed description), related sequences were also identified on chromosomes 1, 9, 10, 12, 16, 18 and Y. All these sequences show high homology to functional β -tubulin genes. Analysis of the *TUBB4Q* sequence compared to the seven known functional genes suggests that this sequence is a novel isotype of the β -tubulin family and might even be a new member [127]. A model suggested for tubulin evolution proposes that structural constraints are lifted after gene duplication, which causes a relatively fast protein change [245]. *TUBB4Q* might thus have changed into a pseudogene very recently or is still evolving towards a functional β -tubulin. At this moment, this gene is most likely in dormant or inactive state. Because sequences related to *TUBB4Q* have been identified on other chromosomes as well, these novel β -tubulin family members were presumably distributed through the genome by duplications and rearrangements [127].

1.3.4.2 Organisation of subtelomeres

Flint and colleagues [114] proposed a model for subtelomeric organisation based on the sequences of chromosome ends 4p, 16p and 22q dividing subtelomeric regions in a distal and a

proximal subdomain separated by degenerated (TTAGGG)_n repeats and putative origin-of-replication consensus sequences. The subdomains of 4p, 16p and 22q show homology to the subtelomeric region on 4q, but further to no other chromosome end. The distal subtelomeric region is composed of small repetitive sequences that show interrupted homology of less than 2 kb with many other chromosome ends, while the proximal subdomain only interacts with a subset of chromosome ends and the blocks of uninterrupted homology tend to be longer (10-40 kb). The function of the repeats that separate the two subdomains is not clear yet, but it can be involved in limiting interactions with non-homologous chromosomal regions.

A chromosomal region, like a subtelomere, can increase gene expression and variation in phenotype by its dynamic behaviour. However, subtelomeric deletions and translocations (reciprocal and/or cryptic) can also cause human diseases. Examples are idiopathic mental retardation [116], developmental delay and learning difficulties [408], Wolf-Hirschhorn syndrome [5], Miller-Dieker syndrome [213], α -thalassaemia [115] and FSHD. For the FSHD locus subtelomeric rearrangements of the D4Z4 repeat can even occur with or without pathological consequences [86, 262, 295, 370, 407, 435].

1.3.5 Subtelomeric region 4q35 and its association with FSHD

Due to the spreading of homologous D4Z4 sequences over the entire genome, and especially the presence of a highly homologous and equally recombinogenic repeat array on 10q26 [15, 82], one can expect the opportunity for mispairing between non-homologous chromosomes and suspect exchanges between 4q35 and 10q26 subtelomeres.

When the restriction enzyme *BlnI* is used in combination with *EcoRI*, one expects that two fragments corresponding to chromosome 4 remain visible after hybridisation with probe p13E-11, whereas chromosome 10-derived repeat units are cleaved by *BlnI* [83] (see **Paragraph 1.2.3.1**; reviewed in [231]). Unexpectedly, in a study applying these double digests to 102 Dutch FSHD patients evidence was acquired for exchanges between the repeat arrays of 4q35 and 10q26 [86]. Besides the expected pattern also individuals displaying none, one, three or four *BlnI*-resistant *EcoRI* fragments were encountered. This suggested that chromosome 4-derived repeat units can be present on chromosome 10, and *vice versa*, as schematically depicted in **Figure 1.4.A**. The exchange of the highly homologous repeat units between chromosomes 4 and 10 could have implications for the molecular diagnosis of FSHD, since, as a consequence of these translocations, a contracted chromosome 10-derived repeat array on 4q35 also results in FSHD. To establish the frequency of these translocations in a healthy population a randomly selected panel of 50 Dutch control males was tested and interchromosomal exchanges were then also detected in 20% of all control individuals ($n = 10$) [86].

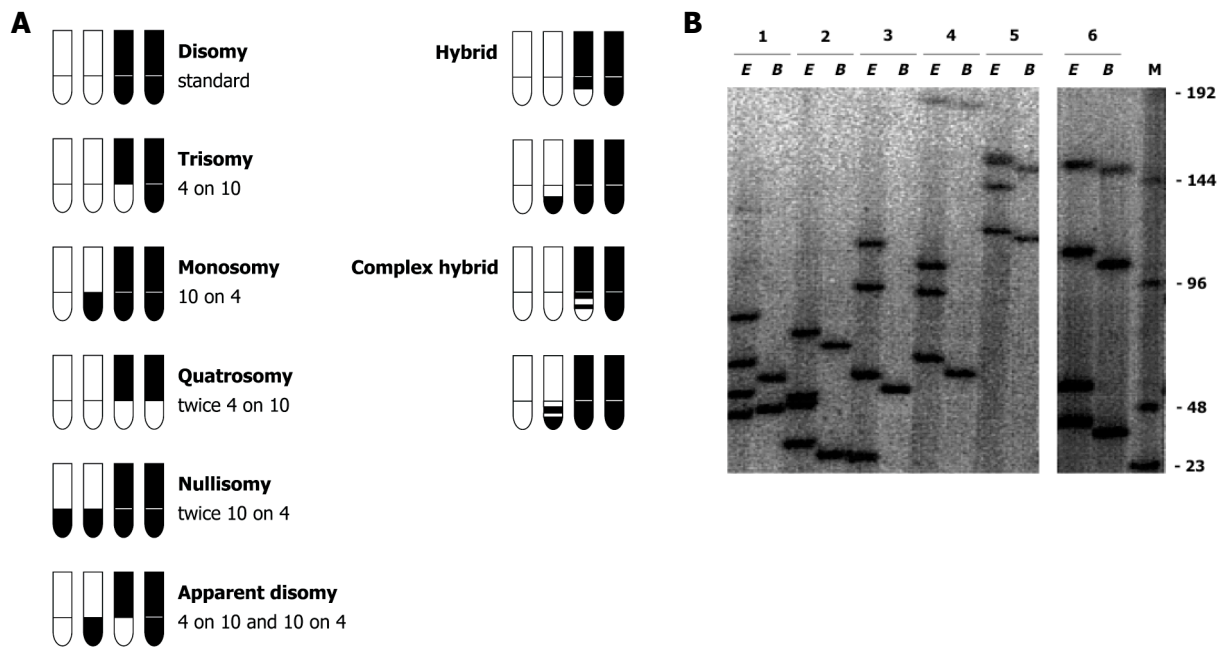


Figure 1.4

(A) Schematic overview of possible exchanges between the subtelomeric regions on chromosomes 4q35 and 10q26. Chromosome 4 sequences are depicted in white, while chromosome 10 sequences are depicted in black. The rounded tip of each chromosome indicates the repeat. Most individuals display a standard repeat array configuration of a chromosome 4 repeat array on chromosome 4 and a chromosome 10 repeat array on chromosome 10 (disomic for D4Z4 repeat, indicated as “disomy”). Exchanges may result in the presence of a chromosome 4-derived repeat array on chromosome 10 (trisomic for D4Z4 repeat; “trisomy”) or a chromosome 10-derived repeat array on chromosome 4 (monosomic for D4Z4 repeat; “monosomy”). The presence of two chromosome 4-derived repeat arrays on chromosome 10 (quatrosomic for D4Z4 repeat; “quatrosomy”) or two chromosome 10-derived repeat arrays on chromosome 4 (nullisomic for D4Z4 repeat; “nullisomy”) may also be observed. The situation in which a chromosome 4-derived repeat array resides on chromosome 10 and a chromosome 10-derived repeat array resides on chromosome 4, termed “apparent disomy”, cannot be discriminated from the standard “disomy” configuration by PFGE of DNA digested with the appropriate restriction enzymes. Furthermore, the repeat arrays on chromosomes 4 or 10 may consist of repeat units derived from both chromosomes, indicating hybrid structures (hybrid and complex hybrid).

(B) Examples of exchanges observed between the repeats on chromosomes 4q35 and 10q26. Molecular analysis of DNA from six individuals by PFGE of DNA digested with *EcoRI* (“E”) or *EcoRI/BlnI* (“B”) followed by Southern blot analysis and hybridisation with probe p13E-11. After *EcoRI* digestion four alleles will be identified: two from chromosome 4 and two from chromosome 10. Treatment with both *EcoRI/BlnI* will fragment chromosome 10 repeat units, while the fragment containing chromosome 4 repeat units remains undigested, apart from digestion at the only *BlnI* restriction site present 3 kb distal to the proximal *EcoRI* site on 4q35. Individuals 1, 2, 4 and 5 display the standard pattern (“disomy”), the most common allele configuration or “apparent disomy”, which is observed in 1% of cases [295]. Individual 3 is an example of “monosomy”, where three repeat arrays are sensitive to *BlnI*. Individual 6 is an example of “trisomy”, where three arrays are resistant to *BlnI*. Marker sizes (“M”; in kb) are indicated on the right. (Adapted from [68]).

1.3.5.1 Identification of complex exchange patterns

To study the structure of the exchanged repeat arrays in these above described 10 control individuals in more detail, the same membranes were rehybridised with probe 9B6A that recognises the D4Z4 repeat instead of the unique region proximal to the repeat array [229]. Thereafter, seven individuals showed a hybridisation pattern consistent with a homogeneous repeat array, which indicates homogeneous chromosome 10-derived repeats on one of the

chromosome 4 alleles or homogeneous chromosome 4-derived repeats on one of the chromosome 10 alleles. In the other three, a more complex hybridisation pattern was obtained, indicating a hybrid structure of the repeat array on one of the chromosomes that consisted of repeat units derived from chromosomes 4 and 10. All possible rearrangements are depicted in *Figure 1.4.A*. Examples of some observed hybridisation patterns probed with probe p13E-11 are depicted in *Figure 1.4.B*.

1.3.5.2 Somatic pairing of 4qter and 10qter

Detailed sequence analysis revealed that the two subtelomeres of 4q and 10q are more than 95% homologous [82], which makes the exchange of genomic material resulting in repeat arrays composed of chromosome 4-derived or chromosome 10-derived repeat units (or a mixture of both) on the other homologous chromosome during meiosis possible. This is supported by the observation that homologous chromosome pairing is initiated in subtelomeric regions during meiosis in humans [19, 61, 404]. The hypothesis that chromosome ends may pair preferentially was further explored by studying spatial orientation in interphase nuclei. Stout and colleagues [367] applied FISH on interphase nuclei to test the hypothesis that chromosome regions located in close proximity will be more often involved in somatic recombinations leading to rearrangements than chromosomes positioned at a greater distance.

The frequency of somatic pairing between subtelomeres was significantly higher than between non-telomeric regions (average of 3.5% versus average of 2.1%) and the pairing frequencies depended to a great extent on specific regions involved. Furthermore, subtelomeric regions tended to replicate asynchronously. The authors then focussed on the 4q subtelomeric region to study alterations in nuclear positioning and interactions with other parts of the genome. This revealed a minor but significant increase in the pairing frequency of 10q and 4q subtelomeres in FSHD patients when compared to control individuals, indicating that this pairing may promote exchanges between the 4q35 and 10q26 subtelomeric repeat arrays. In this way interchromosomal rearrangements might participate in FSHD pathology.

Considering these two phenomena, translocations between 4q35 and 10q26 and somatic pairing of 4qter and 10qter, our laboratory decided to investigate the percentage of exchanges in a large healthy population to gain more insight in the normal occurrence and overall repeat size distributions. These results are described in *Chapters 2 and 3*.

1.4 Mosaicism and its consequences for human disease

The term mosaic was used in ancient times to refer to floors or windows with numerous little colourful pieces of stone or glass representing mostly religious or historical events. In a biological context, mosaicism may refer to two distinct phenomena. Mosaicism can refer to a viral disease of plants, resulting in light and dark areas in the leaves, which then often become shrivelled. An alternative definition, and one more appropriate for molecular biology, is a condition in which an individual has two or more genetically distinct cell lines derived from a single zygote, but differing by virtue of the presence of a mutation or non-disjunction [131]. Mosaicism arises when an error disrupts the correct number of chromosomes segregating to each cell during cell division or creates a mutation in a single gene or locus [313]. The first paper on the existence of chromosomal mosaicism was published in 1959. Here, the phenotype of a case of Klinefelter syndrome was described, in a man with two distinct cell populations in his cultured bone marrow. One cell population contained the normal karyotype (46,XX), whereas the other cell population contained the karyotype (47,XXY) [117]. Mosaicism is a very common phenomenon in multicellular organisms. In humans, for example, female cells display random inactivation of the maternally or paternally derived X-chromosome, and so they exhibit mosaicism for the expression of specific genes on the X-chromosome. Moreover, as the number of cells in the adult human body (10^{14}) exceeds the mutation frequency of most genes, one must conclude that every individual must exhibit some degree of genetic mosaicism. Thus, everyone probably carries mutations in any given gene in some proportion of his or her cells [153].

The range of diseases that have been associated with mosaicism is very broad and includes almost all processes in the human body. Observations of mosaicism in all its forms could thus have important clinical implications. All events that induce mosaicism can influence the phenotype in a negative or positive manner. The effect of a mosaic population is usually masked, but can also result in major phenotypic expression and disease. Mosaicism has already been described for more than 30 different diseases, including both Mendelian and non-Mendelian disorders, like hemophilia, Duchenne muscular dystrophy, Friedreich ataxia and FSHD [144]. The next paragraphs will give a general overview on mosaicism, followed by a description on the occurrence of mosaicism for the 4q35 region and its consequences for FSHD.

1.4.1 Inheritance of mosaicism

Mosaic cell populations can be the result of mutations in nuclear or mitochondrial DNA in post-zygotic cells, epigenetic alterations of DNA, numerical or structural chromosomal abnormalities (such as missing or extra chromosomes) or of spontaneous reversion of inherited

mutations. This may occur in any given cell and can be passed on to all daughter cells at any particular point in development or lifetime of an individual. From here on, any mutation described will be referred to as a mosaicism-inducing event.

One of the most important issues in genetic counselling is whether a mutation is likely to be transmitted to the next generation. This depends on the nature of the mutation and more importantly, when and where it occurs. When the mutation occurs in one of the first few cell divisions after fertilisation, it yields a larger proportion of mutant cells in the total cell population than when the mutation occurs in a later developmental stage [153, 313, 442, 442].

A post-zygotic mutation can be passed on to the offspring only if it is present in cells that give rise to germ cells. Therefore, a distinction is made between germline mosaicism (also termed gonadal mosaicism) and somatic mosaicism. In the latter form, the absence of the mutation in germ cells will prevent transmission to progeny. When mosaicism arises during an early stage in embryogenesis, the same mutation can coexist in both somatic and germ cells within the same individual. This is dependent on the specific cell (type) affected and the developmental timing of the mosaicism-inducing event. A mutation will be passed on through the germline, but will be absent from somatic cells only if the mutation occurred in cells already committed to germline formation [153, 442]. The next paragraph will discuss these two types of mosaicism in more detail.

1.4.2 Germline and somatic mosaicism

When two or more children manifest a phenotype of an autosomal dominant or X-linked disorder in a family with no history of disease, germline mosaicism in a substantial number of germ cells in one of the parents could be the underlying cause and has been demonstrated in disorders like osteogenesis imperfecta, hemophilia A and Duchenne muscular dystrophy [24, 72, 100, 153, 442]. In this type of mosaicism, which is depicted in *Figure 1.5.A*, the mutation can occur (1) only in the germ cell precursor which will undergo meiosis to yield a germ cell that will continue to divide, or (2) very early in embryonic development, before the cell fate determination to germinal or somatic cells, and will thus be present in both cell types. The latter is also called gonosomal mosaicism. Factors such as the gene or locus involved, tissue specificity and the degree of mosaicism will eventually determine if a carrier may be asymptomatic or present various characteristics of a disease [448]. The exact risk of passing on a disease-associated mutation is variable in this case, but can be as high as 50% [448].

A classification in two different groups of germline mosaicism-associated disorders has been proposed by Zlotogora [448]. The first group contains parents in whom mosaicism is suspected, because there is transmission of the mutation to two or more children, while they

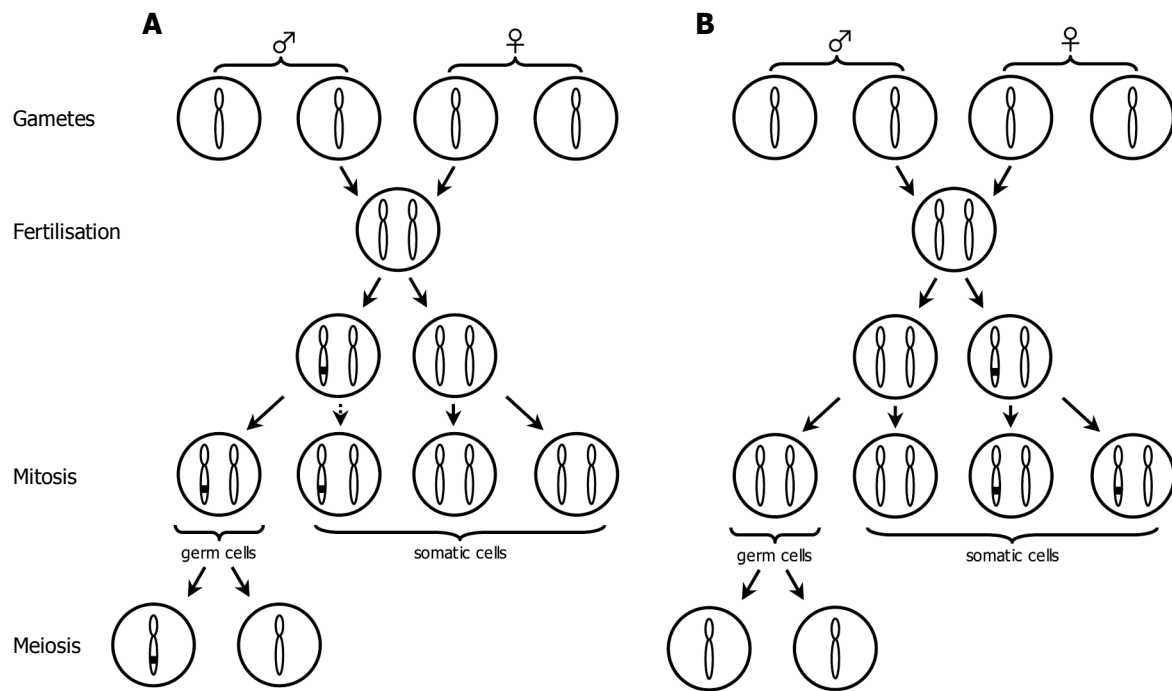


Figure 1.5

(A) Schematic representation of development and inheritance risk for germline mosaicism only, or both somatic and germline mosaicism. The black bar in the chromosome represents the disease-inducing mutation that, depending on the developmental stage in which this mutation occurs, spreads throughout germ cells alone or both somatic and germ cells, which is indicated by a dotted arrow. Depending on the timing, the mutation can already be identified in clinically unaffected parents of non-mosaic patients. The inheritance risk for future offspring depends on the proportion of germ cells that contains the mutation. The maximum risk will be 50% [448].

(B) Schematic representation of development and inheritance risk for somatic mosaicism. The black bar in the chromosome represents the disease-causing mutation that will be visible in somatic cells due to clonal expansion of cells carrying this mutation. In this scenario the offspring is not at risk, since the mutation is not present in the germline.

themselves have very mild symptoms or no disease characteristics at all. In the second group, parents present clinical symptoms of a disease caused by the expression of somatic mosaicism and have at least one child who has received the mutation and is clinically affected, indicative of gonosomal mosaicism. The mutation present in both somatic and germ cells confirms that the mutation most likely occurred very early on in embryonic development.

Whether a mutation will affect only somatic or germ cells or both cell lineages will thus depend on cell fate. During human embryogenesis, at least four cells are allocated to the germline before the emergence of somatic cell lineages [442, 448]. All evidence points out that this must happen at an early developmental stage, but the exact timing is not yet known [448]. Mutations that appear at later time points in the development of somatic cell lineages will only give rise to somatic mosaicism [442, 448] (*Figure 1.5.B*).

The phenotypic consequences of a somatic mutation depend on various factors such as the cell type(s), tissues and/or organs in which the mutation is expressed. A mutation may be tolerated

in some cells, whilst being lethal in others or instead having a selective advantage as observed in malignancies. When a gene is necessary for normal cell function in a certain tissue, non-mutated cells may then select against cells with a mutated gene and hence induce mosaicism, whilst the same mutation is tolerated in other tissues [153]. Therefore, phenotypic variability in expression levels can, besides alterations in DNA sequences in specific genes or mutations in modifying genes, also be influenced by mosaicism. The timing of somatic mutations will then be important, as an early mutation is likely to have a greater effect on the expressed phenotype [144]. As an example, the X-linked dominant disease Rett syndrome is usually lethal in hemizygous males, but some survive due to somatic mosaicism [382]. In these men, normal cells may arise that outgrow the mutant cell population. The healthy cells will spread throughout the developing individual and compensate for the mutant cells, allowing survival of the individual who would otherwise have died during development. Although it is not yet understood why these mutations occur before fertilisation or as a post-zygotic event, the presence of an X-linked disease in a male is rare and is therefore an important diagnostic indicator of the occurrence of somatic mosaicism in that individual [297]. Also numerical or structural abnormalities of chromosomes are resulting from somatic mosaicism, like observed in the Pallister-Kilian syndrome caused by a tetrasomy of the short arm of chromosome 12 [442].

1.4.3 Mosaicism is a common feature in cancer

Cancer pathology is probably the best-known example of mosaicism [29, 101, 442]. Nearly all cancers are monoclonal in origin, and a critical mutation event has (or multiple mutations have) transformed this single cell into a cancer cell with uncontrolled proliferation leading to the formation of a neoplasm. When a mosaic cell population arises due to a mutation in a tumour suppressor gene, this mutation represents the ‘first hit’ towards the development of cancer in these cells. Cells carrying such a mutation are predisposed to malignant formation once a second somatic mutation has inactivated the remaining normal copy of the gene [29, 101, 442]. In the hereditary cancer retinoblastoma, for example, somatic and germline mosaicism for the initial mutation in the retinoblastoma gene is well documented. Not all mosaic individuals have or will develop retinoblastoma [360]. Also material of patients diagnosed with neurofibromatosis type 2, a severe autosomal dominant disorder that predisposes to multiple benign tumours of the nervous system, displays mosaicism. Since the detection rate of mutations in blood samples of neurofibromatosis type 2 founders (i.e. patients with clinically unaffected parents) was lower compared to that in non-founders, the failure to find mutations in blood leucocytes here turned out not to be because of a limitation in the applied screening technique, but was best explained by the high percentage of mosaicism that was observed in tumour specimens in as much as 25% of founders [194]. Germline and somatic mosaicism

associated with human cancer is also observed in, for example, neurofibromatosis type 1, testicular tumour development and von Hippel-Lindau disease [33, 224, 279] and is now regarded as a major contributor to the development of the pathogenesis of cancer.

1.4.4 Detection of mosaicism in FSHD

In 1993, Weiffenbach *et al.* [413] mentioned two families with germline mosaicism. Examination of the parents and offspring in these two families revealed clinically confirmed FSHD in the children, but no signs of muscle weakness in the face or shoulder girdle of both couples of parents. In both families, DNA analysis on peripheral blood lymphocytes confirmed a contracted fragment in all patients, but no contracted bands present in either parent. This suggested that only some or all germ cells of the father or mother contained the mutation and no somatic cells (at least no lymphocytes) were affected. This mutation probably occurred in a germ cell that continued to divide. In the same year, another publication by Griggs *et al.* [148], describing new mutations in sporadic cases, mentioned the same observation in one of their seven families studied.

The first complete survey dedicated to studying germline mosaicism in FSHD was published in 1995 [393]. This article reported three sporadic FSHD families in whom the affected proband had inherited a potential rearranged D4Z4 fragment from one of the parents and was severely affected in all three cases. Analysis of DNA isolated from peripheral blood lymphocytes showed that in all three families the contracted D4Z4 fragment in the child was also visible in the mother, but the fragment was less intense after hybridisation with probe p13E-11. This indicated that these three mothers transmitted the disease allele to their children. Since only a faint *Eco*RI fragment was seen in the lymphocytes of these mothers, also a small percentage of their somatic cells must be mosaic for the contracted D4Z4 repeat. As the mutation is present in both gametes and somatic cells (here only lymphocytes were analysed), the mutation must have occurred before the differentiation of cells into endodermal, ectodermal and mesodermal cell lineages.

Parents of apparently isolated FSHD cases may have insufficient expression of the mutation in different tissues due to mosaicism and therefore show no or only a very mild FSHD phenotype. This may explain the early observation made by Brooke [48] of the unusual manifestations of severe infantile FSHD with a minimally affected parent. With these new insights, germline mosaicism could explain the occurrence of several affected FSHD offspring of unaffected parents and close the debate regarding a possible autosomal recessive inheritance of FSHD [393]. In addition, the occurrence of mosaicism for the FSHD-associated region may, at least partially, explain the anticipation noted in some FSHD families [142, 253, 377, 381, 445].

Subsequent to the first studies, other groups started to report mosaicism as well (*Table 1.1*) [14, 124, 143, 148, 167, 198, 199, 232, 249, 257, 305, 342, 380, 393, 413, 444, 445]. A total of 20 patients have been reported to be mosaic for a chromosome 4 allele associated with FSHD in a total of 983 families tested. However, one must take into account, that most research laboratories tested lymphocyte DNA from the individuals with conventional gel electrophoresis, while mosaic alleles are more easily visualised with PFGE [232]. Seven patients transmitted the mosaic allele to their offspring and are therefore also mosaic in their germ cells. From the other patients gonadal tissue was not analysed and transmission of the mosaic allele has not yet been observed, because these individuals did not have children at the moment of lymphocyte testing. Mosaicism in one of the unaffected parents (or parents that are not aware of having FSHD) was reported in 89 parents of a non-mosaic patient in the 983 families tested. Nine individuals were only mosaic in their germ cells, because no contracted fragment was observed in their lymphocytes. In 80 individuals mosaicism was observed in both somatic and germ cells. Somatic mosaicism was always analysed in lymphocyte DNA. Germline mosaicism in unaffected parents, however, was always suggested by multiple affected offspring. However, in none of the cases gonadal tissue was actually tested to confirm the presence of the mosaic allele.

In general, all studies on mosaicism for the contracted D4Z4 repeat in FSHD indicated a wide range of phenotypic expression of two genetically distinct cell populations for a single chromosome 4 with one of the cell populations carrying a contracted D4Z4 repeat array. Clinical effects of mosaicism can be absent or very mild in adults, and parents could therefore be at risk to pass on the condition to their children if their germ cells contain the mutated allele. Interestingly, in all studies the majority of mosaic parents are female putting asymptomatic mothers at a higher risk of being a carrier of a mosaic FSHD allele.

We decided to analyse multiple Dutch *de novo* FSHD families to determine the occurrence of mosaicism in patients and parents and the effect of “being a mosaic” on development of the disease (*Chapter 4*). Whereas we would expect to observe a mosaic mixture of one contracted allele and an unchanged ancestral allele, in three families we detected two rearranged D4Z4 repeat arrays instead of one. Therefore, we analysed these families in more detail and focussed on the mechanism from which these rearrangements originate (*Chapter 3*).

Table 1.1 Complete summary of published data on mosaicism for the FSHD-associated region.

Reference	Family (n)	Parents (n)			FSHD patients (n)		
		Germline mosaicism		Somatic mosaicism		Germline mosaicism	
		male	female	total	male	female	total
Weiffenbach <i>et al.</i> , 1993	2	nr	nr	2	-	-	0
Griggs <i>et al.</i> , 1993	1	nr	nr	1	-	-	0
Upadhyaya <i>et al.</i> , 1995	3	-	3	3	-	3	3
Padberg <i>et al.</i> , 1995	5 ^a	nr	nr	5	-	-	0
Zatz <i>et al.</i> , 1995	1	-	1	1	-	1	1
Köhler <i>et al.</i> , 1996	11	2	9	11	2	9	11
Bakker <i>et al.</i> , 1996	3	3	-	3	3	-	3
Hsu <i>et al.</i> , 1997	1	nr	nr	1	-	-	0
Lunt, 1998	26 ^b	5	21	26	5	21	26
Roques <i>et al.</i> , 1998	1	1	-	1	1	-	1
Zatz <i>et al.</i> , 1998	4 ^c	1	2	3	1	2	3
Galluzzi <i>et al.</i> , 1999	7 ^d	nr	1	5	nr	1	5
Köhler <i>et al.</i> , 1999	3	1	2	3	1	2	3
Van der Maarel <i>et al.</i> , 2000	14	2	3	5	2	3	5
Tonini <i>et al.</i> , 2004	7 ^e	nr	1	7	1	6	7
Goto <i>et al.</i> , 2004	18 ^f	6	6	12	6	6	12
Lemmers <i>et al.</i> , 2004	11 ^g	-	-	-	-	-	-
Total	118	≥ 21	≥ 49	89	≥ 22	≥ 54	80
		3	4	7			20

The triangle (Δ) indicates the number of reported mosaic FSHD families, while the asterisk (*) indicates the total number of families analysed in each study. Both inherited and *de novo* mutations are included. "Parents" indicates mosaic asymptomatic parents of a non-mosaic FSHD patient. nr: gender not reported. a: Here only DNA from two families was analysed, where no short fragment was observed in lymphocytes. b: Data collected on the 44th ENMC international workshop; this contained a summary of other articles and included results which had been previously presented. Two groups summarised in this report (15/128 and 87/128) may also still include mosaic individuals, only the 26/128 are established mosaic cases. c: One female from this study was already reported in 1995 [445] and was excluded here. d: The gender of one mosaic parent was reported. e: Allele transmission by only one mosaic parent to her children was reported. f: The gender of one patient with somatic mosaicism was not reported. g: This article included 8 families that have been previously published by Van der Maarel *et al.* [257] and were excluded here.

1.5 Epigenetic modifications

Chromatin is a dynamic structure that determines gene activity of transcriptional DNA sequences. Transcription usually occurs when the chromatin structure is open, while tightly packaged chromatin is associated with transcriptional repression. The different epigenetic processes that are each by themselves able to change chromatin conformation, e.g. DNA methylation, histone modifications and ATP-dependent remodelling complexes [11, 169], can also interact with each other to achieve gene expression or stable molecular silencing. Numerous examples in literature have made clear that compaction of genomic DNA into chromatin is fundamental for the regulation of gene expression. Each protein involved in modification of chromatin might, when mutated, play a role in human disease and some proteins have already been associated with diseases. For FSHD, epigenetic modifications of the subtelomere 4q35 have not been studied extensively, although there are several models that propose an alteration of chromatin structure in FSHD patients (see *Paragraph 1.6*). We have investigated chromatin structure present in the 4q35 region by focussing on DNA methylation and histone acetylation. After a description on chromosome architecture, the following paragraphs will therefore primarily focus on the role of DNA methylation and the function of histone modifications in the human genome.

1.5.1 Chromatin architecture

1.5.1.1 Heterochromatin and euchromatin

In the nucleus DNA is assembled with proteins into a nucleoprotein complex, designated chromatin [67, 321]. Most eukaryotic genomes can roughly be divided into two types of chromatin: heterochromatin and euchromatin. Heterochromatin mainly consists of transcriptionally inactive regions in a tightly packed chromatin structure replicating late in S-phase and is usually highly DNA methylated and poorly histone acetylated [89, 282, 338]. Heterochromatin occurs in two varieties: constitutive and facultative heterochromatin. The first is rich in non-coding repetitive sequences and contains few transcribed genes. This type of heterochromatin is often observed at centromeric and (sub)telomeric regions of chromosomes. It tends to be permanent transcriptionally inactive and can induce repression of nearby genes by a position effect [80, 248, 315, 321]. Facultative heterochromatin has a compact conformation and can be specific for one or two homologous chromosomes or for certain cell types or developmental stages. This chromatin form describes the inert state of sequences that contain many genes that also exist in active copies [315, 321] and is only silenced in specific contexts, for example when it contains tissue-specific genes that are silenced in all but the proper tissue [324].

Euchromatin has an open chromatin structure characterised by DNA hypomethylation and histone acetylation and is, due to its accessibility, generally associated with actively transcribed regions [80, 282, 321, 338]. It has a high density of functional genes, mainly comprises single copy sequences and usually replicates early in the cell cycle [89, 405].

1.5.1.2 Nucleosome conformation

The basic unit of chromatin is the nucleosome, which consists out of 146-180 bp of DNA wrapped around a protein core in approximately two left-handed superhelical turns [184]. This protein core of eight histone proteins (also termed octamer) is made up of two copies each of histone 2A (H2A), histone 2B (H2B), histone 3 (H3) and histone 4 (H4) [67]. Nucleosomes are connected by short segments of linker DNA and compacted by linker histones 1 (H1) and other non-histone proteins into chromatin fibres [111, 204]. The position and density of the histone octamers are mediated by ATP-dependent chromatin remodelling complexes that mobilise nucleosomes via alterations of the contact between DNA and histones [388]. Several processes mediate the affinity of the histone octamers for DNA and ATP-dependent chromatin remodelling complexes: acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ADP-ribosylation of histone residues [28, 46, 92]. The position of the histones and these modifications will organise chromatin together with the other epigenetic changes in an open or condensed state and in this manner regulate the accessibility of DNA for the transcription machinery [282].

Four groups of proteins have the ability to modify and remodel chromatin and regulate gene expression: DNA cytosine methyltransferases, methyl-CpG binding proteins, histone modification enzymes and ATP-dependent remodelling complexes. Often these proteins interact with each other to form a higher-order chromatin structure and to regulate the accessibility of DNA [238]. One of the proposed models for this interaction is schematically depicted in *Figure 1.6*.

1.5.1.3 Epigenetics

Eukaryotic cells have the capacity to epigenetically modify their genomes. A current definition of epigenetics is “mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” [345]. Epigenetics can be seen as a system that selectively uses genome information by activating or inactivating functional genes [282] and alters the phenotype without changing the genotype [67]. Three major biochemical epigenetic marks in the formation of chromatin are DNA methylation, histone acetylation and histone methylation [238].

1.5.2 DNA modification by methylation

DNA methylation is the most prevalent endogenous DNA modification in eukaryotes [75, 359] and involves the addition of a methyl group to the carbon 5 position of the cytosine ring within the context of the sequence 5'-CG-3', also called a CpG dinucleotide [35, 359]. Although non-CpG sequences may also be methylated, nearly all methylation of cytosines in eukaryotes is observed in CpG dinucleotides [11, 75, 333].

1.5.2.1 DNA methyltransferases

Cytosine residues in newly synthesized DNA are methylated via a DNA-specific enzymatic reaction by a DNA methyltransferase (DNMT) that catalyses the addition of a methyl group from the methyl donor S-adenosyl-L-methionine to the carbon 5 position of cytosine [336, 359]. Eukaryotic DNMT establishes and maintains the methylation status of transcriptionally

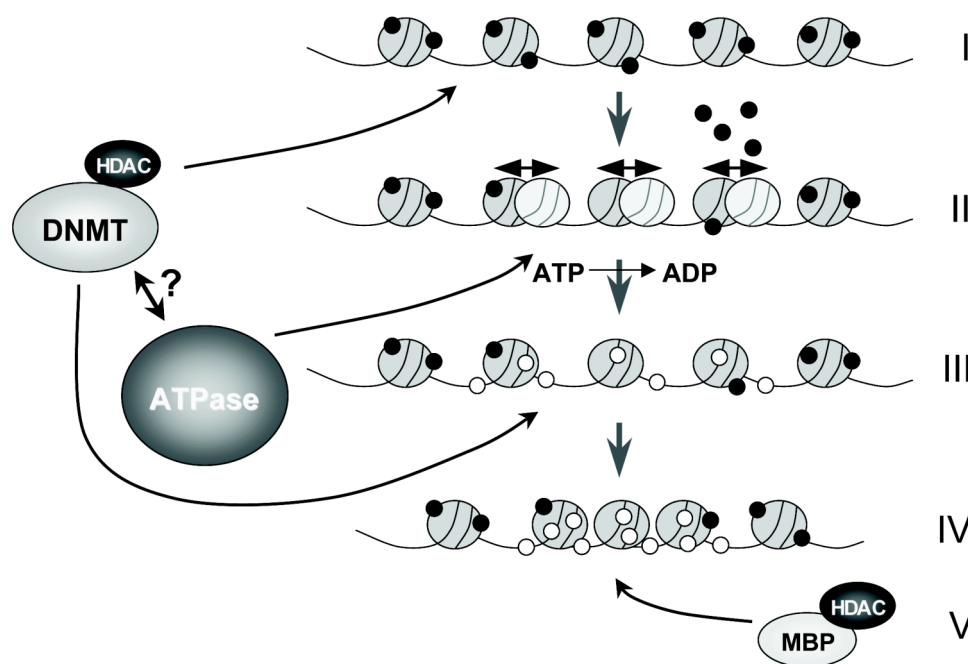


Figure 1.6

Schematic model proposed by Robertson [336] depicting how HDACs, ATP-dependent chromatin remodelling enzymes and DNMTs may cooperate to generate region-specific methylation patterns on DNA wrapped around nucleosomes. The acetylated histones (acetyl groups indicated with black dots) located within a (potential) transcriptionally active region become deacetylated by HDACs. It is not yet known how HDACs are targeted towards the histones and this may require other chromatin remodelling activities. Deacetylation likely starts the transcriptional shutdown of the region (I). ATP-dependent chromatin remodelling enzymes will then recognise the chromatin and alter the chromatin structure or nucleosome positioning (II) so that a DNMT can access its DNA target sites directly or via creating a chromatin signature recognised by a DNMT or a DNMT-associated complex (III). DNMT will then methylate the cytosines in the DNA region (III; depicted with open dots). Evidence exists that DNMT and HDAC are components of a single complex, but it still needs to be determined if the ATP-dependent chromatin remodelling enzyme is also associated with this complex or acts independently. The methylated region then locks the chromatin in a silent state (IV) and the methylated cytosines of the DNA will recruit methyl-CpG binding proteins (MBP) and associated co-repressors to further strengthen the established transcriptional repression and chromatin compaction (V).

inactive sequences such as repetitive elements and imprinted alleles [291]. Until now five human DNMTs have been identified: DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L (reviewed in [336]). DNMT1 has a preference for hemimethylated DNA and is believed to be responsible for the propagation of methylation patterns after each round of replication in somatic cells. The role of DNMT2 is not yet clear, but it is thought to play a role in centromere functioning. DNMT3A and DNMT3B are involved in *de novo* methylation of specific sequences such as pericentromeric satellite sequences, probably during specific stages of gametogenesis and early development. DNMT3B is also required for maintaining the methylation patterns of pericentromeric regions through cell division, which indicates that a separate function in maintenance and *de novo* methylation for each DNMT cannot (always) be defined. Finally, DNMT3L is expressed during gametogenesis and is essential for the establishment of maternal genomic imprinting [31, 258, 374].

1.5.2.2 CpG islands

Approximately 70-80% of all CpG residues in the mammalian genome are methylated [11, 37, 333], except for dense CpG clusters, which are observed in genomic regions of 200 bp to several kb [35, 75]. These regions, termed CpG islands, have a very high CpG frequency (60-70%) [76] and are often located in or near promoter regions or coding regions of nearly all housekeeping genes and of 40% of genes with an expression pattern restricted to specific tissue types or developmental stage [7, 11, 38, 76, 223, 338, 359, 398]. All CpG islands together make up approximately 2% of the genome [76]. The chromatin conformation of these CpG islands is open, which makes the DNA sequence accessible for transcription factors [39, 359]. CpG islands not associated with promoters or coding sequences are often methylated, just like CpGs located outside CpG islands, which are frequently detected in satellite sequences and transposable elements [11, 396].

1.5.2.3 DNA methylation and development

DNA methylation patterns are established during the development of an organism and serve to suppress genes dispensable for the function of the mature cell [333]. After fertilisation, but before implantation, genomic methylation is almost completely converted into unmethylated cytosines [182, 274]. However, imprinted genes do not lose their methylation pattern [236]. This wave of demethylation is then followed by *de novo* methylation upon embryo implantation to re-establish the methylation pattern inherited from the gametes [182, 274, 327]. This *de novo* methylation does not apply to the active copy of imprinted genes, most of the tissue-specific genes that are unmethylated in the tissue of expression, and CpG islands associated with active genes, which remain unmethylated in germline and most somatic tissues [47, 181, 325, 358,

441]. In this manner, a bimodal pattern of methylation is created [358], which is inherited and remains stable through multiple cell divisions [38, 326]. These patterns can only alter during developmental or pathological changes of the cell or as a response to environmental modifications [149].

1.5.2.4 Influence of DNA methylation on silencing

DNA methylation is an effective suppressor of gene expression in many processes and can interfere with transcription in various ways [38, 180, 359]. The repressive effect of DNA methylation depends on the nature of control elements present (like an enhancer or promoter), the density of methylated CpGs, the protein environment in a specific type of cell and the chromosomal context [45, 165, 277, 352]. DNA methylation has a stabilising effect and can cause transcriptional repression in several ways. The first involves direct interference of the methylated cytosines in binding of a protein to its corresponding CpG-containing DNA recognition sequence. In this case methylation alone hinders or eliminates binding of the basal transcription machinery or ubiquitous transcription factors (like E2F, NF- κ B, AP2 and CREB), which require contact with an unmethylated cytosine [38, 185, 282]. Methylation within gene regulatory elements, like promoters, enhancers and insulators, will consequently repress the function of these elements [75]. Secondly, repression may involve methyl-CpG binding proteins that bind specifically to methylated CpGs. By limiting access to regulatory elements they can indirectly inhibit the binding of transcription factors, such as CTCF, Sp1 and YY1, that can even bind to CpG-containing sequences when these are methylated [38, 45, 75, 277, 282]. A third mechanism by which methylation may mediate repression of transcription is by altering the chromatin structure through an interaction between DNA, histones and proteins that together form chromatin [277, 359]. Finally, methylation induces and stabilises the inactive state of chromatin by affecting the positioning of nucleosomes, influencing the DNA sensitivity to DNaseI and targeting sequences for assembly into the condensed state [96, 186, 264]. This mechanism probably also is affected by the action of (or interactions with) methyl-CpG binding proteins [277].

1.5.2.5 Suppression and defence

DNA methylation may also be a mechanism to suppress “parasitic” DNA sequences. In mammalian cells transposable sequences, which are mobile genetic elements that are scattered within and around most genes, represent at least 35% of the genome [441]. These transposons, of which the *Alu* repeat family is the most abundant, are very often methylated and transcriptionally silenced in the genome [38]. Two ideas have been put forward that set this repression in an evolutionary context. The first hypothesised that methylation acts as a stabiliser

for a genome that contains large amounts of repetitive DNA by masking or inhibiting homologous recombination between these repetitive sequences. This prevents repetitive DNA from spreading throughout the genome and thus avoids DNA damage [406, 441]. The other proposed that the irrelevant transcription of all these promoter sequences would generate a high level of transcriptional noise that would interfere with all necessary gene expression. Therefore, repression would be necessary to reduce this transcriptional noise, thereby allowing vertebrates to gather and selectively use genes that are essential for development [36].

Besides the suppression of transposons, cytosine methylation may also be part of a genome defence system in mammals that evolved primarily to oppose to the threats posed by parasitic sequence elements [32, 263]. *De novo* methylation especially affects promoters of viral genes which would thus become inactivated, like those of the herpes viruses [18, 210, 337].

1.5.3 DNA modifications via histones

The highly conserved core histone proteins H2A, H2B, H3 and H4 located in the nucleosome and the linker histone H1 are the major protein components of chromatin in the eukaryotic nucleus and function as blocks to package eukaryotic DNA into higher-order chromatin fibres [65, 184, 203]. Chromatin structure can be disrupted by nucleosome remodelling, the replacement of histones that form the nucleosome by histone variants or by covalent histone modification (reviewed in [106]).

These posttranslational chemical modifications are located on histone residues, such as lysine or arginine, in the flexible amino-terminal domains, which protrude out of the nucleosome surface [106], or on globular carboxy-terminal domains comprising the nucleosome scaffold [4]. These domains are commonly referred to as “tails” [184]. The acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ADP-ribosylation of histones correlates with specific transcriptional states [28, 152]. However, histone residues will only be modified if histone-modifying enzymes are targeted to the histones by DNA-bound activators or repressors, since they are unable to access to the histone residues themselves [204]. Nowadays, multiple residues per histone have been identified as possible modification sites and, taking the variety of modifications into account, this will allow many combinatorial possibilities that can play a role in gene expression (e.g. modulation of transcription during development and X-chromosome inactivation in females), DNA replication, DNA repair, genome stability, chromosome dynamics and apoptosis [140, 211, 369, 384].

1.5.3.1 Histone acetylation

Histone acetylation is the most intensively studied modification, which is reversible and plays a prominent role in higher-order chromatin packing [27, 28, 80, 145, 190]. Furthermore, histone

acetylation is involved in replication, nucleosome assembly and interactions of nucleosomes with non-histone proteins [145]. Histone acetyltransferases (HATs) and histone deacetylases (HDACs) modify histones by the addition or removal of an acetyl group [28, 146, 189]. Sites of acetylation include conserved lysines in the amino-termini of H4 and H3 and less conserved lysine residues in H2A and H2B [212]. Histone acetylation may influence transcription by altering the folding properties of chromatin by changing the accessibility of DNA [66, 162] or provide specific binding places for repressors and activators of gene expression through this specific modification [139, 212]. Because this modification is involved in many cellular processes, different functions may require the acetylation of specific histones [178].

Like DNA methylation, the distribution of acetylated histones is not randomly distributed over the genome. Very low levels of histone acetylation are observed in heterochromatic regions, which are associated with high levels of DNA methylation, while high levels of histone acetylation are seen along coding regions in euchromatin [178]. This concurs with the view that transcriptional repression by DNA methylation is linked to histone acetylation through the recruitment of HDACs by methyl-CpG binding proteins [75, 178, 284, 290, 333].

1.5.3.2 Histone methylation

Primarily the amino-terminal histone tails of H3 and H4 are modified by histone methylation [369]. The modification occurs both at lysine and arginine residues [28], is established by specific histone methyltransferases [140, 204] and is biochemically stable [401]. The effects of lysine and arginine methylation depend on the specific residue and the number of methyl groups in a modification (mono-, di- or trimethylated for lysine and mono- or dimethylated for arginine) [170, 190, 204, 369]. Methylation of lysines (K) in H3 and H4 is involved in gene regulation [27]: H3-K9 dimethylation and H3-K27 trimethylation are associated with gene silencing, while methylation of H3-K4, H3-K36 and H3-K79 are associated with gene activation [204]. Arginine methylation is correlated with the active state of transcription, is involved in signal transduction pathways and recruits methylases as co-activators to promoters [27, 204].

1.5.3.3 Histone phosphorylation

Histone phosphorylation involves serines and threonines, particularly of H3 and H2B [28] and the linker H1 [81]. H1 is phosphorylated at serines and threonines in both amino-terminal domains as carboxy-terminal domains, while phosphorylation of the nucleosome histones is only located in the amino-terminal tails [81]. This modification, which is established by a balance between kinases and phosphatases, is highly reversible [4, 81]. It is an important regulatory mechanism for transcriptional activation, especially of immediate-early response

genes, like *c-fos* and *c-jun* [81]. Furthermore, phosphorylation is often associated with chromosomal condensation that includes mitosis and meiosis, apoptosis and DNA damage [4, 27, 152]. This indicates that one modification is involved both in opening and closing of chromatin, but it is not yet clear how phosphorylation generates the condensed chromatin structure [28].

1.5.3.4 Histone ubiquitination

Ubiquitination (by some authors also called ubiquitylation) is a dynamic process involving the coupling of ubiquitin to lysines in the carboxy-terminal domains of histones, which will generally become mono-ubiquitinated [43, 80, 106, 293]. This coupling has been observed in H2A and H2B. Also ubiquitination of H3 was reported *in vivo* in elongated spermatids of rat testis and H1 ubiquitination is observed in *Drosophila*, but the sites for this modification at H1 and H3 are not yet elucidated [62, 316]. Ubiquitination is associated with many cellular processes, like protein degradation, DNA repair, cell cycle control, stress and heat shock response, endocytosis signalling, ribosome biosynthesis, neural and muscular degradation, and transcription [43, 275]. Levels of ubiquitination are dependent on the activity of a family of ubiquitin-conjugating enzymes, of which Rad6 is a member [175, 183]. Ubiquitination of histones can affect spermatogenesis, chromosomal segregation and condensation, and transcription [43, 175, 434]. This modification could label certain genomic regions to interact with specific proteins or to be packaged in a special manner in metaphase chromosomes [175, 275]. Ubiquitination on H2B is critical for cell cycle progression in mitosis and meiosis [27] and plays a role in transcriptional activation by defining regions of active chromatin [183, 293]. Ubiquitination of H1 also plays a role in transcriptional activation [27].

1.5.3.5 Histone sumoylation

Small ubiquitin-related modifier (SUMO) family proteins, which show resemblance to ubiquitin, are involved in dynamic histone modification by the addition of a SUMO to lysines in the carboxy-terminal domain [278, 287]. These proteins are activated in an ATP-dependent manner and transferred to UBC9, until now the only SUMO-conjugating enzyme known to function *in vivo* [287], after which they are attached to a specific lysine in the target protein [399]. The best-characterised member is the SUMO protein (also termed smt3, sentrin, GMP1 or PIC1; reviewed in [278]) that has been shown to stimulate activity of for example p53 and heat shock factors HSF1 and HSF2, but most frequently correlates with repression of target genes, such as *c-jun* and *Sp3* [278, 399]. Recently, it was suggested that histone sumoylation mediates transcriptional repression and gene silencing through the recruitment of HDACs and heterochromatin protein1, but no clear molecular basis is known yet [357, 399].

1.5.3.6 Histone ADP-ribosylation

Two non-enzymatic protein modifications, glycation and glycooxidation, result in the formation of protein glycooxidation products, termed AGE [60, 375]. These two processes are potentially toxic, which suggests that there should be mechanisms to repair or minimize these modifications [60]. *In vitro* ADP-ribose is an efficient source for glycation and glycooxidation [60], ADP-ribose formed conjugates of histone H1 in cell cultures after DNA damage [207] and Wong and colleagues reported a stable complex that contained a dimer of histone H1 and ADP-ribose polymers [431]. These observations suggest that the covalent linkage of histone molecules is the result of histone glycooxidation initiated by ADP-ribose [60]. Histones organised in a chromatin structure studied in sugar-incubated nuclei were, in contrast to purified histones in solution, less reactive in the glycation reaction, as a consequence of the accessibility of the histones in the chromatin structure. However, H1, H2B and H2A were more reactive than H3 and H4, which are most protected from this reaction. When tested *in vivo*, high concentrations of ADP-ribose were able to induce apoptotic cell death indicating the toxic character of these modifications [375]. By contrast, in sea urchins ADP-ribosylation plays a vital role after fertilisation by protecting the integrity of maternal histone proteins against nuclear proteinases, while sperm-specific histones are degraded and replaced by maternal histone variants. This suggests that ADP-ribosylation can also act as a mechanism to protect nuclear proteins from proteolysis in a specific manner [276]. However, this has not yet been demonstrated in other organisms.

1.5.3.7 Histone interactions

All histone modifications can affect one another and many correlated in a positive or negative manner [106]. In the process of transcriptional activation histone phosphorylation is applied together with histone acetylation, in which phosphorylation occurs before acetylation and even promotes acetylation, resulting in the recruitment of the transcription machinery. Also histone acetylation and histone methylation are thought to influence each other both in a positive and negative manner (reviewed in [27]). Furthermore, histone modifications on H3 and H4 are interrelated to and interdependent on each other, suggesting specific patterns for activation and repression [28]. Besides the interaction of histone modifications, also the interplay of different histone modification enzymes with DNMTs and chromatin remodelling factors can lead to transcriptional activation and repression [106, 152].

1.5.3.8 Histone variants

Remodelling of chromatin can also be established by the exchange of the histones within the nucleosome by variants [152, 349]. These histone variants (also termed replacement histones; [246]) can substitute “normal” histones and this may cause transcriptional changes [106, 162]. The replacement of H2A by H2AZ is associated with increased nucleosome stability, while variant H2AX seems to be involved in recombination events. Furthermore, inactive X-chromosomes in mammals are enriched for variant macroH2A at the onset of differentiation [349]. H3.3 is a variant of H3 and is linked to transcriptionally active genes [164, 371]. Another homologue of H3 is CENP-A observed at centromeric regions of vertebrate chromosomes, which is essential for kinetochore formation and plays an early role in organising centromeric chromatin at interphase [164, 371]. Every variant is involved in local chromatin structures as they can be chemically modified as well. This is illustrated by the carboxy-terminal domain of H2AX that is rapidly phosphorylated in response to double strand breaks and plays an important role by associating with factors essential for DNA-damage repair [170].

1.5.3.9 Function of histone modifications

Currently, there are two models for the function of histone modifications. The first model suggests that these modifications directly affect chromatin or nucleosome structure by preventing the binding of specific transcription factors, changing the mutual interaction of nucleosomes in array or by altering the interaction of histones and DNA in a nucleosome [27, 447]. The second model, the histone code hypothesis, proposes a role for these modifications as a site for recruiting other chromatin-associating proteins in which different histone modifications can act consecutively or in combination to form a histone code. This code will be recognised and read by other proteins to trigger downstream events that will generate various chromatin remodelling effects and translate the code into a particular chromatin state, either active or repressed. Multiple marks on histone domains may amplify these downstream events, subsequently leading to a larger change in chromatin structure of target sequences [27, 204, 369]. The list of histone modifications is increasing and all modifications appear to represent a mechanism for the regulation of chromatin. The exact way in which this histone code is established and maintained and the specific function of each modification has not yet been determined [74, 106, 369].

1.5.4 Imprinting and X-inactivation

The structure of chromatin is the result of DNA methylation, variation in histone composition, histone modifications, presence of non-histone proteins, the balance of all ATP-dependent chromatin remodelling complexes and the interaction of all these processes [11, 169, 189].

These modifications provide a mechanism to permanently turn off transcription of genes whose activity is not necessary in a particular stage of development or differentiated cell type [179]. Two good examples of these modifications are genomic imprinting and X-chromosome inactivation. In both processes the transcriptional repression is strongly associated with epigenetic changes of CpG islands within the promoters of silenced genes [38, 75, 76, 333].

In the case of genomic imprinting, i.e. a chromosomal modification that leads to the preferential expression of a paternal or maternal allele [403], differences in DNA methylation, histone modifications and conformation changes of the maternal or paternal allele will determine which allele will be expressed [104, 107, 119]. The mechanisms for the regulation of this type of gene expression are likely to be specific for each locus or region [119] and also include the involvement of insulator elements, boundary elements, repressor proteins and antisense transcripts [328]. Approximately 1% of all mammalian genes are imprinted [20].

Another group of genes associated with epigenetic modifications are genes located on the inactive X-chromosome of female mammals [309]. This process occurs early in embryogenesis and involves the random transcriptional silencing of one of the two X-chromosomes in females, necessary to ensure comparable levels of gene expression as observed in males (reviewed in [12]). This silencing is mainly characterised by histone hypoacetylation and DNA hypermethylation of the gene promoters [12]. The inactive X-chromosome is then stably inherited through subsequent cell divisions, so that only one X-chromosome is functional per cell [12, 321].

1.5.5 Chromatin remodelling and human disease

Mutations or inactivation of genes that are involved in the process of regulating or remodelling chromatin structure often lead to disease in humans. In cancers, for example, a change in DNA methylation status is one of the most common epigenetic alterations [75, 95, 179]. The regulation of methylation is disturbed that results in a genome-wide decrease in DNA methylation, which often includes repeated DNA sequences, but also shows an increase of DNA methylation in promoter regions of tumour suppressor genes important in the early development of cancer [96, 319]. This leads to a loss in gene function, oncogenic transformation and tumour progression [96, 104, 179], like observed for *pRb* in retinoblastoma [160], *APC* in colorectal carcinoma [160] and *BRCA1* in breast cancer [160]. DNA hypomethylation of proto-oncogenes may contribute to an increase in gene expression and promote translocations and mutations [333]. Furthermore, because DNA methylation prevents homologous recombination between repetitive elements, this overall DNA hypomethylation may also lead to an increase of inappropriate recombinations and eventually cause chromosomal instability, which is often observed in cancers [63, 96, 104, 291].

To date only one human disease is associated with mutations in a DNMT. Mutations in the *DNMT3B* gene located at locus 20q11.2 [418] are observed in patients with the ICF syndrome (i.e. immunodeficiency, centromeric region instability and facial anomalies syndrome), a very rare autosomal recessive disease that has been reported in fewer than 50 patients worldwide (reviewed in [97, 157]). This disease is characterised by facial abnormalities, psychomotor retardation and immunoglobulin deficiencies and, less frequently, mental retardation and digestive disorders [200]. Since mutations detected are either nonsense mutations leading to a premature termination of the translation process, or missense and splice-site mutations in the catalytic domain of the gene, researchers suggest that it is probably a reduced or even absent DNA methyltransferase activity that is responsible for the disease phenotype and the abnormal hypomethylation of specific satellite sequences, repeat sequences *NBL2* and *D4Z4* [176, 200, 270, 417], the *H19* gene [353] and several genes on the inactive X-chromosome [154].

Another inherited condition is Rett syndrome, a X-linked autosomal dominant neurodevelopmental disorder. The molecular defect in a large fraction of Rett syndrome patients are mutations in gene coding for the methyl-CpG binding protein MeCP2 that cause a deceleration in head growth, the loss of language and motor skills during early childhood and sporadic mental retardation in females [11, 122, 157, 209]. This disease is also very rarely observed in males that usually have an extra X-chromosome [157]. In addition, *MeCP2* mutations also cause neonatal-onset encephalopathy and autism [168].

Aberrant ATP-dependent remodelling complexes and histone-modifying enzymes have as well been associated with several human diseases [95, 157, 168]. Mutations of the *ATRX* gene at Xq13, which encodes an ATP-dependent helicase, cause not only mental retardation, but can also cause microcephaly and α -thalassemia together with facial, skeletal and urogenital abnormalities [168]. These characteristics are (in part) observed in the ATR-X syndrome, Juberg-Marsidi syndrome, Sutherland-Haan syndrome and Smith-Fineman-Myers syndrome [135, 157, 168]. The *ATRX* protein may act as a transcriptional regulator [30] and mutations result in unusual gain or loss of methylation of some highly repetitive DNA sequences [134]. Other diseases linked to mutations in ATP-dependent remodelling complexes are for example Cockayne syndrome type B, cerebro-oculo-facio-skeletal syndrome and Schimke immuno-osseous dysplasia, while the Rubinstein Taybi syndrome and mutations of *EP300* in several human cancers are associated with histone-modifying enzymes [168].

1.5.6 Epigenetic modifications and FSHD

The 4q45 region associated with FSHD is located on a subtelomere [419], which appears upon Giemsa staining as a dark band, indicative for a high level of chromatin condensation [378]. The region harbours D4Z4 that contains several sequence motifs associated with heterochromatic regions of the genome [2, 446] and has a high CpG content [159, 433]. Furthermore, D4Z4 is part of a family of repetitive elements that are mainly located at heterochromatic regions in the human genome [256]. These observations suggest that D4Z4 has a heterochromatic character.

Yang and colleagues [439] examined whether there was a heterochromatin-like or other atypical higher-order chromatin structure that could distinguish normal-sized D4Z4 arrays from contracted D4Z4 arrays using immuno-FISH in which the regions studied were in close proximity to D4Z4. The results indicated that no characteristics of constitutive heterochromatin were observed in myoblasts of control individuals nor of FSHD patients, since they observed no co-localisation with heterochromatic foci in interphase nuclei, no co-localisation with chromatin regions staining intensively for antibodies binding to heterochromatin protein 1 α or to H3 trimethylated at lysine 9, which are both enriched in constitutive heterochromatin, and there was no co-localisation of antibodies binding to H4 acetylation of lysine 8 and H3 methylation of lysine 4 that both bind preferentially to regions enriched in transcriptionally active genes. Moreover, no differences were observed between FSHD patients and control individuals in this limited set of samples, demonstrating no altered behaviour due to chromatin structure of FSHD-sized D4Z4 repeat arrays at the resolution of interphase FISH.

Since it is known that epigenetic influences on gene expression can also result from differential nuclear localisation of genomic loci [23, 118, 172, 273], the localisation of 4q35 within the interphase nucleus was examined as well. FISH analysis with a D4Z4-sequence probe revealed cross-hybridisation with nucleolar organising regions, which are heterochromatic sequences located on all acrocentric chromosomes that contain ribosomal DNA copies [2, 426]. The hybridisation signal was located interspersed both within the ribosomal DNA gene cluster and β -satellite sequences [419, 426], which led to the idea that the D4Z4 array might require a positioning of the 4q telomere near the nucleolus for proper gene expression. However, no differences in positioning of chromosome 4qter loci to the nucleolus were observed when FSHD interphase lymphoblasts were compared to those of control individuals [427, 427]. Recently, these findings were confirmed using immuno-FISH and, in addition, revealed that most of the 4q35 FISH signals were located at the nuclear periphery [260, 376, 439]. Furthermore, Masny *et al.* [260] observed that lamin A/C, a major component of the nuclear

lamina underlying the nuclear envelope, was necessary for the localisation of the 4q subtelomere to the nuclear periphery, while the D4Z4 repeat itself was not responsible. As chromatin is anchored to the nuclear periphery via lamin A/C, Lamin B, LAP2 and BAF [55, 294, 424], these observations may indicate that the chromatin structure of the FSHD-associated region may be directly associated with the nuclear envelope.

Given that epigenetic remodelling of chromatin is an ongoing area of research and the disease-causing mechanism for FSHD has not yet been found, we and other laboratories investigated if epigenetic modifications of the subtelomere on 4q might be important for FSHD pathology. In parallel with the localisation studies mentioned above we investigated the chromatin structure in which the D4Z4 repeat array is located and if this local structure would change as a consequence of the rearrangement observed in FSHD patients. Tsien *et al.* [385] observed high levels of DNA methylation of normal-sized D4Z4 arrays in a small control group, so we first concentrated on levels of DNA methylation in healthy controls and FSHD patients using digests of DNA with CpG methylation-sensitive enzymes (*Chapters 5 and 6*). In collaboration with colleagues in New Orleans we also measured histone acetylation patterns of the 4q35 region by performing chromatin immunoprecipitation with an antibody that discriminates between constitutive heterochromatin and unexpressed euchromatin (*Chapter 7*).

1.6 Postulated disease models for FSHD pathology

The forthcoming chapters (*Chapters 2-7*) will present new insights on structure and behaviour of the FSHD-associated D4Z4 repeat addressing (1) interactions of the subtelomeric region 4q35, in which D4Z4 resides, with other regions in the genome; (2) the consequences of mosaicism for FSHD; and (3) epigenetic modifications of the 4q35 region, including DNA methylation and histone acetylation. This section will provide an overview of the proposed models that are currently applied in FSHD research, implicating a direct or indirect role of D4Z4 in the development of FSHD. *Chapter 8* will discuss the validity of these models in relation to the data presented in this thesis.

Several observations made in the 4q35 region point to a possible role of chromatin structure or nuclear positioning of 4qter. First, D4Z4 is located in a subtelomere, which is a type of region that is more often involved in recombination than other parts of the genome [34, 64, 115]. This recombination-sensitivity is also emphasized by the high frequency of exchanges in this region [86, 295]. Second, several sequences within D4Z4 are associated with heterochromatin and, even though D4Z4-related sequences are scattered throughout the genome, they are mainly observed in heterochromatic regions. Finally, contracted repeats have been suggested to influence gene expression of (nearby) genes [121, 177, 428]. Different models have been proposed in which the D4Z4 contraction in this region located near the telomere plays a direct or indirect role in FSHD pathology. All models can be divided in three main groups: (1) direct effect of D4Z4; (2) changes in the chromatin structure of the 4q35 region; and (3) the nuclear localisation of the 4q35 region. These groups will be discussed in more detail below. One must take into account that all models assume that the consequences of the proposed model will eventually cause or influence FSHD pathology. However, supporting data is still not available to conclusively pinpoint all the phenotypic characteristics of FSHD to one defined molecular event.

1.6.1 Direct effect of D4Z4

Several models have suggested a role for the D4Z4 sequence itself in the development of FSHD. The most obvious one is that the gene responsible for the FSHD phenotype is located within the D4Z4 repeat array encoded by one or more repeat units or that the total array is part of a much larger gene. When a segment of the repeat array is contracted, this might result in loss of all or part of the FSHD gene if a minimum number of repeat units is required for proper gene function [85, 110]. Supporting this idea, each repeat unit contains a putative gene *DUX4*, the sequence of which suggests a role in regulation of gene expression possibly in a temporal,

regional or tissue-specific manner [123, 159]. However, haploinsufficiency of the 4q35 region does not cause disease, as individuals with a terminal deletion of 4qter including the whole repeat array and 200 kb proximal of the array do not have any features of FSHD [386].

Gabellini and colleagues [121] suggested, that binding of D4Z4 recognition complexes to a specific sequence in D4Z4 influences the phenotype. In healthy individuals these complexes containing the proteins YY1, HMGB2 and nucleolin would actively suppress gene expression, while in patients the number of bound repressor complexes is reduced and consequently results in a partial or complete decrease of transcriptional repression of the genes on 4q35 leading to the FSHD phenotype.

Lastly, Jiang *et al.* [98, 177] proposed a looping model in which communication between a contracted D4Z4 repeat array and a target gene (or genes) occurs in *cis* by intrachromosomal looping. This looping may extend over long distances and can therefore involve a target gene far away. Repeat arrays longer than 33 kb in size will in this model normally display intra-array looping, while arrays with 5-10 repeat units will form less stable intra-array loops and repeats with 1-4 repeat units will almost never form these interactions. This would explain the increased severity of symptoms and decreased age at diagnosis associated with very short D4Z4 repeat arrays. The abnormal intrachromosomal looping interactions may increase gene expression of the target gene by direct delivery of a transcription factor through this interaction, by changing chromatin structure of the promoter region of the target gene or by influencing the association that a target gene may have with a nuclear membrane or nuclear scaffold. A contracted D4Z4 repeat array may also activate inappropriate gene expression itself by the binding of the transcription factor Sp1 and (some of) the proteins that are located in the D4Z4 recognition complex proposed by Gabellini *et al.* [121], which can all contribute to gene expression through this intrachromosomal looping interaction.

1.6.2 Changes in chromatin structure of the 4q35 region

Expression of a gene can be strongly influenced by its position in the genome. When located in heterochromatin, a gene will not be easily accessible to the transcription machinery and therefore usually be silenced [372]. Most expressed genes are located in euchromatic areas with a more open chromatin structure. A positional effect (i.e. position effect variegation (PEV)) was first defined in *Drosophila* as a deleterious change in the level of gene expression brought about by a change in the position of the gene relative to its normal chromosomal environment, but not associated with intragenic mutations or deletions [193]. PEV can occur when a chromosomal rearrangement causes the juxtaposition of a euchromatic gene with a region of heterochromatin. The heterochromatic organisation spreads as a *cis*-acting silencing activity

into the juxtaposed euchromatic region and silences the nearby gene(s) [193, 372]. Position effects can occur over several hundred kb up to several Mb away from the rearrangement [414]. In humans, several diseases have been linked to a possible involvement of PEV (reviewed in [193]). For FSHD two different models postulate the role of a PEV-like mechanism. Both models result in a change of DNA conformation and subsequently in an aberrant expression of genes in the 4q35 region, which will eventually lead to the development of the FSHD phenotype.

The first model involves heterochromatinisation of the 4q35 region. The telomeric region of all chromosomes is composed of heterochromatin. D4Z4 contraction may disable the function of the array as a buffer zone between telomeric heterochromatin and proximal euchromatin region 4q35. This buffer zone will become smaller when the array is contracted and through spreading of telomeric heterochromatin in a proximal direction a position effect on gene(s) located proximal to the 4q telomere will be induced, thus inactivating genes located in the FSHD region [6, 85, 426].

The second model assumes a loss of heterochromatin and is also called “the loss-of-PEV-model” [177]. Three regions with homology to D4Z4, *LSau*, *hhspm3* and *Sau3A*, are located in heterochromatic regions in the human genome [268] and D4Z4 repeats have been associated with heterochromatic regions as well [256]. If D4Z4 repeat units are important for normal folding of the chromosome in a heterochromatic state, a contraction may result in a change in (local) chromatin conformation, consequently a loss of heterochromatin and thereby having an impact on gene expression by extinguishing or altering gene activity on 4q35 [159, 225, 226, 426]. For example, the contractions in the array may destabilise the heterochromatic character of D4Z4 itself, thereby permitting gene expression of *DUX4*. In muscle cells, the translated *DUX4* protein is then thought to be toxic due to a strong dimerisation potential shown in *in vitro* experiments, which may explain that severity roughly correlates with the residual repeat array size and that haploinsufficiency on 4q does not result in a disease phenotype [123]. This model thus suggests that a heterochromatic chromatin conformation of D4Z4 is a requirement for proper functioning of genes in the 4q35 region.

1.6.3 Nuclear localisation of the 4q35 region

The nuclear localisation model implies a pivotal role in the interactions and position for 4qter in the nucleus. FISH has been used to study the spatial orientation of, among others, the 4qter and 10qter subtelomeric regions in the interphase nucleus and revealed an increased somatic pairing frequency of subtelomeres when compared to interstitial chromosomal sites, depending on the specific regions involved [367]. In addition, a slightly more frequent pairing of the 4q with the 10q subtelomere in FSHD patients was detected that may facilitate possible

trans-interactions [367]. An inappropriate nuclear positioning of the FSHD region was proposed as a model by Winokur *et al.* [428] to explain the genome-wide disruption of genes observed in an expression profiling study. The transcriptional deregulation of genes involved in FSHD myogenesis and oxidative stress may arise from a global effect on gene regulation due to improper nuclear positioning of 4qter. The D4Z4 contractions may affect the localisation of the 4q telomere to a nuclear domain, in which aberrant gene expression can occur [428]. The same group recently examined the nuclear organisation of the FSHD region in patient myoblasts to gain more evidence for this improper nuclear positioning [260]. Masny and colleagues [260] observed that during the cell cycle 4q telomeres of both normal and FSHD alleles in these cells were specifically localised to the nuclear rim, an area in the nuclear periphery rich in nuclear envelope proteins, in contrast to the other telomeres measured (1q, 5q, 10q and 18q) that were dispersed throughout the nucleus. The presence of lamin A/C is necessary for this positioning, as absence of this protein leads to a drastic reduction in localisation of 4q35 to the nuclear rim. In addition, FISH experiments revealed that not the D4Z4 sequence on normal and FSHD alleles directs the location to the nuclear rim, but a sequence closer to the chromosome 4-specific locus D4S139. Based on these findings, the authors suggest that FSHD should be seen as a nuclear envelope disease [260].

1.7 References

1. Aboobaker A and Blaxter M (2003) Hox gene evolution in nematodes: novelty conserved. *Curr Opin Genet Dev* 13(6): 593-598.
2. 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.
3. Akiyama C, Suzuki H, and Nonaka I (1991) A case of facioscapulohumeral muscular dystrophy with infantile spasms, sensorineural deafness and retinal vessel abnormality. *No To Hattatsu* 23(4): 395-399.
4. Allis D and Fischle W (2003) Chromatin biology, histone modifications and human disease. *Upstate Biosignals* 2: 1-3.
5. Altherr MR, Bengtsson U, Elder FF, Ledbetter DH, Wasmuth JJ, McDonald ME, Gusella JF, and Greenberg F (1991) Molecular confirmation of Wolf-Hirschhorn syndrome with a subtle translocation of chromosome 4. *Am J Hum Genet* 49(6): 1235-1242.
6. Altherr MR, Bengtsson U, Markovich RP, and Winokur ST (1995) Efforts toward understanding the molecular basis of facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2: 32-38.
7. Antequera F and Bird A (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci USA* 90(24): 11995-11999.
8. Arahata K, Goto K, Yonemoto K, and Fukanoshi M (1998) DNA rearrangements in Japanese patients with facioscapulohumeral muscular dystrophy (FSHD). *Muscle Nerve Supplement* 7: S25.
9. Arahata K, Ishihara T, Fukunaga H, Orimo S, Lee JH, Goto K, and Nonaka I (1995) Inflammatory response in facioscapulohumeral muscular dystrophy (FSHD): immunocytochemical and genetic analyses. *Muscle Nerve* 2: S56-S66.
10. Attarian S, Malzac P, Azulay J, Pelisser J, Philip N, and Pouget J (1998) Phenotype-genotype correlations studies in facioscapulohumeral muscular dystrophy. *Muscle Nerve Supplement* 7: S111.
11. Attwood JT, Yung RL, and Richardson BC (2002) DNA methylation and the regulation of gene transcription. *Cell Mol Life Sci* 59(2): 241-257.
12. Avner P and Heard E (2001) X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* 2(1): 59-67.
13. Baird DM and Royle NJ (1997) Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. *Hum Mol Genet* 6(13): 2291-2299.
14. Bakker E, Wielen van der MJ, Voorhoeve E, Ippel PF, Padberg GW, Frants RR, and Wijmenga C (1996) Diagnostic, predictive, and prenatal testing for facioscapulohumeral muscular dystrophy: diagnostic approach for sporadic and familial cases. *J Med Genet* 33(1): 29-35.
15. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, Wielen van der M, Rasmussen K, and Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2: 39-44.
16. Ballarati L, Piccini I, Carbone L, Archidiacono N, Rollier A, Marozzi A, Meneveri R, and Ginelli E (2002) Human genome dispersal and evolution of 4q35 duplications and interspersed LSau repeats. *Gene* 296(1-2): 21-27.
17. Barbierato L, Gobbi P, Ferlini A, and Tupler R (1998) Atypical clinical progression of FSH disease in two brothers and their mother: the puzzle of genotype-phenotype correlation. *Muscle Nerve Supplement* 7: S111.
18. Barletta J and Greer SB (1992) Methylation of HSV-1 DNA as a mechanism of viral inhibition: studies of an analogue of methyldeoxycytidine: trifluoromethyldeoxycytidine (F3mdCyd). *Antiviral Res* 18(1): 1-25.

19. Barlow AL and Hulten MA (1996) Combined immunocytogenetic and molecular cytogenetic analysis of meiosis I human spermatocytes. *Chromosome Res* 4(8): 562-573.
20. Barlow DP (1995) Gametic imprinting in mammals. *Science* 270(5242): 1610-1613.
21. Bastress KL, Stajich J, Gilbert J, and Speer MC (2004) Linkage analysis in non-chromosome 4-linked FSHD. Chapter 19: 299-308. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
22. Bastress KL, Stajich JM, Lemmers RJLF, Maarel van der SM, Speer MC, and Gilbert JR (2004) D4Z4 fragment analysis and genome-wide SNP screening in non-chromosome 4-linked facioscapulohumeral muscular dystrophy. FSHD International Research Consortium Workshop.
23. Baxter J, Merckenschlager M, and Fisher AG (2002) Nuclear organisation and gene expression. *Curr Opin Cell Biol* 14(3): 372-376.
24. Becker J, Schwaab R, Moller-Taube A, Schwaab U, Schmidt W, Brackmann HH, Grimm T, Olek K, and Oldenburg J (1996) Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type- dependent sex ratio of mutation frequencies. *Am J Hum Genet* 58(4): 657-670.
25. Beckers M, Gabriels J, Maarel van der SM, Vriese de A, Frants RR, Collen D, and Belayew A (2001) Active genes in junk DNA? Characterization of *DUX* genes embedded within 3.3 kb repeated elements. *Gene* 264(1): 51-57.
26. Belzacq AS, Vieira HL, Kroemer G, and Brenner C (2002) The adenine nucleotide translocator in apoptosis. *Biochimie* 84(2-3): 167-176.
27. Berger SL (2001) An embarrassment of niches: the many covalent modifications of histones in transcriptional regulation. *Oncogene* 20(24): 3007-3013.
28. Berger SL (2002) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 12(2): 142-148.
29. Bernards A and Gusella JF (1994) The importance of genetic mosaicism in human disease. *N Engl J Med* 331(21): 1447-1449.
30. Bérubé NG, Smeenk CA, and Picketts DJ (2000) Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. *Hum Mol Genet* 9(4): 539-547.
31. Bestor TH (2000) The DNA methyltransferases of mammals. *Hum Mol Genet* 9(16): 2395-2402.
32. Bestor TH and Tycko B (1996) Creation of genomic methylation patterns. *Nat Genet* 12(4): 363-367.
33. Bianchi NO, Richard SM, Peltomaki P, and Bianchi MS (2002) Mosaic AZF deletions and susceptibility to testicular tumors. *Mutat Res* 503(1-2): 51-62.
34. Biessmann H and Mason JM (1992) Genetics and molecular biology of telomeres. *Adv Genet* 30: 185-249.
35. Bird AP (1992) The essentials of DNA methylation. *Cell* 70(1): 5-8.
36. Bird AP (1995) Gene number, noise reduction and biological complexity. *Trends Genet* 11(3): 94-100.
37. Bird AP (1999) DNA methylation de novo. *Science* 286(5448): 2287-2288.
38. Bird AP (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16(1): 6-21.
39. Bird AP and Wolffe AP (1999) Methylation-induced repression-belts, braces, and chromatin. *Cell* 99(5): 451-454.
40. Bohmann K, Ferreira J, Santama N, Weis K, and Lamond AI (1995) Molecular analysis of the coiled body. *J Cell Sci Supplement* 19: 107-113.
41. Bohmann K, Ferreira JA, and Lamond AI (1995) Mutational analysis of p80 coilin indicates a functional interaction between coiled bodies and the nucleolus. *J Cell Biol* 131(4): 817-831.
42. Bond VC and Wold B (1993) Nucleolar localization of myc transcripts. *Mol Cell Biol* 13(6): 3221-3230.

43. Bonner WM, Hatch CL, and Wu RS (1988) Ubiquitinated histones and chromatin. Chapter 6: 157-172. In: Ubiquitin. Rechsteiner M (Ed.) Plenum Press, New York and London.
44. Bouju S, Pietu G, Le Cunff M, Cros N, Malzac P, Pellissier JF, Pons F, Leger JJ, Auffray C, and Dechesne CA (1999) Exclusion of muscle specific actinin-associated LIM protein (ALP) gene from 4q35 facioscapulohumeral muscular dystrophy (FSHD) candidate genes. *Neuromusc Disord* 9(1): 3-10.
45. Boyes J and Bird A (1992) Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J* 11(1): 327-333.
46. Bradbury EM (1992) Reversible histone modifications and the chromosome cell cycle. *Bioessays* 14(1): 9-16.
47. Brandeis M, Kafri T, Ariel M, Chaillet JR, McCarrey J, Razin A, and Cedar H (1993) The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J* 12(9): 3669-3677.
48. Brooke MH (1977) A clinician's view of neuromuscular diseases. Williams & Wilkins, Baltimore.
49. Brouwer OF, Padberg GW, Bakker E, Wijmenga C, and Frants RR (1995) Early onset facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2: S67-S72.
50. Brouwer OF, Padberg GW, Ruys CJ, Brand R, Laat de JA, and Grote JJ (1991) Hearing loss in facioscapulohumeral muscular dystrophy. *Neurology* 41(12): 1878-1881.
51. Brouwer OF, Padberg GW, Wijmenga C, and Frants RR (1994) Facioscapulohumeral muscular dystrophy in early childhood. *Arch Neurol* 51(4): 387-394.
52. Brown LY and Brown SA (2004) Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet* 20(1): 51-58.
53. Brown TA (2002) Genome anatomies. Chapter 2: 59-64. In: Genomes 2. Carlson S (Ed.) BIOS Scientific Publishers Limited, Oxford.
54. Brown WR, MacKinnon PJ, Villasante A, Spurr N, Buckle VJ, and Dobson MJ (1990) Structure and polymorphism of human telomere-associated DNA. *Cell* 63(1): 119-132.
55. Burke B and Stewart CL (2002) Life at the edge: the nuclear envelope and human disease. *Nat Rev Mol Cell Biol* 3(8): 575-585.
56. Bushby KM, Pollitt C, Johnson MA, Rogers MT, and Chinnery PF (1998) Muscle pain as a prominent feature of facioscapulohumeral muscular dystrophy (FSHD): four illustrative case reports. *Neuromusc Disord* 8(8): 574-579.
57. Butz M, Koch MC, Muller-Felber W, Lemmers RJ, Maarel van der SM, and Schreiber H (2003) Facioscapulohumeral muscular dystrophy. Phenotype-genotype correlation in patients with borderline D4Z4 repeat numbers. *J Neurol* 250(8): 932-937.
58. Cacurri S, Deidda G, Piazzi N, Novelletto A, La C, I, Servidei S, Galluzzi G, Wijmenga C, Frants RR, and Felicetti L (1994) Chromosome 4q35 haplotypes and DNA rearrangements segregating in affected subjects of 19 Italian families with facioscapulohumeral muscular dystrophy (FSHD). *Hum Genet* 94(4): 367-374.
59. Carmo-Fonseca M (2002) New clues to the function of the Cajal body. *EMBO Rep* 3(8): 726-727.
60. Cervantes-Laurean D, Jacobson EL, and Jacobson MK (1996) Glycation and glycoxidation of histones by ADP-ribose. *J Biol Chem* 271(18): 10461-10469.
61. Chandley AC (1989) Asymmetry in chromosome pairing: a major factor in de novo mutation and the production of genetic disease in man. *J Med Genet* 26(9): 546-552.
62. Chen HY, Sun JM, Zhang Y, Davie JR, and Meistrich ML (1998) Ubiquitination of histone H3 in elongating spermatids of rat testes. *J Biol Chem* 273(21): 13165-13169.
63. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, and Jaenisch R (1998) DNA hypomethylation leads to elevated mutation rates. *Nature* 395(6697): 89-93.
64. Cheng JF, Smith CL, and Cantor CR (1991) Structural and transcriptional analysis of a human subtelomeric repeat. *Nucleic Acids Res* 19(1): 149-154.

65. Cheung P and Lau P (2005) Epigenetic regulation by histone methylation and histone variants. *Mol Endocrinol* 19(3): 563-573.
66. Cheung WL, Briggs SD, and Allis CD (2000) Acetylation and chromosomal functions. *Curr Opin Cell Biol* 12(3): 326-333.
67. Cho KS, Elizondo LI, and Boerkoel CF (2004) Advances in chromatin remodeling and human disease. *Curr Opin Genet Dev* 14(3): 308-315.
68. Cillo C, Cantile M, Faiella A, and Boncinelli E (2001) Homeobox genes in normal and malignant cells. *J Cell Physiol* 188(2): 161-169.
69. Clapp J, Fantes J, and Hewitt JE (2004) Identification of potential homologues of the human repeat D4Z4, deletions of which are causally associated with facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 75(4; Supplement) Abstract #1576.
70. Clark LN, Koehler U, Ward DC, Wienberg J, and Hewitt JE (1996) Analysis of the organisation and localisation of the FSHD-associated tandem array in primates: implications for the origin and evolution of the 3.3 kb repeat family. *Chromosoma* 105(3): 180-189.
71. Clayton-Smith J and Laan L (2003) Angelman syndrome: a review of the clinical and genetic aspects. *J Med Genet* 40: 87-95.
72. Cohn DH, Starman BJ, Blumberg B, and Byers PH (1990) Recurrence of lethal osteogenesis imperfecta due to parental mosaicism for a dominant mutation in a human type I collagen gene (COL1A1). *Am J Hum Genet* 46(3): 591-601.
73. Coppée F, Mattéotti C, Anseau 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.
74. Cosgrove MS, Boeke JD, and Wolberger C (2004) Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* 11(11): 1037-1043.
75. Costello JF and Plass C (2001) Methylation matters. *J Med Genet* 38(5): 285-303.
76. Cross SH and Bird AP (1995) CpG islands and genes. *Curr Opin Genet Dev* 5(3): 309-314.
77. Cummings CJ and Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* 9(6): 909-916.
78. D'Elia AV, Tell G, Paron I, Pellizzari L, Lonigro R, and Damante G (2001) Missense mutations of human homeoboxes: A review. *Hum Mutat* 18(5): 361-374.
79. Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, and Kiss T (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* 21(11): 2746-2756.
80. Davie JR (1998) Covalent modifications of histones: expression from chromatin templates. *Curr Opin Genet Dev* 8(2): 173-178.
81. Davie JR and Spencer VA (1999) Control of histone modifications. *J Cell Biochem Supplement* 32-33: 141-148.
82. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, and Felicetti L (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3: 155-167.
83. Deidda G, Cacurri S, Piazzo N, and Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet* 33(5): 361-365.
84. Deutekom van JC, Hofker MH, Romberg S, Geel van M, Rommens J, Wright TJ, Hewitt JE, Padberg GW, Wijmenga C, and Frants RR (1995) Search for the FSHD gene using cDNA selection in a region spanning 100 kb on chromosome 4q35. *Muscle Nerve* 2: S19-S26.
85. Deutekom van JCT (1996) Towards the molecular mechanism of facioscapulohumeral muscular dystrophy. Thesis, Leiden University.

86. 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.
87. Deutekom van JCT, Lemmers RJLF, Grewal PK, Geel van M, Romberg S, Dauwerse HG, Wright TJ, Padberg GW, Hofker MH, Hewitt JE, *et al.* (1996) Identification of the first gene (*FRG1*) from the FSHD region on human chromosome 4q35. *Hum Mol Genet* 5(5): 581-590.
88. 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.
89. Dillon N and Festenstein R (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet* 18(5): 252-258.
90. Ding H, Beckers MC, Plaisance S, Marynen P, Collen D, and Belayew A (1998) Characterization of a double homeodomain protein (DUX1) encoded by a cDNA homologous to 3.3 kb dispersed repeated elements. *Hum Mol Genet* 7(11): 1681-1694.
91. Dirks RW, Hattinger CM, Molenaar C, and Snaar SP (1999) Synthesis, processing, and transport of RNA within the three-dimensional context of the cell nucleus. *Crit Rev Eukaryot Gene Expr* 9(3-4): 191-201.
92. Dobosy JR and Selker EU (2001) Emerging connections between DNA methylation and histone acetylation. *Cell Mol Life Sci* 58(5-6): 721-727.
93. Doerner A, Pauschinger M, Badorff A, Noutsias M, Giessen S, Schulze K, Bilger J, Rauch U, and Schultheiss HP (1997) Tissue-specific transcription pattern of the adenine nucleotide translocase isoforms in humans. *FEBS Lett* 414(2): 258-262.
94. Doggett NA, Stallings RL, Hildebrand CE, and Moyzis RK (1992) The mapping of chromosome 16. *Los Alamos Science* 20: 182-198.
95. Egger G, Liang G, Aparicio A, and Jones PA (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429: 457-463.
96. Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. *Oncogene* 21(35): 5400-5413.
97. Ehrlich M (2003) The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clin Immunol* 109(1): 17-28.
98. 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.
99. Elder JF, Jr. and Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol* 70(3): 297-320.
100. Essen van AJ, Abbs S, Baiget M, Bakker E, Boileau C, Broeckhoven van C, Bushby K, Clarke A, Claustres M, Covone AE, *et al.* (1992) Parental origin and germline mosaicism of deletions and duplications of the dystrophin gene: a European study. *Hum Genet* 88(3): 249-257.
101. Evans DG, Wallace AJ, Wu CL, Trueman L, Ramsden RT, and Strachan T (1998) Somatic mosaicism: a common cause of classic disease in tumor-prone syndromes? Lessons from type 2 neurofibromatosis. *Am J Hum Genet* 63(3): 727-736.
102. Fan YS, Yang HM, and Lin CC (1992) Assignment of the human muscle adenine nucleotide translocator gene (*ANT1*) to 4q35 by fluorescence in situ hybridization. *Cytogenet Cell Genet* 60(1): 29-30.
103. Faustmann PM, Farahati J, Rupilius B, Dux R, Koch MC, and Reiners C (1996) Cardiac involvement in facio-scapulo-humeral muscular dystrophy: a family study using Thallium-201 single-photon-emission-computed tomography. *J Neurol Sci* 144(1-2): 59-63.

104. Fazzari MJ and Greally JM (2004) Epigenomics: beyond CpG islands. *Nat Rev Genet* 5(6): 446-455.
105. Felice KJ, North WA, Moore SA, and Mathews KD (2000) FSH dystrophy 4q35 deletion in patients presenting with facial-sparing scapular myopathy. *Neurology* 54(10): 1927-1931.
106. Felsenfeld G and Groudine M (2003) Controlling the double helix. *Nature* 421(6921): 448-453.
107. Ferguson-Smith AC and Surani MA (2001) Imprinting and the epigenetic asymmetry between parental genomes. *Science* 293(5532): 1086-1089.
108. Figlewicz DA, Barrett K, Haefele Leskova A, Forrester JR, Sowden JE, and Tawil R (2004) FSHD myoblasts: *in vitro* studies. Chapter 16: 235-251. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
109. Figlewicz DA, Sowden JE, Haefele Leskova A, Forrester JR, Barrett K, Kavcic V, and Tawil R (2002) Facioscapulohumeral dystrophy: premature activation of the myogenic program? *Am J Hum Genet* 71(4; Supplement): 527.
110. Fisher J and Upadhyaya M (1997) Molecular genetics of facioscapulohumeral muscular dystrophy (FSHD). *Neuromusc Disord* 7(1): 55-62.
111. Fitzpatrick DR and Wilson CB (2003) Methylation and demethylation in the regulation of genes, cells, and responses in the immune system. *Clin Immunol* 109(1): 37-45.
112. Fitzsimons RB (1999) Facioscapulohumeral muscular dystrophy. *Curr Opin Neurol* 12(5): 501-511.
113. Fitzsimons RB, Gurwin EB, and Bird AC (1987) Retinal vascular abnormalities in facioscapulohumeral muscular dystrophy. A general association with genetic and therapeutic implications. *Brain* 110 (Pt 3): 631-648.
114. Flint J, Bates GP, Clark K, Dorman A, Willingham D, Roe BA, Micklem G, Higgs DR, and Louis EJ (1997) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet* 6(8): 1305-1314.
115. Flint J, Rochette J, Craddock CF, Dode C, Vignes B, Horsley SW, Kearney L, Buckle VJ, Ayyub H, and Higgs DR (1996) Chromosomal stabilisation by a subtelomeric rearrangement involving two closely related Alu elements. *Hum Mol Genet* 5(8): 1163-1169.
116. Flint J, Wilkie AOM, Buckle VJ, Winter RM, Holland AJ, and McDermid HE (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* 9(2): 132-140.
117. Ford CE, Polani PE, Briggs JH, and Bishop PMF (1959) A presumptive human XXY/XX mosaic. *Nature* 183: 1030-1032.
118. Francastel C, Schubeler D, Martin DI, and Groudine M (2000) Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol* 1(2): 137-143.
119. Fulmer-Smentek SB and Francke U (2001) Association of acetylated histones with paternally expressed genes in the Prader-Willi deletion region. *Hum Mol Genet* 10(6): 645-652.
120. Funakoshi M, Goto K, and Arahata K (1998) Epilepsy and mental retardation in a subset of early onset 4q35- facioscapulohumeral muscular dystrophy. *Neurology* 50(6): 1791-1794.
121. 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.
122. Gabellini D, Tupler R, and Green MR (2003) Transcriptional derepression as a cause of genetic diseases. *Curr Opin Genet Dev* 13(3): 239-245.
123. Gabriels J, Beckers MC, Ding H, Vriese de A, Plaisance S, Maarel van der SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al.* (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 236(1): 25-32.

124. Galluzzi G, Deidda G, Cacurri S, Colantoni L, Piazzo N, Vigneti E, Ricci E, Servidei S, Merico B, Pachi A, *et al.* (1999) Molecular analysis of 4q35 rearrangements in facioscapulohumeral muscular dystrophy (FSHD): application to family studies for a correct genetic advice and a reliable prenatal diagnosis of the disease. *Neuromusc Disord* 9(3): 190-198.
125. Galvao R, Mendes-Soares L, Camara J, Jaco I, and Carmo-Fonseca M (2001) Triplet repeats, RNA secondary structure and toxic gain-of-function models for pathogenesis. *Brain Res Bull* 56(3-4): 191-201.
126. Geel van M (2001) Facioscapulohumeral muscular dystrophy, a complex human disease entangled in dynamic molecular evolution. Thesis, Leiden University.
127. Geel van M, Deutekom van JC, Staalduinen van A, Lemmers RJLF, Dickson MC, Hofker MH, Padberg GW, Hewitt JE, Jong de PJ, and Frants RR (2000) Identification of a novel beta-tubulin subfamily with one member (TUBB4Q) located near the telomere of chromosome region 4q35. *Cytogenet Cell Genet* 88(3-4): 316-321.
128. 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.
129. Geel van M, Heather LJ, Lyle R, Hewitt JE, Frants RR, and Jong de PJ (1999) The FSHD region on human chromosome 4q35 contains potential coding regions among pseudogenes and a high density of repeat elements. *Genomics* 61(1): 55-65.
130. Gelehrter TD, Collins FS, and Ginsburg D (1998) Cytogenetics. Chapter 8: 153-194. In: *Principles of medical genetics*. Kelly PJ (Ed.) Williams & Wilkins, Baltimore, Maryland.
131. Gelehrter TD, Collins FS, and Ginsburg D (1998) Glossary. In: *Principles of medical genetics*. Kelly PJ (Ed.) Williams & Wilkins, Baltimore, Maryland.
132. Gelehrter TD, Collins FS, and Ginsburg D (1998) Structure and behavior of genes and chromosomes. Chapter 2: 9-22. In: *Principles of medical genetics*. Kelly PJ (Ed.) Williams & Wilkins, Baltimore, Maryland.
133. Gelehrter TD, Collins FS, and Ginsburg D (1998) The human genome project. Chapter 10: 231-243. In: *Principles of medical genetics*. Kelly PJ (Ed.) Williams & Wilkins, Baltimore, Maryland.
134. Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, and Higgs DR (2000) Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* 24(4): 368-371.
135. Gibbons RJ, Picketts DJ, Villard L, and Higgs DR (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* 80(6): 837-845.
136. Gilbert JR, Speer MC, Stajich J, Clancy R, Lewis K, Qiu H, Yamaoka L, Kumar A, Vance J, and Stewart C (1995) Exclusion mapping of chromosomal regions which cross hybridise to FSHD1A associated markers in FSHD1B. *J Med Genet* 32(10): 770-773.
137. Gilbert JR, Stajich JM, Speer MC, Vance JM, Stewart CS, Yamaoka LH, Samson F, Fardeau M, Potter TG, Roses AD, *et al.* (1992) Linkage studies in facioscapulohumeral muscular dystrophy (FSHD). *Am J Hum Genet* 51(2): 424-427.
138. Gilbert JR, Stajich JM, Wall S, Carter SC, Qiu H, Vance JM, Stewart CS, Speer MC, Pufky J, Yamaoka LH, *et al.* (1993) Evidence for heterogeneity in facioscapulohumeral muscular dystrophy (FSHD). *Am J Hum Genet* 53(2): 401-408.
139. Gilbert SL and Sharp PA (1999) Promoter-specific hypoacetylation of X-inactivated genes. *Proc Natl Acad Sci USA* 96(24): 13825-13830.
140. Goll MG and Bestor TH (2002) Histone modification and replacement in chromatin activation. *Genes Dev* 16(14): 1739-1742.
141. Gomes-Pereira M, Fortune MT, Ingram L, McAbney JP, and Monckton DG (2004) Pms2 is a genetic enhancer of trinucleotide CAG/CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Genet* 13(16): 1815-1825.
142. Goto K, Lee JH, Matsuda C, Hirabayashi K, Kojo T, Nakamura A, Mitsunaga Y, Furukawa T, Sahashi K, and Arahata K (1995) DNA rearrangements in Japanese

- facioscapulohumeral muscular dystrophy patients: clinical correlations. *Neuromusc Disord* 5(3): 201-208.
143. 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.
 144. Gottlieb B, Beitel LK, and Trifiro MA (2001) Somatic mosaicism and variable expressivity. *Trends Genet* 17(2): 79-82.
 145. Grant PA (2001) A tale of histone modifications. *Genome Biol* 2(4): review 0003.1-6.
 146. Grant PA, Sterner DE, Duggan LJ, Workman JL, and Berger SL (1998) The SAGA unfolds: convergence of transcription regulators in chromatin-modifying complexes. *Trends Cell Biol* 8(5): 193-197.
 147. Grewal PK, Deutekom van JCT, Mills KA, Lemmers RJLF, Mathews KD, Frants RR, and Hewitt JE (1997) The mouse homolog of FRG1, a candidate gene for FSHD, maps proximal to the myodystrophy mutation on Chromosome 8. *Mamm Gen* 8(6): 394-398.
 148. Griggs RC, Tawil R, Storvick D, Mendell JR, and Altherr MR (1993) Genetics of facioscapulohumeral muscular dystrophy: new mutations in sporadic cases. *Neurology* 43(11): 2369-2372.
 149. Grunau C, Renault E, Rosenthal A, and Roizes G (2001) MethDB - a public database for DNA methylation data. *Nucleic Acids Res* 29(1): 270-274.
 150. Gundersen GG (2002) Evolutionary conservation of microtubule-capture mechanisms. *Nat Rev Mol Cell Biol* 3(4): 296-304.
 151. Häfner FM, Salam AA, Linder TE, Balmer D, Baumer A, Schinzel AA, Spillmann T, and Leal SM (2000) A novel locus (DFNA24) for prelingual nonprogressive autosomal dominant nonsyndromic hearing loss maps to 4q35-qter in a large Swiss German kindred. *Am J Hum Genet* 66(4): 1437-1442.
 152. Hake SB, Xiao A, and Allis CD (2004) Linking the epigenetic 'language' of covalent histone modifications to cancer. *Br J Cancer* 90(4): 761-769.
 153. Hall JG (1988) Review and hypotheses: somatic mosaicism: observations related to clinical genetics. *Am J Hum Genet* 43(4): 355-363.
 154. Hansen RS, Stoger R, Wijmenga C, Stanek AM, Canfield TK, Luo P, Matarazzo MR, D'Esposito M, Feil R, Gimelli G, *et al.* (2000) Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. *Hum Mol Genet* 9(18): 2575-2587.
 155. Haraguchi Y, Chung AB, Torroni A, Stepien G, Shoffner JM, Wasmuth JJ, Costigan DA, Polak M, Altherr MR, Winokur ST, *et al.* (1993) Genetic mapping of human heart-skeletal muscle adenine nucleotide translocator and its relationship to the facioscapulohumeral muscular dystrophy locus. *Genomics* 16: 479-485.
 156. Hauser MA, Conde CB, Kowalchow V, Zeppa G, Taratuto AL, Torian UM, Vance J, Pericak-Vance MA, Speer MC, *et al.* (2002) *myotilin* Mutation found in second pedigree with LGMD1A. *Am J Hum Genet* 71(6): 1428-1432.
 157. Hendrich B and Bickmore W (2001) Human diseases with underlying defects in chromatin structure and modification. *Hum Mol Genet* 10(20): 2233-2242.
 158. Hewitt JE and Bolland DJ (2002) The FSHD gene region on chromosome 4q35 - a gene desert? *Neuromusc Disord* 12: 741.
 159. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, Deutekom van JC, Francis F, Sharpe PT, Hofker M, *et al.* (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3(8): 1287-1295.
 160. Hiltunen MO, Alhonen L, Koistinaho J, Myohanen S, Paakkonen M, Marin S, Kosma VM, and Janne J (1997) Hypermethylation of the APC (adenomatous polyposis coli) gene promoter region in human colorectal carcinoma. *Int J Cancer* 70(6): 644-648.
 161. Hirano M and DiMauro S (2001) ANT1, Twinkle, POLG, and TP: new genes open our eyes to ophthalmoplegia. *Neurology* 57(12): 2163-2165.
 162. Horn PJ and Peterson CL (2002) Chromatin higher order folding - wrapping up transcription. *Science* 297(5588): 1824-1827.

163. Horvath JE, Bailey JA, Locke DP, and Eichler EE (2001) Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum Mol Genet* 10(20): 2215-2223.
164. Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, and Choo KH (2000) Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci USA* 97(3): 1148-1153.
165. Hsieh CL (1994) Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol* 14(8): 5487-5494.
166. Hsieh J and Fire A (2000) Recognition and silencing of repeated DNA. *Annu Rev Genet* 34: 187-204.
167. 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.
168. Huang C, Sloan EA, and Boerkoel CF (2003) Chromatin remodeling and human disease. *Curr Opin Gene Dev* 13(3): 246-252.
169. Hung MS and Shen CK (2003) Eukaryotic methyl-CpG-binding domain proteins and chromatin modification. *Eukaryot Cell* 2(5): 841-846.
170. Iizuka M and Smith MM (2003) Functional consequences of histone modifications. *Curr Opin Genet Dev* 13(2): 154-160.
171. Iqbal Z, Roper M, Pericak-Vance MA, Hung W-Y, DeLong R, Cummings WJK, and Siddique T (1992) Genetic heterogeneity in facioscapulohumeral disease. *Am J Hum Genet* 51(4; supplement): A191.
172. Isogai Y and Tjian R (2003) Targeting genes and transcription factors to segregated nuclear compartments. *Curr Opin Cell Biol* 15(3): 296-303.
173. Jardine PE, Koch MC, Lunt PW, Maynard J, Bathke KD, Harper PS, and Upadhyaya M (1994) De novo facioscapulohumeral muscular dystrophy defined by DNA probe p13E-11 (D4F104S1). *Arch Dis Child* 71(3): 221-227.
174. Jasinska A, Michlewski G, Mezer de M, Sobczak K, Kozlowski P, Napierala M, and Krzyzosiak WJ (2003) Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Res* 31(19): 5463-5468.
175. Jason LJ, Moore SC, Lewis JD, Lindsey G, and Ausio J (2002) Histone ubiquitination: a tagging tail unfolds? *Bioessays* 24(2): 166-174.
176. Jeanpierre M, Turleau C, Aurias A, Prieur M, Ledeist F, Fischer A, and Viegas-Pequignot E (1993) An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum Mol Genet* 2(6): 731-735.
177. 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.
178. Johnson CA, O'Neill LP, Mitchell A, and Turner BM (1998) Distinctive patterns of histone H4 acetylation are associated with defined sequence elements within both heterochromatic and euchromatic regions of the human genome. *Nucleic Acids Res* 26(4): 994-1001.
179. Jones PA (1999) The DNA methylation paradox. *Trends Genet* 15(1): 34-37.
180. Jones PA and Laird PW (1999) Cancer epigenetics comes of age. *Nat Genet* 21(2): 163-167.
181. Jones PA and Takai D (2001) The role of DNA methylation in mammalian epigenetics. *Science* 293(5532): 1068-1070.
182. Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, Cedar H, and Razin A (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* 6(5): 705-714.
183. Kao CF, Hillyer C, Tsukuda T, Henry K, Berger S, and Osley MA (2004) Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. *Genes Dev* 18(2): 184-195.

184. Kasinsky HE, Lewis JD, Dacks JB, and Ausio J (2001) Origin of H1 linker histones. *FASEB J* 15(1): 34-42.
185. Kass SU, Pruss D, and Wolffe AP (1997) How does DNA methylation repress transcription? *Trends Genet* 13(11): 444-449.
186. Keshet I, Lieman-Hurwitz J, and Cedar H (1986) DNA methylation affects the formation of active chromatin. *Cell* 44(4): 535-543.
187. Kilmer DD, Abresch RT, Aitkens SG, Carter GT, Fowler WM, Jr., Johnson ER, McDonald CM, and Wright NJ (2004) Impairment and disability of neuromuscular diseases: facioscapulohumeral dystrophy. RTC/Neuromuscular diseases, Department of physical medicine and rehabilitation, University of California, Davis. Internet Communication
188. Kilmer DD, Abresch RT, McCrory MA, Carter GT, Fowler WM, Jr., Johnson ER, and McDonald CM (1995) Profiles of neuromuscular diseases. Facioscapulohumeral muscular dystrophy. *Am J Phys Med Rehabil* 74(5 Supplement): S131-S139.
189. Kimura A, Umehara T, and Horikoshi M (2002) Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nat Genet* 32(3): 370-377.
190. Kinyamu HK and Archer TK (2004) Modifying chromatin to permit steroid hormone receptor-dependent transcription. *Biochim Biophys Acta* 1677(1-3): 30-45.
191. Kirschner MW (1978) Microtubule assembly and nucleation. *Int Rev Cytol* 54: 1-71.
192. Kissel JT (1999) Facioscapulohumeral dystrophy. *Semin Neurol* 19(1): 35-43.
193. Kleinjan DJ and Heyningen van V (1998) Position effects in human genetic disease. *Hum Mol Genet* 7(10): 1611-1618.
194. Kluwe L, Mautner V, Heinrich B, Dezube R, Jacoby LB, Friedrich RE, and MacCollin M (2003) Molecular study of frequency of mosaicism in neurofibromatosis 2 patients with bilateral vestibular schwannomas. *J Med Genet* 40(2): 109-114.
195. Koch MC and Schreiber H (2003) Fazioskapulohumerale Muskeldystrophie - Klinische und molekulargenetische Aspekte. 6. Jahrgang: 1-12. ARCIS Verlag GmbH München. Management of Neuromuscular Disease Letter; Information für medizinische Fachkreise. Dengler, R and Pongratz, D.
196. Koetsier C (1997) Pijn bij FSHD. Een onderschat probleem? Bespreking resultaten van de pijnenquête-FSHD. *VSN Nieuwsbrief FSHD* 17.
197. Koetsier C (1997) Pijn bij FSHD. Een onderschat probleem? Resultaten van de pijnenquête-FSHD. *VSN Nieuwsbrief FSHD* 16.
198. 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.
199. Köhler J, Rupilius B, Otto M, Bathke K, and Koch MC (1996) Germline mosaicism in 4q35 facioscapulohumeral muscular dystrophy (FSHD1A) occurring predominantly in oogenesis. *Hum Genet* 98(4): 485-490.
200. 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.
201. Koningsbruggen van S, Dirks RW, Mommaas AM, Onderwater JJ, Deidda G, Padberg GW, Frants RR, and Maarel van der SM (2004) FRG1P is localised in the nucleolus, Cajal bodies, and speckles. *J Med Genet* 41(4): e46.
202. Korf BR, Bresnan MJ, Shapiro F, Sotrel A, and Abroms IF (1985) Facioscapulohumeral dystrophy presenting in infancy with facial diplegia and sensorineural deafness. *Ann Neurol* 17(5): 513-516.
203. Kornberg RD and Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98(3): 285-294.
204. Kouzarides T (2002) Histone methylation in transcriptional control. *Curr Opin Genet Dev* 12(2): 198-209.

-
205. Krasnianski M, Eger K, Neudecker S, Jakubiczka S, and Zier S (2003) Atypical phenotypes in patients with facioscapulohumeral muscular dystrophy 4q35 deletion. *Arch Neurol* 60: 1421-1425.
 206. Krasnianski M, Neudecker S, Eger K, Schulte-Mattler W, and Zierz S (2003) Fazioskapulohumerale Muskeldystrophie. *Nervenarzt* 74: 151-158.
 207. Kreimeyer A, Wielckens K, Adamietz P, and Hilz H (1984) DNA repair-associated ADP-ribosylation in vivo. Modification of histone H1 differs from that of the principal acceptor proteins. *J Biol Chem* 259(2): 890-896.
 208. Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, and Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 252(5013): 1711-1714.
 209. Kriaucionis S and Bird A (2003) DNA methylation and Rett syndrome. *Hum Mol Genet* 12 Spec No 2: R221-R227.
 210. Kruczek I and Doerfler W (1983) Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters: effect of promoter methylation on gene expression. *Proc Natl Acad Sci USA* 80(24): 7586-7590.
 211. Kurdistani SK and Grunstein M (2003) Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol* 4(4): 276-284.
 212. Kurdistani SK, Tavazoie S, and Grunstein M (2004) Mapping global histone acetylation patterns to gene expression. *Cell* 117(6): 721-733.
 213. Kuwano A, Ledbetter SA, Dobyns WB, Emanuel BS, and Ledbetter DH (1991) Detection of deletions and cryptic translocations in Miller-Dieker syndrome by in situ hybridization. *Am J Hum Genet* 49(4): 707-714.
 214. Kwan J and Hartman SS (2004) Facioscapulohumeral muscular dystrophy (FSHD). Baylor College of Medicine - Department of neurology, Houston. Internet Communication
 215. Laforet P, de Toma C, Eymard B, Becane HM, Jeanpierre M, Fardeau M, and Duboc D (1998) Cardiac involvement in genetically confirmed facioscapulohumeral muscular dystrophy. *Neurology* 51(5): 1454-1456.
 216. Lahue RS and Slater DL (2003) DNA repair and trinucleotide repeat instability. *Front Biosci* 8: s653-s665.
 217. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921.
 218. Landouzy L (1874) Note sur deux cas d'atrophie musculaire progressive de l'enfance. *Memoires de la Société de Biologie*: 103-121.
 219. Landouzy L and Dejerine J (1884) De la myopathie atrophique progressive: myopathie héréditaire débutant, dans l'enfance, par la face, sans altération du système nerveux. *C R Acad Sci (Paris)* 98: 53-55.
 220. Landouzy L and Dejerine J (1885) De la myopathie atrophique progressive (myopathie sans neuropathie, débutant d'ordinaire dans l'enfance, par la face). *Revue de Médecine* 5: 81-117.
 221. Landouzy L and Dejerine J (1885) De la myopathie atrophique progressive (myopathie sans neuropathie, débutant d'ordinaire dans l'enfance, par la face). *Revue de Médecine* 5: 253-366.
 222. Laoudj-Chenivesse D, Carnac G, Bisbal C, Hugon G, Bouillot S, Desnuelle C, Vassetzky Y, and Fernandez A (2004) Increased levels of adenine nucleotide translocator 1 protein and response to oxidative stress are early events in facioscapulohumeral muscular dystrophy muscle. *J Mol Med*.
 223. Larsen F, Gundersen G, Lopez R, and Prydz H (1992) CpG islands as gene markers in the human genome. *Genomics* 13(4): 1095-1107.
 224. Lazaro C, Ravella A, Gaona A, Volpini V, and Estivill X (1994) Neurofibromatosis type 1 due to germ-line mosaicism in a clinically normal father. *N Engl J Med* 331(21): 1403-1407.
-

225. Lee JH, Goto K, Matsuda C, and Arahata K (1995) Characterization of a tandemly repeated 3.3 kb *KpnI* unit in the facioscapulohumeral muscular dystrophy (FSHD) gene region on chromosome 4q35. *Muscle Nerve* 2: S6-S13.
226. 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.
227. Lemmers RJLF, Kievit de P, Geel van M, Wielen van der MJ, Bakker E, Padberg GW, Frants RR, and Maarel van der SM (2001) Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. *Ann Neurol* 50(6): 816-819.
228. Lemmers RJLF, Kievit de P, Sandkuijl L, 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.
229. 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, *et al.* (1998) Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet* 7(8): 1207-1214.
230. Lemmers RJLF, Osborn M, Haaf T, Rogers M, Frants RR, Padberg GW, Cooper DN, Maarel van der SM, and Upadhyaya M (2003) D4F104S1 deletion in facioscapulohumeral muscular dystrophy: Phenotype, size, and detection. *Neurology* 61(2): 178-183.
231. Lemmers RJLF, Wielen van der M, Bakker E, and Maarel van der SM (2004) Molecular diagnosis of FSHD. Chapter 15: 211-234. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
232. Lemmers RJLF, Wielen van der MJ, Bakker E, Padberg GW, Frants RR, and Maarel van der SM (2004) Somatic mosaicism in FSHD often goes undetected. *Ann Neurol* 55(6): 845-850.
233. Lewin B (2000) Clusters and repeats. Chapter 4: 89-115. In: *Genes VII*. Oxford University Press Inc., New York.
234. Lewin B (2000) How many genes are there? Chapter 3: 67-87. In: *Genes VII*. Oxford University Press Inc., New York.
235. Lewin B (2000) Retroviruses and retroposons. Chapter 16: 485-505. In: *Genes VII*. Oxford University Press Inc., New York.
236. Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3(9): 662-673.
237. Li K, Warner CK, Hodge JA, Minoshima S, Kudoh J, Fukuyama R, Maekawa M, Shimizu Y, Shimizu N, and Wallace DC (1989) A human muscle adenine nucleotide translocator gene has four exons, is located on chromosome 4, and is differentially expressed. *J Biol Chem* 264(24): 13998-14004.
238. Li LC and Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18(11): 1427-1431.
239. Li SH and Li XJ (2004) Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends Genet* 20(3): 146-154.
240. Liao D (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 64(1): 24-30.
241. 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.
242. 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.

243. Linardopoulou E, Mefford HC, Nguyen O, Friedman C, Engh van der G, Farwell DG, Coltrera M, and Trask BJ (2001) Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum Mol Genet* 10(21): 2373-2383.
244. Litt M and Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44(397): 401.
245. Little M and Seehaus T (1988) Comparative analysis of tubulin sequences. *Comp Biochem Physiol B* 90(4): 655-670.
246. Loyola A and Almouzni G (2004) Histone chaperones, a supporting role in the limelight. *Biochim Biophys Acta* 1677(1-3): 3-11.
247. Luduena RF (1998) Multiple forms of tubulin: different gene products and covalent modifications. *Int Rev Cytol* 178: 207-275.
248. Lundgren M, Chow CM, Sabbattini P, Georgiou A, Minaee S, and Dillon N (2000) Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* 103(5): 733-743.
249. Lunt PW (1998) 44th ENMC International Workshop: Facioscapulohumeral Muscular Dystrophy: Molecular Studies 19-21 July 1996, Naarden, The Netherlands. *Neuromusc Disord* 8(2): 126-130.
250. 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.
251. Lunt PW and Harper PS (1991) Genetic counselling in facioscapulohumeral muscular dystrophy. *J Med Genet* 28(10): 655-664.
252. Lunt PW, Jardine PE, Koch M, Maynard J, Osborn M, Williams M, Harper PS, and Upadhyaya M (1995) Phenotypic-genotypic correlation will assist genetic counseling in 4q35-facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2: S103-S109.
253. 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.
254. Lunt PW, Upadhyaya M, and Koch M (2004) Genotype-phenotype relationships in FSHD. Chapter 11: 151-168. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
255. Lunt P (2000) Facioscapulohumeral muscular dystrophy: diagnostic and molecular aspects: 44-60. In: *Neuromuscular diseases: from basic mechanisms to clinical management*. Vol. 18. Deymeer F (Ed.) Karger, Basel.
256. Lyle R, Wright TJ, Clark LN, and Hewitt JE (1995) The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* 28(3): 389-397.
257. 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.
258. Margot JB, Ehrenhofer-Murray AE, and Leonhardt H (2003) Interactions within the mammalian DNA methyltransferase family. *BMC Mol Biol* 4(1): 7.
259. Martin R (1992) Classification and evolutionary relationships. Chapter 1.2: 17-23. In: *Human evolution*. Jones S, Martin R, Pilbeam D (Eds.) Cambridge University Press, Cambridge.

260. Masny PS, Bengtsson U, Chung SA, Martin JH, Engelen van B, Maarel van der SM, and Winokur ST (2004) Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum Mol Genet* 13(17): 1857-1871.
261. Mathews KD, Mills KA, Bosch EP, Ionasescu VV, Wiles KR, Buetow KH, and Murray JC (1992) Linkage localization of facioscapulohumeral muscular dystrophy (FSHD) in 4q35. *Am J Hum Genet* 51(2): 428-431.
262. 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).
263. Matzke MA, Mette MF, Aufsatz W, Jakowitsch J, and Matzke AJ (1999) Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. *Genetica* 107(1-3): 271-287.
264. Meehan R, Lewis J, Cross S, Nan X, Jeppesen P, and Bird A (1992) Transcriptional repression by methylation of CpG. *J Cell Sci Supplement* 16: 9-14.
265. Mefford HC, Linardopoulou E, Coil D, Engh van den G, and Trask BJ (2001) Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Hum Mol Genet* 10(21): 2363-2372.
266. Mefford HC and Trask BJ (2002) The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet* 3(2): 91-102.
267. Melese T and Xue Z (1995) The nucleolus: an organelle formed by the act of building a ribosome. *Curr Opin Cell Biol* 7(3): 319-324.
268. Meneveri R, Agresti A, Marozzi A, Saccone S, Rocchi M, Archidiacono N, Corneo G, Della VG, and Ginelli E (1993) Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. *Gene* 123(2): 227-234.
269. Milner EC, Lotshaw CL, Willems-van Dijk K, Charmley P, Concannon P, and Schroeder HW, Jr. (1989) Isolation and mapping of a polymorphic DNA sequence pH30 on chromosome 4. *Nucleic Acids Res* 17(10): 4002.
270. Miniou P, Jeanpierre M, Bourc'h D, Coutinho Barbosa AC, Blanquet V, and Viegas-Pequignot E (1997) alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. *Hum Genet* 99(6): 738-745.
271. Miura K, Kumagai T, Matsumoto A, Iriyama E, Watanabe K, Goto K, and Arahata K (1998) Two cases of chromosome 4q35-linked early onset facioscapulohumeral muscular dystrophy with mental retardation and epilepsy. *Neuropediatrics* 29(5): 239-241.
272. Mizuno Y, Thompson TG, Guyon JR, Lidov HG, Brosius M, Imamura M, Ozawa E, Watkins SC, and Kunkel LM (2001) Desmuslin, an intermediate filament protein that interacts with alpha -dystrobrevin and desmin. *Proc Natl Acad Sci USA* 98(11): 6156-6161.
273. Moen PT, Jr., Johnson CV, Byron M, Shopland LS, Serna de la I, Imbalzano AN, and Lawrence JB (2004) Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol Biol Cell* 15(1): 197-206.
274. Monk M, Boubelik M, and Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99(3): 371-382.
275. Moore SC, Jason LJ, and Ausio J (2002) The elusive structural role of ubiquitinated histones. *Biochem Cell Biol* 80: 311-319.
276. Morin V, Diaz F, Montecino M, Fothergill-Gilmore L, Puchi M, and Imschenetzky M (1999) Poly(ADP-ribosylation) protects maternally derived histones from proteolysis after fertilization. *Biochem J* 343 Pt 1: 95-98.
277. Mostoslavsky R and Bergman Y (1997) DNA methylation: regulation of gene expression and role in the immune system. *Biochim Biophys Acta* 1333(1): F29-F50.
278. Müller S, Ledl A, and Schmidt D (2004) SUMO: a regulator of gene expression and genome integrity. *Oncogene* 23(11): 1998-2008.

-
279. Murgia A, Martella M, Vinanzi C, Polli R, Perilongo G, and Opocher G (1999) Somatic mosaicism in von Hippel-Lindau disease. *Hum Mutat Online*: 114.
 280. Nakagawa M, Matsuzaki T, Higuchi I, Fukunaga H, Inui T, Nagamitsu S, Yamada H, Arimura K, and Osame M (1997) Facioscapulohumeral muscular dystrophy: clinical diversity and genetic abnormalities in Japanese patients. *Intern Med* 36(5): 333-339.
 281. Nakagawa M, Matsuzaki T, Takashima H, Suehara M, Higuchi I, Inui T, Nagamitsu S, Hokezu Y, Uchino M, and Osame M (1998) Facioscapulohumeral muscular dystrophy (FSHD): clinical diversity and genetic abnormalities in Japanese patients. *Muscle Nerve Supplement* 7: S111.
 282. Nakao M (2001) Epigenetics: interaction of DNA methylation and chromatin. *Gene* 278(1-2): 25-31.
 283. Nakayama T, Komiya T, Tyou E, Watanabe S, and Kawai M (1999) Cardiac deformity and dysfunction in facioscapulohumeral dystrophy-electrocardiogram, ECG gate cardiac MRI studies. *Rinsho Shinkeigaku* 39(6): 610-614.
 284. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, and Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393(6683): 386-389.
 285. Napoli L, Bordoni A, Zeviani M, Hadjigeorgiou GM, Sciacco M, Tiranti V, Terentiu A, Moggio M, Papadimitriou A, Scarlato G, *et al.* (2001) A novel missense adenine nucleotide translocator-1 gene mutation in a Greek adPEO family. *Neurology* 57(12): 2295-2298.
 286. Narayanan A, Speckmann W, Terns R, and Terns MP (1999) Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell* 10(7): 2131-2147.
 287. Nathan D, Sterner DE, and Berger SL (2003) Histone modifications: Now summoning sumoylation. *Proc Natl Acad Sci USA* 100(23): 13118-13120.
 288. Nativio DG (2002) The genetics, diagnosis, and management of Prader-Willi syndrome. *J Pediatr Health Care* 16: 298-303.
 289. Nei M, Gu X, and Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94(15): 7799-7806.
 290. Ng HH and Bird A (1999) DNA methylation and chromatin modification. *Curr Opin Genet Dev* 9(2): 158-163.
 291. Ordway JM and Curran T (2002) Methylation matters: modeling a manageable genome. *Cell Growth Differ* 13(4): 149-162.
 292. Orrell RW, Tawil R, Forrester J, Kissel JT, Mendell JR, and Figlewicz DA (1999) Definitive molecular diagnosis of facioscapulohumeral dystrophy. *Neurology* 52(9): 1822-1826.
 293. Osley MA (2004) H2B ubiquitylation: the end is in sight. *Biochim Biophys Acta* 1677(1-3): 74-78.
 294. Ostlund C and Worman HJ (2003) Nuclear envelope proteins and neuromuscular diseases. *Muscle Nerve* 27(4): 393-406.
 295. 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.
 296. Overveld van PGM, Lemmers RJLF, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, Ommen van GB, Frants RR, and Maarel van der SM (2003) Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* 35(4): 315-317.
 297. Ozcelik T (2002) Uncovering the complex mysteries of mosaicism. *Nature* 417(6889): 588.
 298. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
-

299. Padberg GW (1998) Fascioscapulohumeral muscular dystrophy. Chapter 5: 105-121. In: Neuromuscular disorders: clinical and molecular genetics. Emery AEH (Ed.) John Wiley & Sons Ltd, Chichester.
300. Padberg GW (2004). Personal Communication.
301. Padberg GW (2004) Facioscapulohumeral muscular dystrophy: a clinician's experience. Chapter 3: 41-53. In: Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
302. Padberg GW, Brouwer OF, Keizer de RJ, Dijkman G, Wijmenga C, Grote JJ, and Frants RR (1995) On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2: S73-S80.
303. Padberg GW, Brouwer OF, Keizer de RJW, Gruter AM, Wijmenga C, Grote JJ, and Frants RR (1992) Retinal vascular disease and sensorineural deafness are part of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 51(4; Supplement): A104.
304. Padberg GW, Eriksson AW, Volkers WS, Bernini L, Loghem van E, Meera-Khan P., Nijenhuis LE, Pronk JC, and Schreuder GM (1984) Linkage studies in autosomal dominant facioscapulohumeral muscular dystrophy. *J Neurol Sci* 65(3): 261-268.
305. 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.
306. 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.
307. Padberg GW, Vogels OJ, and van der Kooi EL (1998) The clinical picture of FSHD. *Muscle Nerve Supplement* 7: S25.
308. Padberg GW and Wohlgemuth M (2003) Facioscapulohumeral muscular dystrophy (FSHD). University Medical Center St. Radboud, Nijmegen, The Netherlands. Internet Communication
309. Park Y and Kuroda MI (2001) Epigenetic aspects of X-chromosome dosage compensation. *Science* 293(5532): 1083-1085.
310. Parniewski P and Staczek P (2002) Molecular mechanisms of TRS instability. *Adv Exp Med Biol* 516: 1-25.
311. Pavelitz T, Rusche L, Matera AG, Scharf JM, and Weiner AM (1995) Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. *EMBO J* 14(1): 169-177.
312. Pearson CE, Ewel A, Acharya S, Fishel RA, and Sinden RR (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum Mol Genet* 6(7): 1117-1123.
313. Pearson H (2002) Dual identities. *Nature* 417(6884): 10-11.
314. Pederson T (1998) The plurifunctional nucleolus. *Nucleic Acids Res* 26(17): 3871-3876.
315. Perrod S and Gasser SM (2003) Long-range silencing and position effects at telomeres and centromeres: parallels and differences. *Cell Mol Life Sci* 60(11): 2303-2318.
316. Pham AD and Sauer F (2000) Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. *Science* 289(5488): 2357-2360.
317. Piccini I, Ballarati L, Bassi C, Rocchi M, Marozzi A, Ginelli E, and Meneveri R (2001) The structure of duplications on human acrocentric chromosome short arms derived by the analysis of 15p. *Hum Genet* 108(6): 467-477.
318. Pirrotta V (1999) Transvection and chromosomal trans-interaction effects. *Biochim Biophys Acta* 1424(1): 1-8.
319. Plass C (2002) Cancer epigenomics. *Hum Mol Genet* 11(20): 2479-2488.
320. Purandare SM and Patel PI (1997) Recombination hot spots and human disease. *Genome Res* 7(8): 773-786.
321. Rakyant VK, Preis J, Morgan HD, and Whitelaw E (2001) The marks, mechanisms and memory of epigenetic states in mammals. *Biochem J* 356(Pt 1): 1-10.

-
322. Randolph-Anderson BL, Stajich JM, Graham FL, Pericak-Vance MA, Speer MC, and Gilbert JR (2002) Evidence consistent with linkage to 15q of a non-chromosome 4 linked FSHD family. *Am J Hum Genet* 71(4; Supplement): 530.
 323. Ranum LP and Day JW (2004) Pathogenic RNA repeats: an expanding role in genetic disease. *Trends Genet* 20(10): 506-512.
 324. Rasmussen TP (2003) Embryonic stem cell differentiation: a chromatin perspective. *Reprod Biol Endocrinol* 1(1): 100.
 325. Razin A and Cedar H (1991) DNA methylation and gene expression. *Microbiol Rev* 55(3): 451-458.
 326. Razin A and Shemer R (1995) DNA methylation in early development. *Hum Mol Genet* 4 Spec No: 1751-1755.
 327. Reik W, Dean W, and Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293(5532): 1089-1093.
 328. Reik W and Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2(1): 21-32.
 329. Réveillère C, Diaz C, and Uritzbera JA (1995) FSH: Premiers resultats. *Association Française contre les Myopathies (AFM)* (68): 22-24.
 330. Ricci E, Galluzzi G, Deidda G, Cacurri S, Colantoni L, Merico B, Piazza N, Servidei S, Vigneti E, Pasceri V, Silvestri G, *et al.* (1999) Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of *KpnI* repeats at the 4q35 locus and clinical phenotype. *Ann Neurol* 45(6): 751-757.
 331. Richards RI (2001) Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum Mol Genet* 10(20): 2187-2194.
 332. Richards RI and Sutherland GR (1994) Simple repeat DNA is not replicated simply. *Nat Genet* 6(2): 114-116.
 333. Richardson B (2003) Impact of aging on DNA methylation. *Ageing Res Rev* 2(3): 245-261.
 334. Rijkers T, Deidda G, Koningsbruggen van S, Geel van M, Lemmers RJLF, Deutekom van JC, Figlewicz D, Hewitt JE, Padberg GW, Frants RR, *et al.* (2004) *FRG2*, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J Med Genet* 41(11): 826-836.
 335. Rijkers T, Deidda G, Koningsbruggen van S, Lemmers RJLF, Deutekom van JC, Figlewicz D, Hewitt JE, Padberg GW, Frants RR, and Maarel van der SM (2002) A novel gene in the FSHD region, *FRG2*, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *Neuromusc Disord* 12: 741.
 336. Robertson KD (2002) DNA methylation and chromatin - unraveling the tangled web. *Oncogene* 21(35): 5361-5379.
 337. Robertson KD and Ambinder RF (1997) Methylation of the Epstein-Barr virus genome in normal lymphocytes. *Blood* 90(11): 4480-4484.
 338. Robertson KD and Wolffe AP (2000) DNA methylation in health and disease. *Nat Rev Genet* 1(1): 11-19.
 339. Rogers MT (2004) Facioscapulohumeral muscular dystrophy: historical background and literature review. Chapter 2: 17-39. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
 340. Rogers MT, Sewry CA, and Upadhyaya M (2004) Histological, immunocytochemical, molecular and ultrastructural characteristics of FSHD muscle. Chapter 18: 277-298. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
 341. Rogers MT, Zhao F, Harper PS, and Stephens D (2002) Absence of hearing impairment in adult onset facioscapulohumeral muscular dystrophy. *Neuromusc Disord* 12(4): 358-365.
 342. Roques I, Pedespan JM, Boisserie-Lacroix V, Ferrer X, and Fontan D (1998) Facioscapulohumeral myopathy and germinal mosaicism. *Arch Pediatr* 5(8): 880-883.
-

- 343. Rouyer F, de la CA, Andersson M, and Weissenbach J (1990) An interspersed repeated sequence specific for human subtelomeric regions. *EMBO J* 9(2): 505-514.
- 344. Royle NJ (1995) The proterminal regions and telomeres of human chromosomes. *Adv Genet* 32: 273-315.
- 345. Russo VEA, Martienssen RA, and Riggs AD (1996) Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- 346. Saito A, Higuchi I, Nakagawa M, Saito M, Uchida Y, Inose M, Kasai T, Niiyama T, Fukunaga H, Arimura K, *et al.* (2000) An overexpression of fibroblast growth factor (FGF) and FGF receptor 4 in a severe clinical phenotype of facioscapulohumeral muscular dystrophy. *Muscle Nerve* 23(4): 490-497.
- 347. Sarfarazi M, Upadhyaya M, Padberg G, Pericak-Vance M, Siddique T, Lucotte G, and Lunt P (1989) An exclusion map for facioscapulohumeral (Landouzy-Dejerine) disease. *J Med Genet* 26(8): 481-484.
- 348. Sarfarazi M, Wijmenga C, Upadhyaya M, Weiffenbach B, Hyser C, Mathews K, Murray J, Gilbert J, Pericak Vance M, Lunt PW, *et al.* (1992) Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: combined analysis of an international consortium. *Am J Hum Genet* 51(2): 396-403.
- 349. Sarma K and Reinberg D (2005) Histone variants meet their match. *Nat Rev Mol Cell Biol* 6(2): 139-149.
- 350. Scheer U and Hock R (1999) Structure and function of the nucleolus. *Curr Opin Cell Biol* 11(3): 385-390.
- 351. Schiebel K, Weiss B, Wohrle D, and Rappold G (1993) A human pseudoautosomal gene, ADP/ATP translocase, escapes X-inactivation whereas a homologue on Xq is subject to X-inactivation. *Nat Genet* 3(1): 82-87.
- 352. Schubeler D, Lorincz MC, Cimborá DM, Telling A, Feng YQ, Bouhassira EE, and Groudine M (2000) Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Mol Cell Biol* 20(24): 9103-9112.
- 353. Schuffenhauer S, Bartsch O, Stumm M, Buchholz T, Petropoulou T, Kraft S, Belohradsky B, Hinkel GK, Meitinger T, and Wegner RD (1995) DNA, FISH and complementation studies in ICF syndrome: DNA hypomethylation of repetitive and single copy loci and evidence for a trans acting factor. *Hum Genet* 96(5): 562-571.
- 354. Shaffer LG and Lupski JR (2000) Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu Rev Genet* 34: 297-329.
- 355. Sharma S and Namrata S (2004) Association of schizophrenia and mental retardation with facioscapulohumeral muscular dystrophy. *Neurol India* 52(1): 104-105.
- 356. Shigeto H, Tamura T, Oya Y, Ogawa M, and Kawai M (2002) Facioscapulohumeral muscular dystrophy with sinus dysfunction. *Rinsho Shinkeigaku* 42(9): 881-884.
- 357. Shiio Y and Eisenman RN (2003) Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci USA* 100(23): 13225-13230.
- 358. Siegfried Z and Cedar H (1997) DNA methylation: a molecular lock. *Curr Biol* 7(5): R305-R307.
- 359. Singal R and Ginder GD (1999) DNA methylation. *Blood* 93(12): 4059-4070.
- 360. Sippel KC, Fraioli RE, Smith GD, Schalkoff ME, Sutherland J, Gallie BL, and Dryja TP (1998) Frequency of somatic and germ-line mosaicism in retinoblastoma: implications for genetic counseling. *Am J Hum Genet* 62(3): 610-619.
- 361. Sleeman J, Lyon CE, Platani M, Kreivi JP, and Lamond AI (1998) Dynamic interactions between splicing snRNPs, coiled bodies and nucleoli revealed using snRNP protein fusions to the green fluorescent protein. *Exp Cell Res* 243(2): 290-304.
- 362. Sleeman JE and Lamond AI (1999) Nuclear organization of pre-mRNA splicing factors. *Curr Opin Cell Biol* 11(3): 372-377.
- 363. Slim R, Levilliers J, Ludecke HJ, Claussen U, Nguyen VC, Gough NM, Horsthemke B, and Petit C (1993) A human pseudoautosomal gene encodes the ANT3 ADP/ATP translocase and escapes X-inactivation. *Genomics* 16(1): 26-33.

-
364. Speer MC, Pericak-Vance MA, Stajich JM, Sarrica J, Jordan M, Roses AD, Vance JM, and Gilbert JR (1997) Further exclusion of FSHD1B from the telomeric region of 10q. *Neurogenetics* 1(2): 151-152.
 365. Stankiewicz P and Lupski JR (2002) Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* 12(3): 312-319.
 366. Stevenson WG, Perloff JK, Weiss JN, and Anderson TL (1990) Facioscapulohumeral muscular dystrophy: evidence for selective, genetic electrophysiologic cardiac involvement. *J Am Coll Cardiol* 15(2): 292-299.
 367. Stout K, Maarel van der SM, Frants RR, Padberg GW, Ropers H-H, and Haaf T (1999) Somatic pairing between subtelomeric regions: implications for human genetic disease? *Chromosome Res* 7: 323-329.
 368. Strachan T and Read AP (2001) Organization of the human genome. Chapter 7: 139-168. In: *Human molecular genetics 2*. Kingston F (Ed.) BIOS Scientific Publishers Limited, Oxford.
 369. Strahl BD and Allis CD (2000) The language of covalent histone modifications. *Nature* 403(6765): 41-45.
 370. 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.
 371. Sullivan KF (2001) A solid foundation: functional specialization of centromeric chromatin. *Curr Opin Gene Dev* 11(2): 182-188.
 372. Sun FL, Cuaycong MH, and Elgin SC (2001) Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin. *Mol Cell Biol* 21(8): 2867-2879.
 373. Sutherland GR and Richards RI (1995) Simple tandem DNA repeats and human genetic disease. *Proc Natl Acad Sci USA* 92(9): 3636-3641.
 374. Szyf M (2003) Targeting DNA methylation in cancer. *Ageing Res Rev* 2(3): 299-328.
 375. Talasz H, Wasserer S, and Puschendorf B (2002) Nonenzymatic glycation of histones in vitro and in vivo. *J Cell Biochem* 85(1): 24-34.
 376. Tam R, Smith KP, and Lawrence JB (2004) The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J Cell Biol* 167(2): 269-279.
 377. Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, and Figlewicz D (1996) Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39(6): 744-748.
 378. Thain M and Hickman M (2001) *Biology*. Penguin Group, London.
 379. Tim RW, Gilbert JR, Stajich JM, Rampersaud E, Viles KD, Tawil R, Padberg GW, Frants RR, Maarel van der SM, Bossen EH, *et al.* (2001) Clinical studies in non-chromosome 4-linked facioscapulohumeral muscular dystrophy. *J Clin Neuromusc Dis* 3(1): 1-7.
 380. 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.
 381. Tonini MM and Zatz M (2004) Facioscapulohumeral muscular dystrophy: gender differences and genetic counseling in a complex disorder. Chapter 20: 309-324. In: *Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology*. Upadhyaya M, Cooper DN (Eds.) Garland Science/BIOS Scientific Publishers Limited, Abingdon.
 382. Topcu M, Akyerli C, Sayi A, Toruner GA, Kocoglu SR, Cimbis M, and Ozcelik T (2002) Somatic mosaicism for a MECP2 mutation associated with classic Rett syndrome in a boy. *Eur J Hum Genet* 10(1): 77-81.
 383. Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, Blankenship J, Collins C, Giorgi D, Iadonato S, Johnson F, Kuo *et al.* (1998) Members of the olfactory receptor
-

- gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* 7(1): 13-26.
384. Trievel RC (2004) Structure and function of histone methyltransferases. *Crit Rev Eukaryot Gene Expr* 14(3): 147-169.
385. 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.
386. Tupler R, Berardinelli A, Barbierato L, Frants R, 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.
387. Tupler R, Perini G, Pellegrino MA, and Green MR (1999) Profound misregulation of muscle-specific gene expression in facioscapulohumeral muscular dystrophy. *Proc Natl Acad Sci USA* 96(22): 12650-12654.
388. Tyler JK and Kadonaga JT (1999) The "dark side" of chromatin remodeling: repressive effects on transcription. *Cell* 99(5): 443-446.
389. Tzankov B, Tournev I, and Jeanpierre M (2002) FSHD in the Bulgarian population - clinical and genetic heterogeneity. *Neuromusc Disord* 12: 742.
390. Uncini A, Galluzzi G, Di Muzio A, De Angelis MV, Ricci E, Scoppetta C, and Servidei S (2002) Facioscapulohumeral muscular dystrophy presenting isolated monomelic lower limb atrophy. Report of two patients with and without 4q35 rearrangement. *Neuromusc Disord* 12(9): 874-877.
391. Upadhyaya M and Cooper DN (2002) Molecular diagnosis of facioscapulohumeral muscular dystrophy. *Expert Rev Mol Diagn* 2(2): 160-171.
392. Upadhyaya M, Jardine P, Maynard J, Farnham J, Sarfarazi M, Wijmenga C, Hewitt JE, Frants R, Harper PS, and Lunt PW (1993) Molecular analysis of British facioscapulohumeral dystrophy families for 4q DNA rearrangements. *Hum Mol Genet* 2(7): 981-987.
393. Upadhyaya M, Maynard J, Osborn M, Jardine P, Harper PS, and Lunt P (1995) Germinal mosaicism in facioscapulohumeral muscular dystrophy (FSHD). *Muscle Nerve* 2: S45-S49.
394. Upadhyaya M, Maynard J, Rogers MT, Lunt PW, Jardine P, Ravine D, and Harper PS (1997) Improved molecular diagnosis of facioscapulohumeral muscular dystrophy (FSHD): validation of the differential double digestion for FSHD. *J Med Genet* 34(6): 476-479.
395. Uritzbera JA, Reveillere C, Diaz C, Cure S, and Riviere H (1998) Clinical and psychosocial aspects of facioscapulohumeral muscular dystrophy from the patient's perspective: a survey in 270 French FSHD patients. *Muscle Nerve Supplement* 7: S112.
396. Urnov FD (2002) Methylation and the genome: the power of a small amendment. *J Nutr* 132(8 Supplement): 2450S-2456S.
397. Vandebrouck C, Imbert N, Constantin B, Duport G, Raymond G, and Cognard C (2002) Normal calcium homeostasis in dystrophin-expressing facioscapulohumeral muscular dystrophy myotubes. *Neuromusc Disord* 12(3): 266-272.
398. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.* (2001) The sequence of the human genome. *Science* 291: 1304-1351.
399. Verger A, Perdomo J, and Crossley M (2003) Modification with SUMO. A role in transcriptional regulation. *EMBO Rep* 4(2): 137-142.
400. Verheggen C, Lafontaine DL, Samarsky D, Mouaikel J, Blanchard JM, Bordonne R, and Bertrand E (2002) Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J* 21(11): 2736-2745.
401. Vermaak D, Ahmad K, and Henikoff S (2003) Maintenance of chromatin states: an open-and-shut case. *Curr Opin Cell Biol* 15(3): 266-274.
402. Voit T, Lamprecht A, Lenard HG, and Goebel HH (1986) Hearing loss in facioscapulohumeral dystrophy. *Eur J Pediatr* 145(4): 280-285.

-
403. Vu TH and Hoffman AR (2000) Comparative genomics sheds light on mechanisms of genomic imprinting. *Genome Res* 10(11): 1660-1663.
 404. Wallace BM and Hulten MA (1985) Meiotic chromosome pairing in the normal human female. *Ann Hum Genet* 49(Pt 3): 215-226.
 405. Wallrath LL (1998) Unfolding the mysteries of heterochromatin. *Curr Opin Genet Dev* 8(2): 147-153.
 406. Walsh CP and Bestor TH (1999) Cytosine methylation and mammalian development. *Genes Dev* 13(1): 26-34.
 407. 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.
 408. Warburton P, Mohammed S, and Ogilvie CM (2000) Detection of submicroscopic subtelomeric chromosome translocations: A new case study. *Am J Med Genet* 91: 51-55.
 409. Warren ST (1996) The expanding world of trinucleotide repeats. *Science* 271(5254): 1374-1375.
 410. Weber JL, Kwitek AE, May PE, and Killary AM (1990) Dinucleotide repeat polymorphism at the D4S174 locus. *Nucleic Acids Res* 18(15): 4636.
 411. Weber JL and May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44(3): 388-396.
 412. Weiffenbach B, Bagley R, Falls K, Hyser C, Storvick D, Jacobsen SJ, Schultz P, Mendell J, Willems van Dijk K, Milner EC, *et al.* (1992) Linkage analyses of five chromosome 4 markers localizes the facioscapulohumeral muscular dystrophy (FSHD) gene to distal 4q35. *Am J Hum Genet* 51(2): 416-423.
 413. Weiffenbach B, Dubois J, Storvick D, Tawil R, Jacobsen SJ, Gilbert J, Wijmenga C, Mendell JR, Winokur S, Altherr MR, *et al.* (1993) Mapping the facioscapulohumeral muscular dystrophy gene is complicated by chromosome 4q35 recombination events. *Nat Genet* 4(2): 165-169.
 414. Weiler KS and Wakimoto BT (1995) Heterochromatin and gene expression in *Drosophila*. *Annu Rev Genet* 29: 577-605.
 415. Wijmenga C, Brouwer OF, Padberg GW, and Frants RR (1992) Transmission of de-novo mutation associated with facioscapulohumeral muscular dystrophy. *Lancet* 340(8825): 985-986.
 416. Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, and Padberg GW (1990) Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* 336(8716): 651-653.
 417. 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.
 418. Wijmenga C, Heuvel van den LP, Strengman E, Luyten JA, Burgt van der I, Groot de R, Smeets DF, Draaisma JM, Dongen van JJ, Abreu de RA, *et al.* (1998) Localization of the ICF syndrome to chromosome 20 by homozygosity mapping. *Am J Hum Genet* 63(3): 803-809.
 419. 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.
 420. Wijmenga C, Padberg GW, Moerer P, Wiegant J, Liem L, Brouwer OF, Milner EC, Weber JL, Ommen van GB, Sandkuyl LA, and *et al.* (1991) Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and *in situ* hybridization. *Genomics* 9(4): 570-575.
 421. Wijmenga C, Winokur ST, Padberg GW, Skraastad MI, Altherr MR, Wasmuth JJ, Murray JC, Hofker MH, and Frants RR (1993) The human skeletal muscle adenine nucleotide translocator gene maps to chromosome 4q35 in the region of the facioscapulohumeral muscular dystrophy locus. *Hum Genet* 92(2): 198-203.
-

- 422. Wijmenga C, Wright TJ, Baan MJ, Padberg GW, Williamson R, Ommen van GJ, Hewitt JE, Hofker MH, and Frants RR (1993) Physical mapping and YAC-cloning connects four genetically distinct 4qter loci (D4S163, D4S139, D4F35S1 and D4F104S1) in the FSHD gene-region. *Hum Mol Genet* 2(10): 1667-1672.
- 423. Wilkie AO, Higgs DR, Rack KA, Buckle VJ, Spurr NK, Fischel-Ghodsian N, Ceccherini I, Brown WR, and Harris PC (1991) Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* 64(3): 595-606.
- 424. Wilson KL (2000) The nuclear envelope, muscular dystrophy and gene expression. *Trends Cell Biol* 10(4): 125-129.
- 425. Winokur ST, Barrett K, Martin JH, Forrester JR, Simon M, Tawil R, Chung SA, Masny PS, and Figlewicz DA (2003) Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromusc Disord* 13(4): 322-333.
- 426. Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, *et al.* (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosome Res* 2(3): 225-234.
- 427. Winokur ST, Bengtsson U, Vargas JC, Wasmuth JJ, and Altherr MR (1996) The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum Mol Genet* 5(10): 1567-1575.
- 428. Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, Tapscott SJ, Maarel van der SM, Hayashi Y, and Flanigan KM (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet* 12(22): 2895-2907.
- 429. Wohlgemuth M, Kooi van der EL, Kesteren van RG, Maarel van der SM, and Padberg GW (2004) Ventilatory support in facioscapulohumeral muscular dystrophy. *Neurology* 63(1): 176-178.
- 430. Wohlgemuth M, Kooi van der EL, Vogels OJ, Maarel van der SM, Kesteren van RG, and Padberg GW (2002) Respiratory insufficiency in facioscapulohumeral muscular dystrophy. *Neuromusc Disord* 12: 742.
- 431. Wong M, Kanai Y, Miwa M, Bustin M, and Smulson M (1983) Immunological evidence for the in vivo occurrence of a crosslinked complex of poly(ADP-ribosylated) histone H1. *Proc Natl Acad Sci USA* 80(1): 205-209.
- 432. Woodage T, Basrai MA, Baxevanis AD, Hieter P, and Collins FS (1997) Characterization of the CHD family of proteins. *Proc Natl Acad Sci USA* 94(21): 11472-11477.
- 433. Wright TJ, Wijmenga C, Clark LN, Frants RR, Williamson R, and Hewitt JE (1993) Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. *Hum Mol Genet* 2(10): 1673-1678.
- 434. Wu J and Grunstein M (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem Sci* 25: 619-623.
- 435. 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.
- 436. Xia H, Winokur ST, Kuo WL, Altherr MR, and Bredt DS (1997) Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs. *J Cell Biol* 139(2): 507-515.
- 437. Yamanaka G, Goto K, Hayashi YK, Miyajima T, Hoshika A, and Arahata K (2002) Clinical and genetical features of Japanese early-onset facioscapulohumeral muscular dystrophy. *No To Hattatsu* 34(4): 318-324.
- 438. Yamanaka G, Goto K, Matsumura T, Funakoshi M, Komori T, Hayashi YK, and Arahata K (2001) Tongue atrophy in facioscapulohumeral muscular dystrophy. *Neurology* 57(4): 733-735.

- 439. Yang F, Shao C, Vedanarayanan V, and Ehrlich M (2004) Cytogenetic and immuno-FISH analysis of the 4q subtelomeric region, which is associated with facioscapulohumeral muscular dystrophy. *Chromosoma* 112(7): 350-359.
- 440. Yip DJ and Picketts DJ (2003) Increasing D4Z4 repeat copy number compromises C2C12 myoblast differentiation. *FEBS Lett* 537(1-3): 133-138.
- 441. Yoder JA, Walsh CP, and Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13(8): 335-340.
- 442. Youssoufian H and Pyeritz RE (2002) Mechanisms and consequences of somatic mosaicism in humans. *Nat Rev Genet* 3(10): 748-758.
- 443. Zakian VA (1989) Structure and function of telomeres. *Annu Rev Genet* 23: 579-604.
- 444. 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.
- 445. Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campiotto S, Cerqueira A, Wijmenga C, Padberg G, and Frants R (1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. *Am J Hum Genet* 56(1): 99-105.
- 446. 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.
- 447. Zheng C and Hayes JJ (2003) Structures and interactions of the core histone tail domains. *Biopolymers* 68(4): 539-546.
- 448. Zlotogora J (1998) Germ line mosaicism. *Hum Genet* 102(4): 381-386.

2

Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity

*PGM van Overveld¹, RJFL Lemmers¹, G Deidda^{1,2},
LA Sandkuijl¹, GW Padberg³, RR Frants¹ and SM
van der Maarel¹*

Originally published in:
Human Molecular Genetics
(2000) 9: 2879-2884

- 1 Department of Human Genetics, Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
- 2 Institute of Cell Biology, National Research Council, Rome, Italy.
- 3 Department of Neurology, University Medical Center St. Radboud, Nijmegen, The Netherlands.

2.1 Abstract

Chromosomal rearrangements occur more frequently in subtelomeric domains than in other regions of the genome and are often associated with human pathology. To further elucidate the plasticity of subtelomeric domains, we examined the 3.3 kb D4Z4 repeat array on chromosome 4 and its homologue on chromosome 10 in 208 Dutch blood donors by pulsed-field gel electrophoresis. These subtelomeric repeats are known to rearrange and partial deletions of this polymorphic array on chromosome 4 are associated with facioscapulohumeral muscular dystrophy (FSHD), an autosomal dominant myopathy. Our results show that mitotic rearrangements occur frequently as 3% of individuals display somatic mosaicism for a repeat expansion or contraction explaining the high variability of subtelomeric repeat array sizes. Translocated 4-type repeat arrays on chromosome 10 and the reverse configuration of 10-type repeat arrays on chromosome 4 are observed in 24% of individuals. The translocated repeat arrays on chromosome 4 tend to be more heterogeneous than 4-type repeats on chromosome 10. The repeat length on chromosome 4 is on average larger than on chromosome 10. But on both chromosomes we observe a multi-modal repeat length distribution with equidistant peaks at intervals of 65 kb, possibly reflecting a higher-order chromatin structure. Interestingly, in as many as six random blood donors (3%) we identified FSHD-sized 4-type repeat arrays. Assuming that these individuals are clinically unaffected, these results imply an incomplete penetrance in the upper range of FSHD alleles. Overall, the observed dynamic characteristics of these homologous domains may serve as a model for subtelomeric plasticity.

2.2 Introduction

The chromosomes of many organisms contain subtelomeric regions, which include single copy sequences, highly repetitive sequences and non-homologous chromosome ends [23, 28]. These regions lie immediately adjacent to the dynamic (TTAGGG)_n telomere repeat arrays and display a variable size distribution ranging from a few hundred base pairs (e.g. human XpYpter) [2] to >100 kb (4qter [11] and 16pter [43]). In yeast, chromosome ends are composed of different subtelomeric repeated sequences, such as X and Y' [7, 36, 39], and rearrangements of Y' elements can easily be detected in mitosis and meiosis [20, 22]. Data from other systems also indicate that subtelomeric regions of chromosomes are dynamic structures and may in fact support high levels of meiotic recombination in the initial homology searches, which precede chromosome synapsis [13, 32, 38, 41, 45].

In 1997, a model was proposed in which subtelomeric regions can be divided in a distal and a proximal subdomain separated by degenerated (TTAGGG)_n repeats [15]. The distal subdomain contains repetitive sequences that interact with all chromosome ends, whereas the proximal domain only interacts with a subset of (non-) homologous chromosome ends. It was suggested that subtelomeric regions act as a buffer between functional telomeric repeats and genes proximal of the subtelomeric regions, thereby protecting these genes from heterochromatinisation and silencing, as demonstrated in yeast [4, 6, 18, 21, 40].

Chromosomal rearrangements occur more frequently in (sub)telomeric domains than in other regions of the genome and often result in disease. Translocations of chromosome ends have been reported to cause for example α -thalassaemia mental retardation syndrome, Wolf-Hirschhorn syndrome and Miller-Dieker syndrome [1, 24, 25]. Subtelomeric rearrangements may also account for 7% of the idiopathic mental retardation cases and congenital anomalies [16]. Probably the most intriguing example of disease-associated subtelomeric rearrangements is seen in facioscapulohumeral muscular dystrophy (FSHD), an autosomal dominant myopathy mapped to 4qter. This proximal subtelomeric domain harbours the polymorphic D4Z4 repeat array, which is composed of 11 up to 150 3.3 kb repeated elements in normal individuals. FSHD patients carry a repeat array of <11 units, due to a partial deletion of the D4Z4 repeat array [12, 42].

The subtelomere of chromosome 10 is highly similar to 4q35 and also contains a polymorphic repeat array highly homologous to D4Z4 [3, 8]. A specific *BlnI* site within each repeat unit of chromosome 10 allows discrimination between both chromosomes [9]. In a survey of 50 healthy Dutch male controls we could demonstrate the presence of chromosome 10-type repeats on chromosome 4 in 10% of the Dutch population. The reverse configuration of chromosome 4-type repeats on chromosome 10 is equally frequent [10]. Hybrid repeats, consisting of both 4-type and 10-type repeat units, have also been identified. Several studies demonstrated that only short repeat arrays on chromosome 4 are causally related to FSHD [5, 26].

Recently, we demonstrated that the partial D4Z4 repeat array deletion in FSHD arises mitotically in at least 40% of cases and that the presence of 4-type repeat arrays on chromosome 10 in these mosaic individuals is increased almost five times [30]. We hypothesised that the FSHD deletion occurs mainly by somatic interchromosomal gene-conversion in which the presence of a fully homologous repeat array on a non-homologous chromosome is a predisposing factor.

In this study, we examined the subtelomeric repeat configurations on chromosomes 4 and 10 of 208 unrelated healthy individuals of a Dutch control population. We confirmed the 20% exchanges between 4q35 and 10q26 for both sexes and demonstrated a significant

difference in repeat array length between these two chromosomes. Preferred repeat array lengths were observed possibly reflecting a higher order chromatin structure. Our results indicate a high incidence of mitotic rearrangements as part of the dynamic subtelomere characteristics. This study sheds new light on subtelomeric plasticity and may serve as a model for other subtelomeric loci.

2.3 Subjects and methods

2.3.1 Control samples

Genomic DNA was isolated from peripheral blood lymphocytes from 208 unrelated individuals, which were obtained in a coded form via the Dutch Bloodbank in Leiden after informed consent. DNA was extracted essentially as described by Miller *et al.* [31].

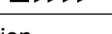
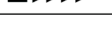
2.3.2 Digestion, PFGE and Southern blotting

Five micrograms of DNA was double digested with restriction enzymes *EcoRI-HindIII* (MBI Fermentas, St Leon-Rot, Germany) or *EcoRI-BlnI* (Amersham, Little Chalfont, UK) according to the manufacturer's instructions. After digestion, DNA was separated by PFGE on a 0.8% agarose gel (MP agarose; Boehringer, Mannheim, Germany). The electrophoresis was performed in 0.5× TBE and was run for 20 h at 8 V/cm². In four identical cycles, switch times increased linearly from 1 to 16 s at the end of each cycle. A pause interval of 2% of the switch time was included. After electrophoresis, the gel was stained with ethidium bromide and blotted to a Nytran⁺ membrane (Schleicher & Schuell, Dassel, Germany). The probes used for hybridisation, p13E-11 (D4F104S1) [42] and 9B6A (D4Z4) [42], were labelled by random priming with [³²P]dCTP, using the megaprime DNA labelling system (Amersham). Hybridisations were performed overnight at 65°C in a buffer containing 7% SDS, 10% PEG 6000 and 50 mM Na₂HPO₄, and the blots were washed at 65°C with 2× SSC/0.1% SDS (p13E-11) or 1× SSC/0.1% SDS (9B6A). The blots were then exposed to phosphorimager screens. After exposure, the alleles were assigned to their respective chromosomes based on their *BlnI* sensitivity. Sizes were estimated independently by two individuals according to a 48 kb marker and λ concatamers.

2.3.3 Classification of individuals

All individuals were analysed for their allele sizes and origin of repeats and classified according to *Table 2.1*. Class A represents a standard repeat array distribution with 4-type arrays on chromosome 4 and 10-type arrays on chromosome 10. Classes B-E represent translocated 10-

Table 2.1 Subtelomeric repeat array constitutions on chromosomes 4q and 10q of 208 unrelated healthy Dutch individuals.

	chr.4	chr.4	chr.10	chr.10	class	n
standard					A	158
10 on 4					B	6
					C	5
					D	7
					E	0
4 on 10					F	18
					G	4
					H	1
					I	2
complex allele constitution					X	1
somatic mosaicism					M	6

The region hybridising to p13E-11 is represented by a square, whereas 4-type and 10-type repeat arrays are indicated by four white and black triangles, respectively. Dots indicate a similar repeat structure as on the top of the column. Individuals are classified in 11 different groups. Class A, the most common allele configuration, is a standard repeat array distribution with 4-type arrays on chromosome 4 and 10-type arrays on chromosome 10. Classes B-E are individuals that carry a 10-type repeat array on chromosome 4, either homogeneous (B and E) or heterogeneous (C and D). The same holds true for the individuals carrying either homogeneous (F and I), or heterogeneous (G and H) 4-type repeats on chromosome 10. Mosaic individuals are scored as M, whereas one individual had such complex repeat array structure (X) that we could not assign this individual to one of the previously defined classes. The last column (*n*) indicates the number of individuals in each class.

type arrays on chromosomes 4, either homogeneous (B and E) or heterogeneous (C and D). The same holds true for translocated homogeneous (F and I) and heterogeneous (G and H) 4-type arrays on chromosome 10. Class X displays complex repeat array configurations, which could not be assigned to classes A-I. Class M is a separate class in which all mosaic cases are represented.

2.3.4 Statistical analyses

Standard allele sizes of males (*n* = 100; class A) and females (*n* = 58; class A) were compared using the Mann-Whitney *U*-test. Allele-size distributions on chromosomes 4 and 10 in these 158 individuals were analysed according to the Kolmogorov-Smirnov test. Sensing on chromosome 4 was performed for alleles of <38 kb (FSHD range). Differences in the composition of the translocated repeats (categories B, C and D versus F, G and H; i.e. only 4-type repeats, only 10-type repeats or a combination of both) were tested against a χ^2 distribution.

In order to determine whether the peaks in the multi-modal distribution of the allele sizes were equidistant, the following procedure was applied: allele sizes on a given chromosome

were assumed to follow a mixture of two or three normal distributions with identical standard deviations but different means. Three hypotheses (H) were compared.

H_2 . The distribution of the allele sizes was explained as a mixture of two normal distributions (four estimated parameters: mean of first distribution, difference between the means of the two distributions, standard deviation and proportion of alleles in the first distribution).

H_{3e} . A mixture of three normal distributions with equidistant means (with one additional estimated parameter: the proportion of alleles in the second distribution).

H_{3u} . As H_{3e} but with non-equidistant means (one more estimated parameter: the distance between the means of the second and the third distribution).

A comparison between likelihoods obtained for H_{3e} and H_{3u} yields a likelihood ratio test with 1 degree of freedom and represents the evidence for the means of the three distributions being at different distances.

2.4 Results

2.4.1 Repeat array configurations

We examined the subtelomeric repeat array configurations on chromosome 4q35 and 10q26 in 208 unrelated healthy individuals: 128 males and 80 females. Alleles were sized by pulsed-field gel electrophoresis (PFGE) and assigned to their chromosomal origin based on their *BlnI* sensitivity. The various allele configurations were classified as indicated in *Table 2.1*. According to the Mann-Whitney *U*-test, no significant differences between repeat array distribution of males and females were observed, neither for chromosome 4 ($P = 0.84$) nor chromosome 10 ($P = 0.67$). Accordingly, data of all individuals were pooled for further analyses.

Most of the individuals (76%) displayed a standard pattern of 4-type arrays on chromosome 4 and 10-type arrays on chromosome 10, designated as class A. The remaining individuals (24%) displayed non-standard configurations with translocated repeat arrays.

2.4.2 Repeat size distribution

Only alleles of class A individuals were included in the analysis of allele size distribution (*Figure 2.1*) since this class represents the normal unbiased allele size distribution, whereas in other

classes the allele sizes may be influenced by the translocated repeat arrays. All alleles of class A individuals were grouped into intervals of 20 kb each.

Excluding allele sizes of <38 kb (associated with FSHD when residing on chromosome 4), the median of 4-type repeat arrays is 96 kb, while the median of 10-type repeat arrays is 75 kb. After correction for the difference in median repeat size, no significant difference was detected between both repeat size distributions according to the Kolmogorov-Smirnov test. As plotted in *Figure 2.1*, the repeat sizes on both chromosomes do not display a uniform distribution. The observed distributions for chromosomes 4 and 10 both agreed significantly better with a mixture of three normal distributions with equidistant means than with a mixture of two normal distributions ($\chi^2 = 27.25$, $P < 10^{-6}$ and $\chi^2 = 19.76$, $P < 10^{-5}$, for chromosomes 4 and 10, respectively). No significant further improvement of the likelihoods was obtained when a mixture of three normal distributions with non-equidistant means was fitted ($\chi^2 = 0.54$, $P \approx 0.46$ and $\chi^2 = 0.84$, $P \approx 0.36$ for chromosomes 4 and 10). The most parsimonious model therefore indicates that the allele sizes follow distributions with periodicity.

2.4.3 FSHD-sized repeat arrays

Repeat arrays <38 kb and residing on chromosome 4 are associated with FSHD. In this survey, we identified 4-type repeat arrays <38 kb in six individuals (25, 25, 30, 32, 35 and 35 kb). One

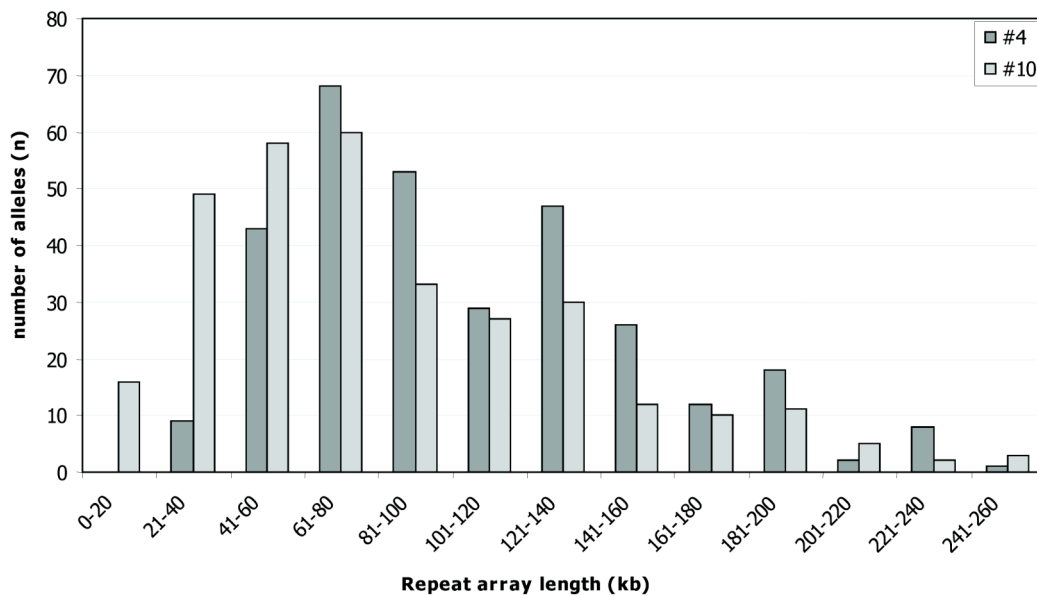


Figure 2.1

Repeat array size distribution of 4-type and 10-type repeat arrays from individuals in class A. On the x-axis the repeat array length is plotted in intervals of 20 kb. The y-axis displays the number of alleles of class A in each size interval. Chromosome 4-type repeat arrays are represented in dark gray bars, whereas 10-type repeats are represented in light gray bars.

of the individuals carrying the 30 kb 4-type repeat array was classified as F, carrying only 4-type repeat units on chromosome 10. The other individuals carried a standard repeat array configuration (class A).

2.4.4 Homogeneity of translocated repeat arrays

Twenty-one percent of individuals displayed a non-standard repeat array configuration on chromosomes 4 and 10. Ten-type repeat arrays on chromosome 4, termed B-E, were identified in 18 individuals (9%), whereas the reverse configuration i.e. 4-type repeat arrays on chromosome 10, termed F-I, was present in 25 individuals (12%) (*Table 2.1*). One individual (class X) had such complex repeat array configuration that it could not be assigned to one of the previously defined classes.

Studying all alleles in individuals carrying translocated repeat arrays (36 in classes B-E and 50 in classes F-I), 10-type arrays on chromosome 4 tended to be more heterogeneous than 4-type arrays on chromosome 10. Of the translocated 10-type arrays on chromosome 4 (classes B-E), 17% (6/36) were composed of a homogeneous array of 10-type repeat units whereas 33% (12/36) were hybrid arrays, consisting of clusters of 4-type and 10-type repeat units. Conversely, 44% (22/50) of chromosomes 10 carried a homogeneous translocated array of 4-type units only, whereas 10% (5/50) were hybrid arrays.

There was a significant difference in homogeneity of translocated repeat arrays between classes D and H ($P = 0.006$). These classes carry a comparable repeat array structure: a chromosome 4 backbone with 10-type repeats followed by a cluster of 4-type repeats (class D) and a chromosome 10 backbone with 4-type repeats followed by a cluster of 10-type repeats (class H). The other types of comparable configurations (classes C and G) do not differ significantly.

2.4.5 Somatic mosaicism

In this survey, somatic mosaicism (defined as a fifth repeat array fragment on PFGE) was observed in 3% of individuals (6/208), 4 males and 2 females. Five individuals displayed somatic mosaicism with a standard allele configuration whereas the sixth individual carried an extra 4-type repeat array on chromosome 10. Two of these individuals were mosaic for chromosome 4 and three for chromosome 10-type arrays. The presence of an extra 4-type repeat in the sixth individual did not allow us to assign the arrays to their respective chromosomes.

2.5 Discussion

In this survey, we analysed the subtelomeric D4Z4 repeat array configuration of the human chromosomes 4qter and 10qter in 208 individuals of the Dutch population. This study corroborates on an earlier study [10] and provides new insights in the behaviour of these subtelomeric domains.

The end of human chromosome 4q displays a similar structural organisation as reported for 4pter, 16pter and 22qter [15]. The distal subdomain on 4qter contains repetitive sequences as found on many chromosome ends (JE Hewitt, personal communication), whereas the proximal domain is largely comprised of the D4Z4 repeat array studied here.

The proximal and distal subtelomeric domains on chromosomes 4qter and 10qter are >95% homologous. The distal subdomain comprises 25 kb, whereas the proximal domain may vary from 50 to >500 kb due to the polymorphic 3.3 kb repeat array. Despite their high homology, 10-type repeat units can be discriminated from 4-type repeat units by an internal *BlnI* site [9]. This *BlnI*-based array discrimination makes these subtelomeres ideal models to study subtelomere plasticity. Employing PFGE, we as well as others have already demonstrated the dynamic behaviour of these subtelomeres indicative for frequent exchanges between both chromosome ends [5, 10, 26, 30].

By studying a population of >200 individuals, we obtained detailed information on the size, distribution and behaviour of these subtelomeric repeat arrays. From this study, it becomes evident that both chromosome ends present similar characteristics with regard to their size distributions and compositions, but also display discrete differences.

We found that the median of the repeat array on chromosome 4 is 21 kb larger than that on chromosome 10. A telomere position effect, in which telomeric heterochromatin has the capacity to silence genes in a distance-dependent manner, is well documented in yeast (reviewed in [19]). Similar to this telomeric gene silencing, it has been postulated that the D4Z4 repeat on chromosome 4 and its homologue on chromosome 10 prevents spreading of heterochromatin into (euchromatic) regions proximal to the repeat and that partial deletion of the chromosome 4 repeat array in FSHD undermines this spacer function [44]. In this light, the length difference between both subtelomeric repeats may reflect a requirement for larger subtelomeric domains on chromosome 4 to prevent gene silencing of critical gene(s) in the region proximal to the subtelomere.

An unexpected high frequency of 4-type repeat arrays <38 kb was observed in this survey (6 individuals = 3%). Five of them belong to class A. Whilst residing on chromosome 4, these arrays are associated with FSHD. Several possibilities may explain this finding. First, it is possible that the array does not reside on chromosome 4. Approximately 10% of individuals carry a 4-

type repeat on chromosome 10 and an equal frequency of 10-type repeats on chromosome 4 is observed, we would expect the presence of ‘double exchanged’ alleles (i.e. a 4-type array on chromosome 10 and a 10-type allele on chromosome 4) in 1% of individuals. Second, one of the individuals carrying a 4-type array of 30 kb was assigned to class F and therefore carries three 4-type repeat arrays. It may well be that in this individual the potential pathogenic repeat array resides on chromosome 10. Definite chromosomal assignments of these repeat arrays can only be obtained by additional PFGE analyses on agarose-embedded DNA plugs using chromosome 4- and 10-specific probes [26]. Since DNA plugs of these random individuals are not available we are unable to pursue this procedure. Thirdly, ~30% of the gene carriers is asymptomatic or have subclinical FSHD signs [33, 34]. Moreover, a correlation has been established between the residual repeat size and the age at onset and severity of FSHD [29, 37]. Individuals carrying small sized FSHD alleles are generally more severely affected than individuals with larger FSHD alleles and it may well be that there is an incomplete penetrance for larger FSHD alleles. Since this unselected group was not clinically evaluated, it is very possible that individuals with a repeat array <38 kb have subclinical characteristics or are non-penetrant. In this light, it is noteworthy that the 4-type arrays in three of the five class A individuals are >30 kb.

We confirmed that 4-type repeat arrays on chromosome 10 occur equally frequently as the reverse configuration of 10-type arrays on chromosome 4 (12% and 9%, respectively). However, the composition of these translocated arrays on both chromosomes differs significantly. Although 4-type repeats on chromosome 10 tend to be homogeneous (44% of all chromosome 10 alleles), 10-type repeats on chromosome 4 are mostly not homogeneous (only 17%) but a combination of 4- and 10-type repeats (33% of all chromosome 4 alleles). This suggests a biological difference between both repeat arrays resulting in a preference of 4-type units on chromosome 4. However, that would also predict a significant difference between classes C and G, which is not observed. Therefore, we favour a mechanistic explanation.

Most repetitive sequences, whether coding or non-coding, undergo concerted evolution. This process homogenises repetitive sequences and is thought to be important for the maintenance of the integrity of each repeat unit. Concerted evolution is achieved by a much faster intrachromosomal homogenisation than interchromosomal recombination (reviewed in [27]). The heterogeneity observed for the D4Z4 subtelomeric repeats suggests that these loci escape concerted evolution and evolve by rather unconstrained inter- and intrachromosomal recombination. This may imply that the putative open reading frame, *DUX4*, present in each repeat unit and for which no expression has been observed *in vivo* [17], may have lost its function during the rapid expansion of this repeat array.

Our results are indicative for a multi-modal allele size distribution on both chromosomes with three equidistant peaks at intervals of ~65 kb. For minisatellite repeats,

bimodal and trimodal distributions have been observed [35]. This is mostly attributed to a founder effect of two ancestral alleles that have a similar dynamic behaviour and show little inter-chromosomal interactions. Since we observe the same unequal size distribution on both chromosomes and have already demonstrated a frequent interchromosomal cross-talk, a founder effect for these loci is highly unlikely. An alternative explanation that the chromatin structure of repeat sequences may impose restrictions on repeat lengths seems more likely. Degradation of high molecular weight DNA by nuclease digestion has suggested that chromatin loops contain ~50 kb of DNA [14]. We have already demonstrated in mosaic individuals of *de novo* FSHD families that usually the shortest allele rearranges to an FSHD-sized allele [30]. This suggests that the unequal distribution on chromosome 4 separates the repeat arrays in a premutation domain (first peak) and a normal domain (larger than the first peak).

Previously, we demonstrated that the FSHD deletion occurs mitotically in at least 40% of cases. Moreover, we showed that the presence of supernumerary fully homologous repeat arrays is a predisposing factor (in this case, the presence of 4-type repeats on chromosome 10) [30]. In this study, we have identified somatic mosaicism for one of the alleles in as many as 3% of the individuals. Mosaicism was observed for both 4-type and 10-type alleles. This high mutation frequency emphasises the importance of mitotic recombination in subtelomeric homogenisation.

In conclusion, this study provides a detailed insight into the complex dynamic characteristics of the proximal subtelomeric domains on chromosomes 4 and 10. Although these domains share many properties and may frequently interact, they also display distinct differences in size and homogeneity. Their plasticity is emphasised by a very high somatic mutation frequency. It will be interesting to examine whether this dynamic behaviour is just a side effect of mitosis and meiosis, or whether it supports distinct subtelomeric functions.

2.6 Acknowledgements

This study was supported by the Dutch Organisation of Scientific Research (NWO), the Prinses Beatrix Fonds, the Muscular Dystrophy Association (USA) and the Dutch FSHD Foundation. SM van der Maarel is a Marjorie Bronfman fellow of the FSH Society.

2.7 References

1. Altherr MR, Bengtsson U, Elder FF, Ledbetter DH, Wasmuth JJ, McDonald ME, Gusella JF, and Greenberg F (1991) Molecular confirmation of Wolf-Hirschhorn syndrome with a subtle translocation of chromosome 4. *Am J Hum Genet* 49(6): 1235-1242.
2. Baird DM and Royle NJ (1997) Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. *Hum Mol Genet* 6(13): 2291-2299.
3. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, Wielen van der M, Rasmussen K, and Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2: 39-44.
4. Biessmann H and Mason JM (1992) Genetics and molecular biology of telomeres. *Adv Genet* 30: 185-249.
5. Cacurri S, Piazzo N, Deidda G, Vigneti E, Galluzzi G, Colantoni L, Merico B, Ricci E, and Felicetti L (1998) Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric *KpnI* repeat units implicated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 63: 181-190.
6. Carlson M, Celenza JL, and Eng FJ (1985) Evolution of the dispersed SUC gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* 5(11): 2894-2902.
7. Chan CS and Tye BK (1983) Organization of DNA sequences and replication origins at yeast telomeres. *Cell* 33(2): 563-573.
8. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, and Felicetti L (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3: 155-167.
9. Deidda G, Cacurri S, Piazzo N, and Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet* 33(5): 361-365.
10. 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.
11. Deutekom van JCT, Lemmers RJLF, Grewal PK, Geel van M, Romberg S, Dauwerse HG, Wright TJ, Padberg GW, Hofker MH, Hewitt JE, *et al.* (1996) Identification of the first gene (*FRG1*) from the FSHD region on human chromosome 4q35. *Hum Mol Genet* 5(5): 581-590.
12. 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.
13. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, *et al.* (1987) A genetic linkage map of the human genome. *Cell* 51(2): 319-337.
14. Filipinski 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.
15. Flint J, Bates GP, Clark K, Dorman A, Willingham D, Roe BA, Micklem G, Higgs DR, and Louis EJ (1997) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet* 6(8): 1305-1314.
16. Flint J, Wilkie AOM, Buckle VJ, Winter RM, Holland AJ, and McDermid HE (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* 9(2): 132-140.

17. Gabriels J, Beckers MC, Ding H, Vriese de A, Plaisance S, Maarel van der SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al.* (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 236(1): 25-32.
18. Gottschling DE, Aparicio OM, Billington BL, and Zakian VA (1990) Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63(4): 751-762.
19. Guarente L (1999) Diverse and dynamic functions of the Sir silencing complex. *Nat Genet* 23(3): 281-285.
20. Horowitz H and Haber JE (1984) Subtelomeric regions of yeast chromosomes contain a 36 base-pair tandemly repeated sequence. *Nucleic Acids Res* 12(18): 7105-7121.
21. Horowitz H and Haber JE (1985) Identification of autonomously replicating circular subtelomeric Y' elements in *Saccharomyces cerevisiae*. *Mol Cell Biol* 5(9): 2369-2380.
22. Horowitz H, Thorburn P, and Haber JE (1984) Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Mol Cell Biol* 4(11): 2509-2517.
23. Kipling D (1995) The telomere, Oxford University Press.
24. Kuwano A, Ledbetter SA, Dobyns WB, Emanuel BS, and Ledbetter DH (1991) Detection of deletions and cryptic translocations in Miller-Dieker syndrome by in situ hybridization. *Am J Hum Genet* 49(4): 707-714.
25. Lamb J, Wilkie AO, Harris PC, Buckle VJ, Lindenbaum RH, Barton NJ, Reeders ST, Weatherall DJ, and Higgs DR (1989) Detection of breakpoints in submicroscopic chromosomal translocation, illustrating an important mechanism for genetic disease. *Lancet* 2(8667): 819-824.
26. 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, *et al.* (1998) Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet* 7(8): 1207-1214.
27. Liao D (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 64(1): 24-30.
28. Louis EJ (1995) The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 11(16): 1553-1573.
29. 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.
30. 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.
31. Miller SA, Dykes DD, and Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3): 1215.
32. NIH-CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. *Science* 258(5079): 67-86.
33. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
34. Padberg GW, Lunt PW, Koch M, and Fardeau M (1991) Diagnostic criteria for facioscapulohumeral muscular dystrophy. *Neuromusc Disord* 1(4): 231-234.
35. Stead JD and Jeffreys AJ (2000) Allele diversity and germline mutation at the insulin minisatellite. *Hum Mol Genet* 9(5): 713-723.
36. Szostak JW and Blackburn EH (1982) Cloning yeast telomeres on linear plasmid vectors. *Cell* 29(1): 245-255.
37. Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, and Figlewicz D (1996) Evidence for anticipation and association of deletion size with

- severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39(6): 744-748.
38. Wallace BM and Hulten MA (1985) Meiotic chromosome pairing in the normal human female. *Ann Hum Genet* 49(Pt 3): 215-226.
 39. Walmsley RW, Chan CS, Tye BK, and Petes TD (1984) Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* 310(5973): 157-160.
 40. Weiner AM, Deininger PL, and Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* 55: 631-661.
 41. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, and Lathrop M (1992) A second-generation linkage map of the human genome. *Nature* 359(6398): 794-801.
 42. 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.
 43. Wilkie AO, Higgs DR, Rack KA, Buckle VJ, Spurr NK, Fischel-Ghodsian N, Ceccherini I, Brown WR, and Harris PC (1991) Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* 64(3): 595-606.
 44. Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, *et al.* (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosome Res* 2(3): 225-234.
 45. Zakian VA (1989) Structure and function of telomeres. *Annu Rev Genet* 23: 579-604.

3

Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy

*RJLF Lemmers¹, PGM van Overveld¹, LA Sandkuijl¹,
H Vrieling², GW Padberg³, RR Frants¹ and SM van
der Maarel¹*

Originally published in:
American Journal of Human Genetics
(2004) 75: 44-53

¹ Department of Human Genetics,
Center for Human and Clinical
Genetics, Leiden University Medical
Center, Leiden, The Netherlands.

² Department of Toxicogenetics,
Leiden University Medical Center,
Leiden, The Netherlands.

³ Department of Neurology, University
Medical Center St. Radboud,
Nijmegen, The Netherlands.

3.1 Abstract

Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD1A) is associated with contractions of the polymorphic D4Z4 repeat on chromosome 4qter. Almost half of new FSHD mutations occur post-fertilisation, resulting in somatic mosaicism for D4Z4. Detailed D4Z4 analysis of 11 mosaic individuals with FSHD revealed a mosaic mixture of a contracted FSHD-sized allele and the unchanged ancestral allele in 8 cases, which is suggestive of a mitotic gene conversion without crossover. However, in 3 cases, the D4Z4 rearrangement resulted in two different-sized D4Z4 repeats, indicative of a gene conversion with crossover. In all cases, DNA markers proximal and distal to D4Z4 showed no allelic exchanges, suggesting that all rearrangements were intrachromosomal. We propose that D4Z4 rearrangements occur via a synthesis-dependent strand annealing model that relatively frequently allows for crossovers. Furthermore, the distribution of different cell populations in mosaic patients with FSHD suggests that mosaicism here results from D4Z4 rearrangements occurring during the first few zygotic cell divisions after fertilisation.

3.2 Introduction

A large proportion of eukaryotic genomes consists of repetitive DNA. The head-to-tail tandem repetitive sequences, categorized according to repeat unit length, are highly polymorphic. Micro-, mini-, macro- and megasatellite repeats are distinguished, each with their own mutation characteristics. Several hypothetical rearrangement models have been proposed, mainly on the basis of micro- and minisatellite instability. These often include gene conversions with or without crossover [12, 35]. However, little is known about macro- and megasatellite instability in human cells.

Facioscapulohumeral muscular dystrophy (FSHD1A [MIM 158900], hereafter referred to as “FSHD”) is an early onset, autosomal dominant myopathy mainly characterised by a progressive and variable atrophy and weakness of the facial, shoulder and upper arm muscles [32]. The FSHD phenotype also presents with extramuscular symptoms such as retinovasculopathy, hearing loss, and, in severe cases, mental retardation [1, 5, 27, 33]. FSHD is associated with a macrosatellite repeat instability: the contraction of the subtelomeric D4Z4 repeat on 4q35. This polymorphic D4Z4 repeat is highly recombinogenic, since somatic mosaicism for a rearrangement of D4Z4 is found in as much as 3% of the general population [31]. The D4Z4 repeat consists of identical *KpnI* units, each 3.3 kb in size, ordered in a head-to-

tail fashion and varying between 11 and 100 units on healthy chromosomes [4]. Patients with FSHD carry a repeat of 1-10 units on one of their chromosomes 4 [40]. A rough and inverse correlation has been observed between the severity and age at onset of the disease and the residual repeat unit number [20, 38].

The subtelomere of chromosome 10q is nearly identical to chromosome 4qter and also contains a highly homologous and equally recombinogenic repeat. D4Z4 repeats on chromosomes 4 and 10 are visualised in *EcoRI*-digested DNA with probe p13E-11, which recognizes the region proximal to the D4Z4 repeat within the *EcoRI* fragment [40]. To discriminate between 4qter- and 10qter-derived repeats, in addition to *EcoRI*, the restriction enzyme *BlnI* is used, which digests only chromosome 10-derived repeat units [2]. Conversely, the restriction enzyme *XapI* is specific to 4qter-derived repeat units and can be used to complement *BlnI* [15]. DNA separated by pulsed-field gel electrophoresis (PFGE), combined with differential digestion with *BlnI* and *XapI*, allows sizing and chromosomal assignment of all D4Z4 alleles ranging from 10 to >350 kb.

In 10% of the population, translocated 4-type repeats reside on chromosome 10q and, vice versa, translocated 10-type repeats on chromosome 4q are equally frequent. This was first found in a Dutch survey [3, 31], but comparable translocation frequencies have now also been reported in the Japanese, Korean and Chinese populations [23]. These translocated repeats can be homogeneous or heterogeneous (consisting of both 4-type and 10-type units). Although this observation may suggest frequent interactions between the regions of homology of both non-homologous chromosomes, *de novo* exchanges between repeat arrays originating from chromosomes 4 and 10 have never been described. In contrast to chromosome 4, repeat contractions on chromosome 10 are non-pathogenic.

Further complicating our understanding of FSHD pathogenesis, a 4qter polymorphism distal to D4Z4 was recently described, giving rise to two distinct 4q chromosome ends: 4qA and 4qB [8]. Both alleles are almost equally common and equally recombinogenic in controls. Nevertheless, FSHD alleles are always of the 4qA type [16].

As many as 10-30% of patients with FSHD carry a new mutation [19, 34, 44], with D4Z4 repeat contractions varying from 1 to >50 units [22]. Several studies have shown that almost half of these D4Z4 rearrangements are of mitotic origin, resulting in somatic mosaicism for the FSHD allele in either patients or asymptomatic parents [7, 14, 22, 39, 43]. A relationship has been established between the severity of the disease and the percentage of peripheral blood lymphocytes (PBLs) carrying the disease allele in combination with the residual repeat size [22]. This suggests that the proportion of mosaicism detected in blood is representative of the mosaicism in muscle tissue. Furthermore, patients with FSHD have been described who displayed mosaicism for the disease allele in PBLs and germline [7, 17], as well as

in fibroblasts (RJLF Lemmers, unpublished data). Thus, mitotic D4Z4 rearrangements very likely occur early during embryogenesis.

In all mosaic cases reported, two PBL cell populations were identified: one carrying the parental, non-rearranged allele, and one in which the D4Z4 repeat had been rearranged to FSHD size. Such an allele distribution generally is suggestive of a rearrangement by gene conversion without crossover, since the hallmark of this mechanism is an unchanged donor allele and a changed acceptor allele [35].

Although many macrosatellite repeats have been identified in the human genome, little is known about the mechanism by which they rearrange. Mitotic minisatellite rearrangements were thoroughly studied in yeast, in which gene conversion, usually without crossover, was shown to be the main mechanism of rearrangement [35].

Similar to D4Z4, the polymorphic human RS447 megasatellite repeat, which is comprised of 4.7 kb units displays both mitotic and meiotic instability. For RS447, a high frequency of repeat contraction and expansion has been shown [29]. The tandemly repeated human U2 snRNA genes (RNU2 locus) have been studied more extensively and provided insight into the mechanism of concerted evolution. These 6.1 kb U2 repeats display repeat homogenisation by rare interchromosomal gene conversion followed by rapid intrachromosomal homogenisation [18].

The mechanism by which D4Z4 rearranges was further examined in the presented study, through use of the 4qA/4qB polymorphism as distal flanking marker in combination with a newly identified RFLP within D4Z4 proper, serving as a proximal flanking marker. Studying these proximal and distal markers provided insight into the mechanism and timing of D4Z4 rearrangements and may stand as a model for macrosatellite repeat mutation in general.

3.3 Subjects and methods

3.3.1 Patients and control individuals

All families were ascertained via one of the Dutch Neuromuscular Centers. In all but one family, the mosaic child received a clinical diagnosis of FSHD, and the parents did not display any FSHD features. In family 55611, the FSHD allele of the affected mother was further somatically contracted in the affected child. Blood from these families was collected after informed consent was obtained. The *Pvu*II polymorphism was determined in a previously described population of anonymised Dutch blood donors with a standard allele configuration of homogeneous 4-type repeats on chromosome 4 and 10-type repeats on chromosome 10 [31]. Four of 11 mosaic

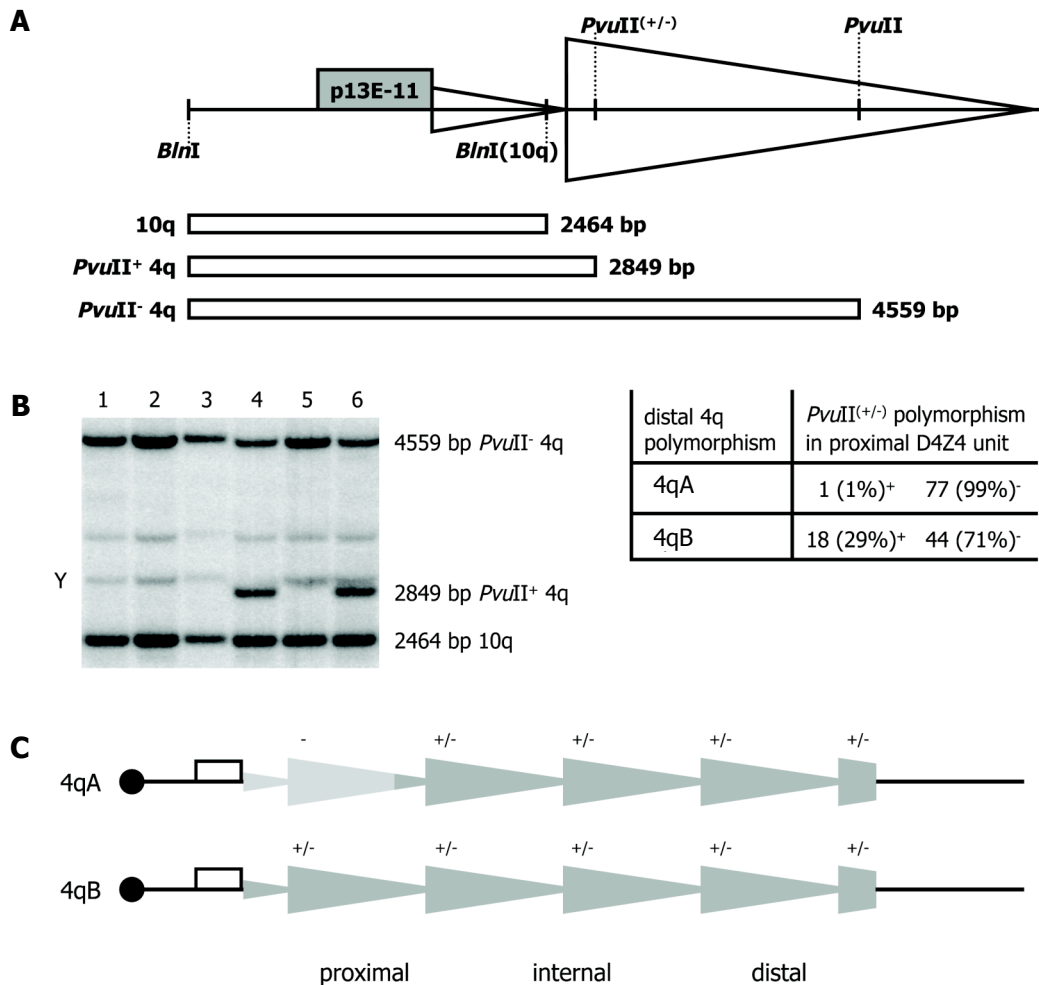


Figure 3.1

(A) Polymorphic *PvuII* site at position 6044 of GenBank accession number AF117653 was used to study recombination between 4qA and 4qB alleles. A double digestion with *PvuII*/*BlnI* normally gives rise to fragments of 4559 bp (chromosome 4 alleles) and 2464 bp (chromosome 10 alleles). However, the presence of an extra *PvuII* site within the proximal D4Z4 unit yields a fragment of 2849 bp.

(B) Example of the modified *PvuII*/*BlnI* dosage test after hybridisation with probe p13E-11. Samples in lanes 1-3 carry two 4qA alleles and two 10q alleles, all of which are *PvuII*-resistant (*PvuII*⁻). Samples in lanes 4-6 carry two 4qB alleles and two 10q alleles; in lane 4 and lane 6, a *PvuII*-sensitive (*PvuII*⁺) fragment is visible. "Y" indicates the cross-hybridising Y chromosome. The table on the right shows dosage experiment results on the *PvuII* RFLP in the proximal D4Z4 unit of 78 different 4qA and 62 different 4qB alleles. Twenty-nine percent of the 4qB alleles are *PvuII*⁺, whereas 4qA alleles are almost exclusively *PvuII*⁻ ($P < 10^{-5}$).

(C) Distribution of the *PvuII* RFLP on 4qA and 4qB alleles. The *PvuII*⁻ proximal D4Z4 unit on 4qA alleles is indicated by " - ", whereas D4Z4 units that were heterozygous for *PvuII* RFLP are indicated by " +/- ".

patients with FSHD studied carried a homogeneous translocated 4-type repeat on chromosome 10.

3.3.2 Somatic cell hybrids and DNA clones

The following chromosome 4qA sources were used: YAC Y25C-2E [41]; the monochromosomal rodent somatic cell hybrids HHW1494, SU10 (a gift from S Winokur, Irvine, CA), and GM11448 (Coriell Institute for Medical Research); and lambda clones λ 42, λ 68, and λ 260201

[4]. All sources, except for Y25C-2E and GM11448, represent patient alleles. Somatic cell hybrid GM11448 contains a human chromosome 4qA with a D4Z4 repeat of 125 kb (data not shown).

As chromosome 4qB sources, we used the monochromosomal rodent somatic cell hybrids GM11687 (Coriell Institute for Medical Research), 4L-10 (a gift from E Stanbridge, Irvine, CA), HHW416 (a gift from M Altherr, Los Alamos, NM) and 361-9 (a gift from S Winokur, Irvine, CA), with D4Z4 repeats of 140 kb, 85 kb, 96 kb, and 290 kb, respectively.

3.3.3 DNA isolation

DNA was isolated from PBLs essentially as described by Miller *et al.* [26]. If possible, PBLs were embedded in agarose plugs (InCert agarose, [FMC]) at a concentration of 7.5×10^5 cells per plug.

3.3.4 D4Z4 analysis

For allele sizing, DNA samples were double digested with *EcoRI/HindIII* and with *EcoRI/BlnI*. For analysis of the allelic variation at 4qter (probes 4qA and 4qB, [16]), DNA was digested with *HindIII* only. All digestions were performed according to the manufacturer's instructions. *EcoRI* and *HindIII* were purchased from MBI Fermentas, *PvuII* and *BlnI* were purchased from Amersham-Pharmacia. DNA was separated in 22 hours on a 0.8% agarose gel (MP agarose [Boehringer]) by pulsed-field gel electrophoresis (PFGE) at 8.5 V/cm at 21°C in 0.5× tris-borate EDTA (TBE) supplemented with 150 pg/ml ethidium bromide. PFGE was performed in four identical cycles, with a switch time increasing linearly from 1 s at the start to 16 s at the end of each cycle. Subsequently, DNA was transferred to a Hybond XL membrane (Amersham-Pharmacia) by Southern blotting and was hybridised in a buffer containing 0.125 M Na_2HPO_4 (pH 7.2), 10% PEG6000, 0.25 M NaCl, 1 mM EDTA, and 7% SDS, for 16-24 h at 65°C. Hybridisations with probe p13E-11 (D4F104S1, [40]) were washed in 2× SSC/0.1% SDS, and those with probes 4qA or 4qB were washed more stringently in 1× SSC/0.1% SDS. The blots were exposed for 16-24 hours to phosphorimager screens and were analysed with the ImageQuant software program (Molecular Dynamics).

The *PvuII* polymorphism in the first proximal unit of the D4Z4 repeat (position 6044 of GenBank accession number AF117653) was analysed by applying a modified *BglII/BlnI* dosage test [21] on genomic DNA samples by double digestion with *PvuII* and *BlnI* (**Figure 3.1**). After Southern blot and hybridisation with probe p13E-11, blots were washed in 0.3× SSC/0.1% SDS. Chromosome 4-derived fragments are 2849 bp or 4559 bp in size, depending on the presence or absence of the *PvuII* restriction site, whereas chromosome 10-derived fragments are 2464 bp in size, because of the presence of *BlnI*. Blots were exposed to phosphorimager screens, and signals

were quantified with the ImageQuant software program. Restriction analysis of this polymorphism in internal D4Z4 units was performed on chromosome 4-only sources. A set of seven 4qA- and four 4qB-derived repeats were hybridised with the D4Z4 repeat probe 9B6A [41] after digestion with *PvuII*, separation on a 0.8% TBE agarose gel, and transfer to Hybond XL membranes. The 3.3 kb fragment is derived from *PvuII* units, whereas the 1.6 kb fragment represents *PvuII*⁺ units. The most distal D4Z4 unit of the repeat array was analysed by specific PCR amplification of this unit, followed by a *PvuII* digestion (primer sequences available upon request).

3.3.5 Haplotyping

Haplotype analysis for family 36 was performed using 24 polymorphic markers, including D4S163, D4S139, D10S555, and D10S595 (P de Knijff, personal communication).

3.3.6 Statistical analysis

The presence of the *PvuII* site in the proximal unit of the D4Z4 repeat of 4qA and 4qB alleles was compared using the Pearson χ^2 test. Allele size distributions in 88 control individuals (84 4qA and 92 4qB alleles) were analysed according to one-way ANOVA and a non-parametric two-sample Kolmogorov-Smirnov test.

3.4 Results

3.4.1 Chromosomal partner for D4Z4 rearrangements

Repeat rearrangements usually occur between sister chromatids or homologous chromosomes. To reveal the common partner for D4Z4 rearrangements, the allele size distributions of 4qA- and 4qB-derived D4Z4 repeats were compared. A different D4Z4 repeat length distribution of 4qA- and 4qB-type alleles would indicate that these chromosomes do not often interact. The sizing was performed by PFGE on homogeneous 4-type D4Z4 repeats obtained from high-quality agarose-embedded DNA plugs. In total, 84 4qA- and 92 4qB-type alleles were analysed. Statistical analyses revealed that D4Z4 repeats derived from 4qA alleles (mean $\pm$ 95% CI: 136 $\pm$ 7 kb) were significantly longer than those derived from 4qB alleles (mean $\pm$ 95% CI: 93 $\pm$ 4 kb) and that they had a different size distribution ($P < 0.001$). This indicates that, in general, D4Z4 rearrangements occur intrachromosomally.

Table 3.1 Sizing and typing of chromosome 4-derived D4Z4 alleles from mosaic patients with FSHD.

Family	Sex of patient	Length and proportion of D4Z4 alleles (4qA-type) after involvement in mitotic rearrangement									
		Ancestral allele		Contracted allele ^a		Rearranged allele 2 ^b		Chromosomal		Homologous allele	
		length (kb)	%	length (kb)	%	length (kb)	%	origin	length (kb)	type	
1	Male	70	37	14	26	96	37	Paternal	96	4qB	
5	Male	130	40	14	60			Maternal	80	4qA	
6	Male	280	60	10	40			Paternal	175	4qA	
7	Male	160	70	16	30			Paternal	65	4qB	
12	Female	80	10	23	90			Maternal	45	4qB	
24	Male	50	50	17	50			Maternal	80	4qB	
26	Male	130	50	14	50			Maternal	80	4qB	
35	Male	90	50	33	50			Maternal	220	4qA	
36	Male	240	0	16	60	384	40	Maternal	135	4qA	
Rf120	Female	100	20	20	70	55	10	Maternal	50	4qB	
55611	Male	33	80	30	20			Maternal	130	4qB	

a: Contracted FSHD-causing allele. b: Mitotic gene conversions with crossover produce two rearranged D4Z4 alleles (contracted and rearranged allele 2).

To further elucidate the mechanism underlying D4Z4 rearrangements, we studied a polymorphic *Pvu*II site (position 6044 in GenBank accession number AF117653 [6]) in the proximal D4Z4 unit of the repeat through use of a modified *Bgl*II-*Bln*I dosage test (*Figure 3.1*) [21]. A *Pvu*II/*Bln*I double digestion was performed to separate chromosome 4 *Pvu*II-sensitive (2849 bp) from *Pvu*II-resistant (4559 bp) alleles. In addition, *Bln*I was used to avoid interference of chromosome 10-type alleles (2464 bp). Analysis of this *Pvu*II-RFLP in 78 independent 4qA alleles and 62 independent 4qB alleles (*Figures 3.1.A* and *3.1.B*) showed that the *Pvu*II site is present in 29% of 4qB alleles and only in 1% of 4qA alleles ($P < 10^{-5}$), suggesting a marked linkage disequilibrium (LD) between the proximal and distal ends of the D4Z4 repeat. Next, internal D4Z4 units were analysed for the *Pvu*II-RFLP on DNA sources containing only a single chromosome 4-derived D4Z4 repeat. A set of seven 4qA- and four 4qB-derived D4Z4 repeats revealed that the repeat units following the first unit were polymorphic for the *Pvu*II polymorphism on both alleles (data summarised in *Figure 3.1.C*).

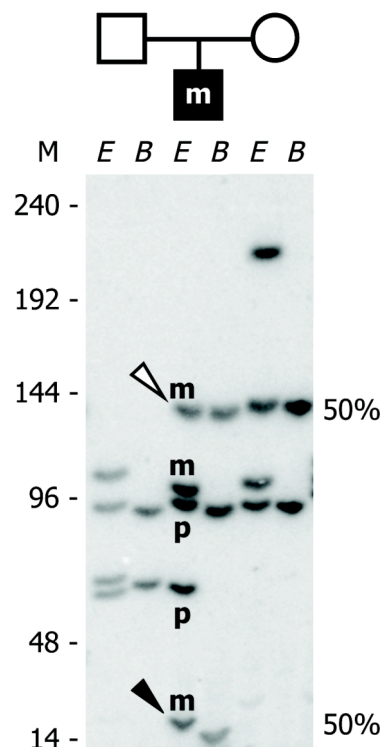


Figure 3.2

Example of p13E-11 hybridisation after PFGE and Southern blot analysis: analysis of mosaic family 26 with FSHD (GC-CO = gene conversion without crossover). DNA was double digested with *Eco*RI/*Hind*III (“E”) and *Eco*RI/*Bln*I (“B”), separated by PFGE and hybridised with p13E-11. The patient inherited the paternal (p) 60 kb chromosome 10 allele and 90 kb 4qB allele and, from his mother (m) the 100 kb chromosome 10 allele and the 140 kb 4qA allele (unblackened arrowhead), which is present in only 50% of his PBLs. The 14 kb 4qA allele (blackened arrowhead) causing FSHD is derived from the 140 kb maternal allele and is detected in 50% of his PBLs. Estimation of the proportion of mosaic alleles is based on the signal intensities of the D4Z4 fragments, with a CI of 5%. Allele types are not shown. Marker sizes (M; in kb) are indicated on the left.

3.4.2 D4Z4 repeat analysis of mosaic FSHD kindreds

Eleven kindreds with FSHD, consisting of father, mother and mosaic child with FSHD, were analysed for their D4Z4 repeat size and distal 4qA/4qB origin (*Table 3.1*). From most of these families, D4Z4 allele sizes have been described elsewhere [22]. In all mosaic individuals, we identified three non-mosaic alleles: two alleles derived from chromosome 10 and one derived from the unaffected chromosome 4 homologue. In all cases, the *de novo* mosaic alleles had the same distal flanking marker (4qA) as their ancestral alleles, despite heterozygosity for this distal flanking marker in seven of them [16] (*Table 3.1*). This suggests that D4Z4 rearrangements generally do not occur between both chromosomes 4. In 8 of these 11 mosaic individuals, we observed a mosaic mixture of a *de novo* FSHD allele and an unchanged parental allele, suggestive of a mitotic D4Z4 rearrangement by gene conversion without crossover. An example is shown in *Figure 3.2*. The remaining three mosaic patients displayed a more complex mosaic

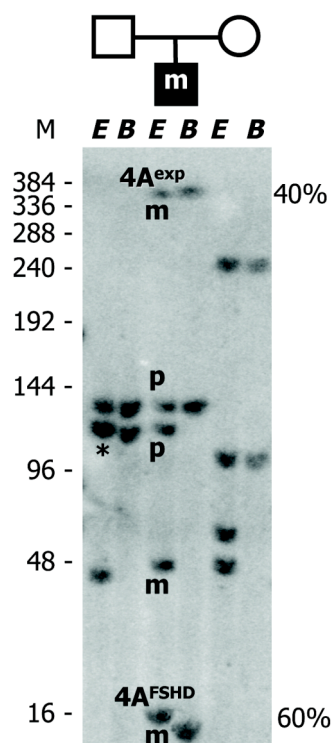


Figure 3.3 (part 1)

Example of p13E-11 hybridisation after PFGE and Southern blot. Analyses, as in *Figure 3.2*, of mitotic interchromatid gene conversion with crossover (GC+CO) in families 36 and 1 with FSHD. Estimation of the proportion of mosaic alleles is based on the signal intensities of the D4Z4 fragments, with a CI of 5%. Allele types are not shown.

EcoRI/HindIII (E) and *EcoRI/BlnI* (B) digestion of DNA from family 36 with *de novo* FSHD. The affected male inherited 10q alleles of 50 kb and 130 kb and a 140 kb 4qA allele. He also has mosaic alleles of 384 kb (40%; 4A^{exp}) and 16 kb (60%; 4A^{FSHD}), which are expanded and contracted alleles, respectively, most probably produced by an interchromatid gene conversion with crossover in the maternal 100 kb or 240 kb allele. Maternally and paternally inherited alleles are indicated by “m” and “p”, respectively. Co-migrating chromosome 4 and 10 alleles of 130 kb in the father are marked with an asterisk. Marker sizes (M in kb) are indicated on the left.

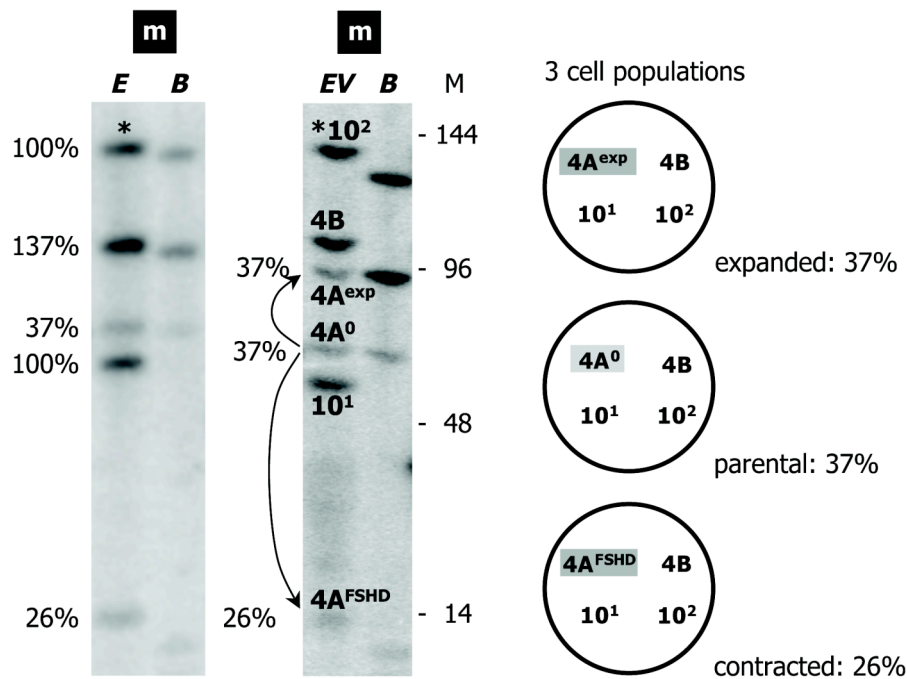


Figure 3.3 (part 2)

Example of p13E-11 hybridisation after PFGE and Southern blot. Analyses, as in *Figure 3.2*, of mitotic interchromatid gene conversion with crossover (GC+CO) in families 36 and 1 with FSHD. Estimation of the proportion of mosaic alleles is based on the signal intensities of the D4Z4 fragments, with a CI of 5%. Allele types are not shown.

The left panel shows *EcoRI/HindIII* (“E”) and *EcoRI/BlnI* (“B”) digestion of DNA from family 1 with *de novo* FSHD. The patient from this family has three mosaic alleles, of 14 kb, 70 kb and 96 kb, that originated from the paternal 70 kb 4qA allele. The 96 kb 4qA mosaic allele is co-migrating with a maternal 96 kb 4qB allele. Furthermore, he inherits a paternal 60 kb 10q allele and a maternal 135 kb 4qB allele (marked with an asterisk), which originated from chromosome 10 (translocated 4-type repeat (t10;4)) [3]. Quantification of all alleles revealed 100% intensity for the 135 kb (t10;4) 4qB allele, 137% for the co-migrating 96 kb 4qA and 4qB alleles, 37% for the 70 kb 4qA allele, 100% for the 60 kb 10q allele, and 26% for the 14 kb 4qA allele. An *EcoRV* (“EV”) digestion allows separation of the co-migrating 4qA and 4qB alleles at 96 kb (137%). On the basis of *EcoRV* digestion (right panel) the patient clearly has six alleles; of these, the contracted 14 kb 4qA allele (26%; FSHD allele, 4A^{FSHD}) and the expanded 96 kb 4qA allele (37%; 4A^{exp}) originate from the mitotic rearranged parental 70 kb allele (37%; 4A⁰). Two arrows indicate D4Z4 expansion and contraction. Allele quantifications after *EcoRV* digestion are indicated in the right example panel. Marker sizes (M; in kb) are indicated at the right. In the box, the constitution of the three different cell populations of this patient is schematically depicted.

pattern, since they carry two somatically rearranged D4Z4 alleles, suggestive of rearrangements with crossover.

In family 36, we observed a mosaic patient with a mosaic mixture of a 384 kb-sized expanded and a 16 kb-sized contracted allele in almost equal proportions of PBLs (*Figure 3.3 part 1*). Haplotyping of the patient and his parents excluded non-parenthood (data not shown). In family 1, the presence of three mosaic alleles was initially overlooked due to the co-migration of one of these mosaic alleles with the non-rearranged chromosome 4qB homologue (*Figure 3.3 part 2*). However, sequence comparison of the region distal to D4Z4 on 4qA (GenBank accession numbers. U74496 and U74497) and 4qB (GenBank accession number AF017466, [8]) revealed an RFLP for *EcoRV*. This *EcoRV* restriction site is directly adjacent to D4Z4 on

4qA, whereas its location on 4qB is 10 kb distal to the D4Z4 repeat (data not shown), allowing separation of these co-migrating 4qA and 4qB alleles and visualisation of the previously undetected mosaic allele. The patient now clearly displayed three mosaic alleles: one FSHD-sized allele of 14 kb, one expanded allele of 96 kb, and the parental unchanged allele of 70 kb, each of them equally frequent in the PBLs (*Figure 3.3 part 2*). The third complex mosaic patient, in family Rf120 (data not shown), displays a mixture of three mosaic alleles: two *de novo* contracted alleles of 20 kb and 55 kb in 70 and 10% of PBLs, respectively, and the unchanged parental-sized allele of 100 kb in the remainder of cells.

3.5 Discussion

To elucidate the mechanism by which D4Z4 rearranges, we analysed the exact repeat size, composition (*Pvu*II-RFLP), and origin (4qA/4qB) of all D4Z4 repeats in individuals in whom we had established that a D4Z4 rearrangement had occurred *de novo* through the presence of mitotically rearranged D4Z4 repeats. We observed that the mechanism by which D4Z4 rearranges shares features common to that of other tandemly repeated DNA sequences.

3.5.1 General features of D4Z4 rearrangements

To reveal the common partner in D4Z4 rearrangements, D4Z4 repeat length distributions on 4qA and 4qB alleles were compared. Statistical analyses show a significant difference in the distribution between 4qA- and 4qB-type alleles, excluding a frequent recombination between D4Z4 repeats on 4qA and 4qB chromosomes. Most probably, D4Z4 rearrangements preferentially occur between sister chromatids, the most common partners for rearrangements in mammalian cells [13]. In addition, the repetitive nature of D4Z4 enables an intrachromatid rearrangement by the formation of an intra-allelic DNA loop.

In all mosaic cases studied we observed no change of distal flanking marker 4qA/4qB between *de novo* and ancestral alleles, despite the fact that the majority of mosaic individuals were heterozygous for this distal marker (*Table 3.1*). The different D4Z4 repeat length distributions of 4qA and 4qB alleles and the observed LD between the proximal *Pvu*II RFLP and the 4qB allele is suggestive of a strong suppression of recombination between 4qA and 4qB alleles. This suppression of recombination between 4qA and 4qB appears not to be sustained throughout the entire repeat, on the basis of the observation that internal D4Z4 units are equally polymorphic for the *Pvu*II restriction site on 4qA and 4qB chromosomes. This latter may be explained by ongoing or incomplete concerted evolution, as has been described for the RNU2 locus [18]. Apparently, a rare interchromosomal gene conversion copied *Pvu*II from 4qB

to 4qA and subsequently spread over this repeat by interallelic rearrangements. It is possible that the proximal D4Z4 unit has escaped this homogenisation until now (*Figure 3.1.C*), which is suggestive of a 3' polarity of D4Z4 rearrangements, as described for meiotic minisatellite rearrangements [11].

3.5.2 D4Z4 rearrangements by gene conversion without interchromatid crossover

The results of the present study strongly suggest that most mitotic D4Z4 rearrangements occur via an interchromatid gene conversion mechanism without crossover in which the donor allele remains unchanged, since the majority of mosaic patients with FSHD carry two distinct cell populations: one that encompasses the parental-sized D4Z4 alleles prior to rearrangement and one that encompasses the *de novo* disease-causing D4Z4 allele (for an example, see *Figure 3.2*).

Gene conversions not associated with crossover can be explained by synthesis-dependent strand annealing (SDSA) models, as proposed for recombination in yeast [28]. In these models, a 3' end of a resected double-strand break (DSB) invades the donor template and primes DNA synthesis. Next, the newly synthesized DNA strands are unwound from the template and returned to the broken molecule, allowing them to anneal to each other. Out-of-frame annealing of the newly synthesized strand can lead to repeat contractions or expansion, while the donor template remains unchanged. Alternatively, intrachromatid looping-related or single-strand annealing rearrangements could also explain our observations. However, these mechanisms can give rise only to D4Z4 contractions, which cannot explain the rapid expansion of D4Z4 repeats in the hominoid lineage [9] or our observation of D4Z4 expansions in the population (RJLF Lemmers, unpublished data).

3.5.3 D4Z4 rearrangements by gene conversion with interchromatid crossover

The presence of two rearranged D4Z4 repeats in 3 of 11 mosaic families with FSHD is indicative of the relatively frequent occurrence of gene conversion with interchromatid crossover (as shown in *Figure 3.3*). Gene conversion with crossover is commonly observed in only a small proportion of rearrangements [36]. It is possible that the telomeric localization of D4Z4 repeats is underlying the increased crossover rate. In family 36, a mitotic D4Z4 rearrangement of the maternal 100 kb or 240 kb allele resulted in alleles of 384 kb and 16 kb, which are observed in almost equal proportions in the patient's PBLs. The patient with *de novo* FSHD from family 1 has three PBL cell populations, one of which carries the original allele (37%), whereas the other two contain contracted and expanded alleles of the same chromosome in 26% and 37% of cells, respectively. Regardless of the exact mechanism, to our knowledge this is the first observation of a *de novo* expansion of the D4Z4 repeat. In family Rf120, the original 100 kb allele is mitotically reduced to a 55 kb allele and an FSHD-sized allele.

Gene conversions that allow for crossovers are often explained by a DSB repair model [37] but can also be explained by a SDSA model that includes the possibility of crossover [36]. In both models, DSB formation is followed by 5'-to-3' resection, leaving two recombinogenic 3' ends that can invade the donor template. In both models, the resolution is postulated to occur through cutting and resolving of two Holliday junctions, which may result in either crossover or no crossover [10]. The DSB repair model was initially proposed to explain gene conversions in yeast that were accompanied by crossover in half of the cases [30]. When subsequent studies showed much lower crossover rates, other models were proposed that did not require Holliday junctions. However, the suppression of crossover during the resolution of these junctions can also explain the overrepresentation of gene conversion without crossover [42]. On the basis of our findings, we propose that mitotic D4Z4 rearrangements occur preferentially via a SDSA model that allows for crossovers.

3.5.4 Timing of D4Z4 rearrangements

The coexistence of mosaicism for D4Z4 in PBLs, germline and fibroblast cells of patients with FSHD already indicated that mitotic D4Z4 rearrangements occur early during embryogenesis. More specifically, the absence of the mosaic parental-sized repeat in family 36 might suggest that the rearrangement had occurred at the one-cell stage, before the first embryonic cell division. Also, all other somatic rearrangements most probably occurred during the first few zygotic divisions after fertilization, resulting in a proportion of affected cells that depended on the timing of the rearrangement and on stochastic events related to which cells contribute to the embryo proper. Mitotic D4Z4 rearrangements that occur at a later stage of development will generally result in *de novo* mosaic alleles in <25% of the cells, as detected in asymptomatic carriers of the FSHD allele [22].

Different mechanisms could underlie this early occurrence of D4Z4 rearrangements. During the first replication steps after fertilisation, paternal and maternal genomes still display some characteristics of their progeny cells [24, 25]. It is tempting to speculate that chromatin conformational changes during the first rounds of cell division may result in DNA strand breaks that initiate these mitotic D4Z4 rearrangements.

3.6 Acknowledgements

We dedicate this work to the memory of Lodewijk Sandkuijl, who unexpectedly passed away during preparation of this manuscript. The authors thank all patients and relatives for their cooperation. We thank Dr. P de Knijff for helping with the haplotype analysis and Dr. GJB van

Ommen and Dr. LH Mullenders for critical reading and useful comments. FSHD research is made possible by the Prinses Beatrix Fonds, the Muscular Dystrophy Association USA, the FSH Society, the Stichting FSHD, the Shaw Family and the National Institutes of Health (NIH).

3.7 Electronic database information

Accession numbers and URL for data presented herein are as follows:

Genbank, <http://www.ncbi.nlm.nih.gov/GenBank/> (for the D4Z4 allele [accession number AF117653], distal 4qB sequence [accession number AF017466], and distal 4qA sequence [accession numbers U74496 and U74497]).

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for FSHD1A).

3.8 References

1. Brouwer OF, Padberg GW, Ruys CJ, Brand R, Laat de JA, and Grote JJ (1991) Hearing loss in facioscapulohumeral muscular dystrophy. *Neurology* 41(12): 1878-1881.
2. Deidda G, Cacurri S, Piazzo N, and Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet* 33(5): 361-365.
3. 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.
4. 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.
5. Funakoshi M, Goto K, and Arahata K (1998) Epilepsy and mental retardation in a subset of early onset 4q35- facioscapulohumeral muscular dystrophy. *Neurology* 50(6): 1791-1794.
6. Gabriels J, Beckers MC, Ding H, Vriese de A, Plaisance S, Maarel van der SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al.* (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 236(1): 25-32.
7. Galluzzi G, Deidda G, Cacurri S, Colantoni L, Piazzo N, Vigneti E, Ricci E, Servidei S, Merico B, Pachi A, *et al.* (1999) Molecular analysis of 4q35 rearrangements in facioscapulohumeral muscular dystrophy (FSHD): application to family studies for a correct genetic advice and a reliable prenatal diagnosis of the disease. *Neuromusc Disord* 9(3): 190-198.

8. 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.
9. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, Deutekom van JC, Francis F, Sharpe PT, Hofker M, *et al.* (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3(8): 1287-1295.
10. Holliday R (1964) A mechanism for gene conversion in fungi. *Genet Res* 5: 282-304.
11. Jeffreys AJ, Neil DL, and Neumann R (1998) Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J* 17(14): 4147-4157.
12. Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, and Armour JAL (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* 6: 136-145.
13. Johnson RD and Jasin M (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* 19(13): 3398-3407.
14. Köhler J, Rupilius B, Otto M, Bathke K, and Koch MC (1996) Germline mosaicism in 4q35 facioscapulohumeral muscular dystrophy (FSHD1A) occurring predominantly in oogenesis. *Hum Genet* 98(4): 485-490.
15. Lemmers RJLF, Kievit de P, Geel van M, Wielen van der MJ, Bakker E, Padberg GW, Frants RR, and Maarel van der SM (2001) Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. *Ann Neurol* 50(6): 816-819.
16. Lemmers RJLF, Kievit de P, Sandkuijl L, 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.
17. Lemmers RJLF, Wielen van der MJ, Bakker E, Padberg GW, Frants RR, and Maarel van der SM (2004) Somatic mosaicism in FSHD often goes undetected. *Ann Neurol* 55(6): 845-850.
18. Liao D (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 64(1): 24-30.
19. Lunt PW (1998) 44th ENMC International Workshop: Facioscapulohumeral Muscular Dystrophy: Molecular Studies 19-21 July 1996, Naarden, The Netherlands. *Neuromusc Disord* 8(2): 126-130.
20. 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.
21. Maarel van der SM, Deidda G, Lemmers RJ, 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.
22. 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.
23. 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).
24. Mayer W, Niveleau A, Walter J, Fundele R, and Haaf T (2000) Demethylation of the zygotic paternal genome. *Nature* 403(6769): 501-502.
25. McLay DW and Clarke HJ (2003) Remodelling the paternal chromatin at fertilization in mammals. *Reproduction* 125(5): 625-633.

26. Miller SA, Dykes DD, and Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3): 1215.
27. Miura K, Kumagai T, Matsumoto A, Iriyama E, Watanabe K, Goto K, and Arahata K (1998) Two cases of chromosome 4q35-linked early onset facioscapulohumeral muscular dystrophy with mental retardation and epilepsy. *Neuropediatrics* 29(5): 239-241.
28. Nassif N, Penney J, Pal S, Engels WR, and Gloor GB (1994) Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol Cell Biol* 14(3): 1613-1625.
29. Okada T, Gondo Y, Goto J, Kanazawa I, Hadano S, and Ikeda JE (2002) Unstable transmission of the RS447 human megasatellite tandem repetitive sequence that contains the USP17 deubiquitinating enzyme gene. *Hum Genet* 110(4): 302-313.
30. Orr-Weaver TL and Szostak JW (1983) Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc Natl Acad Sci USA* 80(14): 4417-4421.
31. 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.
32. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
33. Padberg GW, Brouwer OF, Keizer de RJ, Dijkman G, Wijmenga C, Grote JJ, and Frants RR (1995) On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2: S73-S80.
34. 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.
35. 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.
36. Paques F, Leung WY, and Haber JE (1998) Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol Cell Biol* 18(4): 2045-2054.
37. Szostak JW, Orr-Weaver TL, Rothstein RJ, and Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* 33(1): 25-35.
38. Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, and Figlewicz D (1996) Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39(6): 744-748.
39. Upadhyaya M, Maynard J, Osborn M, Jardine P, Harper PS, and Lunt P (1995) Germinal mosaicism in facioscapulohumeral muscular dystrophy (FSHD). *Muscle Nerve* 2: S45-S49.
40. 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.
41. Wright TJ, Wijmenga C, Clark LN, Frants RR, Williamson R, and Hewitt JE (1993) Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. *Hum Mol Genet* 2(10): 1673-1678.
42. Wu L and Hickson ID (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426(6968): 870-874.
43. 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.
44. Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campioto S, Cerqueira A, Wijmenga C, Padberg G, and Frants R (1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. *Am J Hum Genet* 56(1): 99-105.

4

De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10

*SM van der Maarel*¹, *G Deidda*², *RJFL Lemmers*¹,
*PGM van Overveld*¹, *M van der Wielen*¹, *JE Hewitt*³,
*LA Sandkuijl*¹, *B Bakker*¹, *G-JB Ommen*¹, *GW*
*Padberg*⁴ and *RR Frants*¹

Originally published in:
American Journal of Human Genetics
(2000) 66: 26-35

- ¹ Department of Human Genetics, Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
- ² Institute of Cell Biology, National Research Council, Rome, Italy.
- ³ Institute of Genetics, Queen's Medical Center, University of Nottingham, Nottingham, UK.
- ⁴ Department of Neurology, University Medical Center St. Radboud, Nijmegen, The Netherlands.

4.1 Abstract

Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) is caused by a deletion of most copies of the 3.3 kb subtelomeric D4Z4 repeat array on chromosome 4q. The molecular mechanisms behind the deletion and the high proportion of new mutations have remained elusive. We surveyed 35 *de novo* FSHD families and found somatic mosaicism in 40% of cases, in either the patient or an asymptomatic parent. Mosaic males were typically affected; mosaic females were more often the unaffected parent of a non-mosaic *de novo* patient. A genotypic-severity score, composed of the residual repeat size and the degree of mosaicism, yields a consistent relationship with severity and age at onset of disease. Mosaic females had a higher proportion of somatic mosaicism than did mosaic males. The repeat deletion is significantly enhanced by supernumerary homologous repeat arrays. In 10% of normal chromosomes, 4-type repeat arrays are present on chromosome 10. In mosaic individuals, 4-type repeats on chromosome 10 are almost five times more frequent. The reverse configuration, also 10% in normal chromosomes, was not found, indicating that mutations may arise from transchromosomal interaction, to which the increase in 4-type repeat clusters is a predisposing factor. The somatic mosaicism suggests a mainly mitotic origin; mitotic interchromosomal gene conversion or translocation between fully homologous 4-type repeat arrays may be a major mechanism for FSHD mutations.

4.2 Introduction

Facioscapulohumeral muscular dystrophy (FSHD [MIM 158900]) is a neuromuscular disorder with an autosomal dominant pattern of inheritance. Clinically, the disease is characterised mainly by a progressive wasting of the facial, shoulder, and upper arm muscles, and it displays substantial inter- and intrafamilial variation. FSHD has an incidence of 1 in 20000 in the European population, and there is a relatively high proportion of new mutations (10-30%) [20, 23, 24, 27, 30].

All patients with a confirmed diagnosis and for who detailed molecular studies have been performed carry a chromosomal rearrangement within the subtelomere of chromosome 4q (4q35). This subtelomere is composed mainly of a polymorphic repeat structure consisting of 3.3 kb repeated elements (D4Z4). The number of repeat units varies from 10 to >100 in the population, and, in FSHD patients, an allele of 1-10 residual units is observed because of the deletion of an integral number of these units [10, 28].

A highly homologous polymorphic repeat array is located near the telomere of chromosome 10q [4, 7], and a specific *BlnI* site within each chromosome 10-derived repeat unit allows discrimination between the arrays [8]. This *BlnI* side-dependent discrimination demonstrated the presence of 10-type repeats on chromosome 4 and, vice versa, 4-type repeats on chromosome 10, suggesting a dynamic exchange between these chromosomes. Of 50 healthy control males studied, 5 out of 100 chromosomes carried a 4-type repeat on chromosome 10, and 5 out of 100 chromosomes carried a 10-type repeat on chromosome 4 (*Figure 4.1*) [9]. Hybrid repeat arrays consisting of clusters of both 4-type and 10-type repeat units are also found. Only short repeat arrays on chromosome 4 cause FSHD, irrespective of the type of repeat units. Small repeat arrays on chromosome 10 are non-pathogenic [6, 17].

To understand the molecular basis of FSHD, the repeat array and adjacent regions have been scrutinised for expressed sequences [12, 16]. Also, the severity and age at onset of the disease have been correlated with the size of the residual repeat array on chromosome 4 [22, 27]. However, this has not yet clarified the genetic mechanism underlying the deletion event, nor have the relatively high mutation frequency and the inter- and intrafamilial variation in the

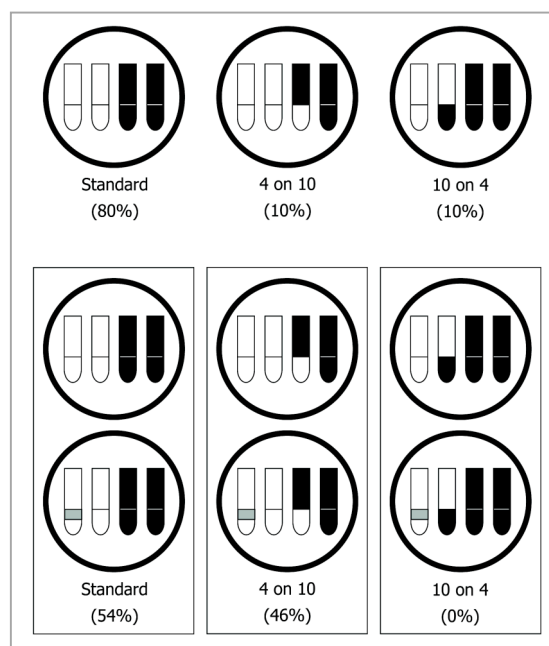


Figure 4.1

Subtelomeric repeat array constitutions on chromosomes 4 and 10 in the Dutch control and FSHD population. Chromosomes 4 are white, whereas chromosomes 10 are blackened.

Top: In the control population, 80% of the individuals carry a standard configuration, with 4-type repeats on chromosome 4 and 10-type repeats on chromosome 10. In 10%, a 4-type repeat is also present on one of the chromosomes 10 (4 on 10), and 10% carry a 10-type repeat on chromosome 4 (10 on 4).

Bottom: The repeat array constitutions of mosaic individuals from *de novo* FSHD families. The deletion is indicated by a gray bar. These individuals carry two cell populations indicated within a box. In the original population, no FSHD-associated rearrangement is present, whereas, in the other population, a deletion has occurred on chromosome 4. In mosaic individuals, 54% carry a standard allele configuration, and 46% carry a “4 on 10” allele configuration.

clinical expression of the disease been explained. To elucidate the deletion process and the clinical variability of the disease, we have focussed our attention on the repeat constitutions of *de novo* patients and their parents. The high degree of somatic mosaicism described here implies that the deletion is mainly mitotic. Furthermore, we find a distinct and yet unexplained relationship between sex and affection status in mosaic carriers. Finally, we find that the numerical excess of 4-type repeats on chromosome 10 is a significant if not the major predisposing factor for the occurrence of the FSHD-type deletion.

4.3 Subjects and methods

4.3.1 Patients and controls

After informed consent, DNA isolated from peripheral blood lymphocytes (PBLs) of 35 sporadic FSHD patients and their parents was obtained via one of the Dutch Neuromuscular Centers. All patients had clinical diagnosis of FSHD. In all cases, diagnosis was confirmed by the presence of a short (<35 kb) D4Z4 repeat array. The repeat array constitutions of 50 healthy control males have been described elsewhere [9].

4.3.2 Clinical manifestations

The age at onset of the various stages of FSHD was established in 12 mosaic individuals. Clinical severity scores were based on the age at onset of the initial clinical symptoms: 0 (no signs, no symptoms), 1 (facial weakness, no symptoms), 2 (onset of symptoms at age ≥ 20 years), 3 (onset of symptoms at age 15-19 years), 4 (onset of symptoms at age 10-14 years), and 5 (onset of symptoms at age <10 years).

4.3.3 PFGE

Five micrograms of DNA was double digested with *EcoRI/HindIII* (E/H) or double digested with *EcoRI/BlnI* (E/B) according to the manufacturer's instructions. DNA was separated by PFGE. In brief, DNA was loaded on a 0.8% agarose gel (Boehringer MP agarose) in 0.5× TBE and separated by PFGE for 20 h at 8.5 V/cm [9]. After blotting to a Nytran⁺ membrane (Schleicher & Schuell), the DNA was hybridised sequentially with p13E-11 (D4F104S1) and 9B6A (D4Z4) [28] as described elsewhere [17]. After exposure to phosphorimager screens, repeat arrays were assigned to their chromosomal location on the basis of their *BlnI* sensitivity. Somatic mosaicism was defined as a fifth fragment hybridising with p13E-11 in one individual. The proportion of cells carrying the mosaic alleles was estimated by comparison of the observed signal intensity

Table 4.1 Repeat array constitution of 35 Dutch *de novo* FSHD families.

Family	Sex	D4Z4 repeat length (units)	Complete allele information	Presence of mosaicism in			Inheritance	Allele constitution in mosaic individuals
				father	mother	patient		
1	male	2	+			+		4 on 10
2	male	5					maternal	
3	male	4					paternal	
4	female	2		+			paternal	standard
5	male	2				+		standard
6	male	1	+			+		4 on 10 ^a
7	male	4				+		standard
8	male	3	+				paternal	
9	male	1	+		+		maternal	4 on 10 ^b
10	female	3			+		maternal	
11	male	5	+		+		maternal	standard
12	female	4	+			+		4 on 10
13	female	4				+		standard
14	female	4					maternal	
15	female	3					?	
16	female	4					?	
17	female	6					?	
18	female	3	+				maternal	
19	female	4	+				maternal	
20	female	4	+				maternal	
21	male	4	+				paternal	
22	male	3	+				?	
23 ^c	male	3	+				paternal	
24	male	3	+			+		standard
25	female	4	+				maternal	
26	male	2	+			+		standard
27	female	4	+				?	
28	female	6	+				paternal	
29	male	3	+				maternal	
30	male	2	+				paternal	
31	female	6	+				maternal	
32	female	2	+				paternal	
33	male	4	+				maternal	
34	male	4/5 ^d	+	+			paternal	4 on 10
35	male	2/6 ^d				+		

A plus sign (+) indicates presence, and a question mark (?) indicates unknown. a: This individual carries a hybrid allele on chromosome 10, consisting of both 4-type and 10-type repeat units. b: This individual carries only 4-type repeat arrays (*Figure 4.3.B*). c: In this patient, the partial deletion of the D4Z4 repeat array includes the p13E-11 region. d: These individuals are mosaic for two short repeat arrays.

with the expected intensity. Signal intensities of the various hybridising fragments were obtained with the ImageQuant program (Molecular Dynamics).

4.3.4 Statistical evaluation

The allele constitutions on mosaic and control individuals were evaluated according to Fisher's exact test. D4Z4 repeat array lengths in mosaic and control individuals were compared by means of the Mann-Whitney *U* test (one sided).

A relationship between age at onset and the proportion of cells carrying the deleted allele was calculated by linear regression analysis. For this purpose, we constructed a composite genotypic-severity scale: (5 – remaining number of repeat units) × the fraction of cells showing

that deletion. The number 5 is the largest repeat array observed in mosaic individuals in this study.

4.4 Results

4.4.1 Allele identification by PFGE

The repeat array constitutions of 35 Dutch sporadic FSHD patients (19 males and 16 females) and their parents were analysed by PFGE (*Table 4.1*). For all but two patients, DNA from parents was studied. In 23 families we could determine the chromosomal origin and the size of each of the four repeat arrays (on chromosomes 4 and 10) in both patients and parents. In the 10 remaining families with incomplete information, the DNA quality of one of the individuals was not sufficient to assign all alleles.

All patients showed one *BlnI*-insensitive repeat array <35 kb (9 units), consistent with FSHD diagnosis [3, 20] (*Table 4.1*). The lengths of these arrays varied from 8 kb (1 unit and flanking sequences) to 25 kb (6 units). One patient inherited a 6 unit repeat array from his clinically unaffected father.

4.4.2 Mosaicism

We observed 14 cases of somatic mosaicism (defined by a fifth band on PFGE; *Tables 4.1* and *4.2* and *Figure 4.1*). In 3 of 23 fully informative families, we observed mosaicism for the disease allele in unaffected parents (2 mothers and 1 father), and in 5 more families the patients were

Table 4.2 Repeat allele information for the 35 FSHD families studied.

Type of FSHD and characteristic	Families (<i>n</i>)	Individuals (<i>n</i>)
Familial	1	
<i>De novo</i>	34	
Mosaicism:		14 ^a
male carriers		2
female carriers		3
male affecteds		7
female affecteds		2
Repeat array constitutions in mosaic individuals: ^a		
standard		7
4 on 10		6
10 on 4		0

a: Thirteen individuals with complete information of both alleles.

mosaic (4 males and 1 female; *Tables 4.2* and *4.3*). In the remaining 12 families, for which we did not have full information on all repeat lengths, we observed mosaicism in one father, one mother, and four patients (three males and one female; *Tables 4.2* and *4.3*).

The ImageQuant program was used to estimate the percentage of the cell population carrying the deleted allele in mosaic individuals; this varied from 15 to 95% (*Table 4.3*). Where possible, the parental origin of the mosaic allele was determined, as well as the original repeat array size and the size of the deletion (*Table 4.3*). In 8 of 11 mosaic individuals, the smallest D4Z4 allele was reduced to an FSHD-sized repeat array. A genotypic-severity score composed of the size of the residual repeat array and the proportion of cells carrying this array could be established in 11 mosaic individuals (*Table 4.3*).

The age at onset of the clinical manifestations was used for a clinical severity score for the mosaic individuals (*Table 4.3*). A relationship could be established between the age at onset of the disease, on the one hand, and the fraction of cells carrying the deleted allele and the residual repeat array lengths (*Figure 4.2*), on the other hand. Individuals with a large proportion of cells carrying the disease allele have an earlier age at onset of the disease ($P \approx 0.02$).

4.4.3 Repeat array inheritance

To determine whether the presence of either 4-type repeat arrays on chromosome 10 or 10-type repeat arrays on chromosome 4 might play a role in the deletion mechanism, we analysed the repeat array constitutions in those mosaic individuals in which we could score all alleles (13 of 14). Of these 13, 6 carried one or more 4-type repeat arrays on chromosomes 10 (*Figure 4.1* and *Table 4.3*). Typically, in these cases, PFGE analysis revealed the presence of five p13E-11 hybridising fragments: three *BlnI*-insensitive (4-type) repeat arrays >35 kb; one *BlnI*-sensitive (10-type) repeat array; and the mosaic, truncated, disease-associated repeat array <35 kb (*Figures 4.1* and *4.3.A* and *Table 4.2*). In one of these six individuals, the “non-standard” repeat array consisted of both 4-type and 10-type repeat units (data not shown). One mosaic individual even carried only 4-type repeat arrays (*Figure 4.3.B*). According to the Fisher’s exact test, the frequency of 4-type repeats on chromosomes 10 in the mosaic individuals (5 of 13 [40%]) differs significantly from the Dutch control population ($P = 0.0125$), previously determined at 10% [9]. In contrast, 0 of 13 mosaic individuals showed the reverse configuration of three or more *BlnI*-sensitive arrays indicative of the presence of 10-type repeat units on chromosome 4.

Table 4.3 Mosaic individuals in *de novo* FSHD families.

Family	Sex	Affection status	D4Z4 repeat length (units)	Original allele		Disease allele		Rearranged allele ^a	Δ^b (kb)	Inheritance	Allele constitution	Age at onset of weakness ^c (years)			Severity score	
				length (kb)	%	length (kb)	%					shoulder	foot extensor	pelvic	clinical	genotype
1	male	affected	2	70	50	12	50	small	58	paternal	4 on 10	15	19	40	3	1.50
5	male	affected	2	96	80	12	20	small	84	paternal	standard	18	23	-	3	0.60
6	male	affected	1	230	66	8	34	large	222	paternal	4 on 10	4	4	4	5	1.36
7	male	affected	4	65	73	18	27	small	47	maternal	standard	18	?	34	3	0.27
12	female	affected	4	80	5	18	95	large	62	maternal	4 on 10	12	7	12	5	0.95
13	female	affected	4	62	35	17	65	small	45	?	standard	14	?	?	4	0.65
24	male	affected	3	40	55	15	45	small	25	maternal	standard	14	30	30	4	0.90
26	male	affected	2	130	60	12	40	large	118	maternal	standard	13	11	18	4	1.20
35	male	affected	2/6								4 on 10	73	48	72	2	
4	male	carrier	2	40	67	12	33	small	28		standard	?	?	?		
34	male	carrier	4/5	22	60	19	40	small	3		4 on 10	-	-	-	0	0.00
9	female	carrier	1	110	85	10	15		100		4 on both 10	- ^d	-	-	1	0.60
10	female	carrier	3									?	?	?		
11	female	carrier	4	48	62	21	38	small	27		standard	-	-	-	0	0.38

a: Smallest or largest chromosome 4 allele rearranged. b: Size of the deletion. c: Minus sign (-) indicates that the individual is not affected; a question mark (?) indicates onset unknown. d: This individual has a mild facial weakness.

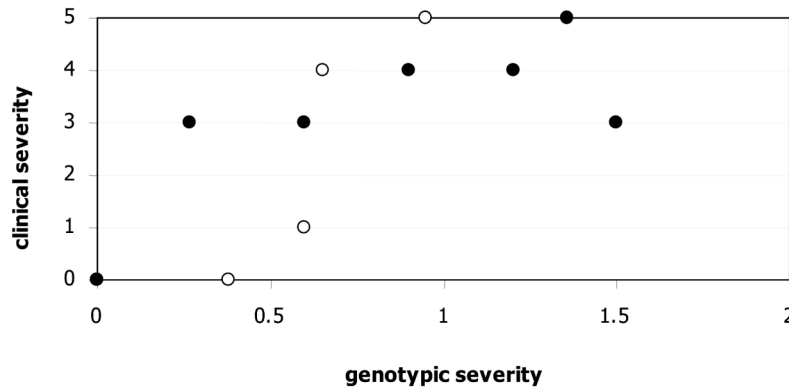


Figure 4.2

Relationship between clinical and genotypic severity in mosaic individuals. The clinical severity is on the y -axis, whereas the genotypic severity, calculated by $[(5 - \text{number of remaining repeat units}) \times \text{fraction of cells with that deletion}]$ is on the x -axis. Males are represented by blackened dots, whereas females are represented by open dots.

4.5 Discussion

Although the deletion of an integral proportion of D4Z4 repeat arrays has been well established as the causal mutation of FSHD, little is known about the mechanism by which the deletion arises. We recently found that highly homologous repeat units on chromosomes 4 and 10 may interact, resulting in exchanged repeat units on both chromosomes in 20% of the population [9, 17]. Recently, we could confirm this finding in a larger sample of both males and females (SM van der Maarel, unpublished data). To obtain more insight into the deletion mechanism and the putative role of interchromosomal repeat interference, we studied the repeat array constitution of chromosomes 4 and 10 in 35 *de novo* FSHD families ascertained via the Dutch neuromuscular centers. The results highlight several aspects of FSHD, not previously reported, that are relevant to the basic and clinical insight in FSHD aetiology.

4.5.1 Mitosis or meiosis

We identified somatic mosaicism in 40% of our *de novo* families. Mosaicism has previously been reported for several human X-linked and autosomal dominant diseases, including Duchenne muscular dystrophy and haemophilia A [2, 5, 25]. Also, for several tumor-prone syndromes, such as neurofibromatosis type 2 and retinoblastoma, mosaicism has been reported [11, 26]. However, the high proportion of mosaicism identified in FSHD, as reported here, is unprecedented. Considering that we had full allele information on only 23 of 35 *de novo* families and that PFGE reveals only a significant degree of mosaicism, we estimate that $\geq 40\%$,

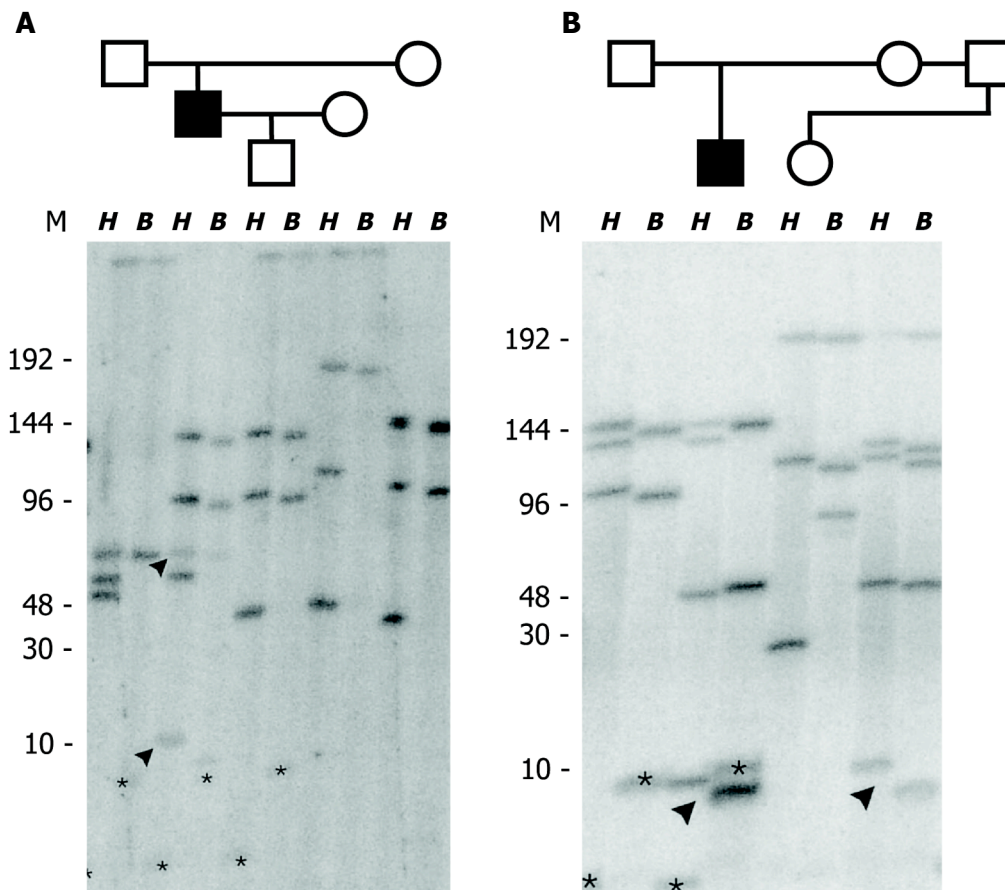


Figure 4.3

(A) Somatic mosaicism in a *de novo* FSHD patient (FSHD1). DNA, digested by *EcoRI/HindIII* (“H”) or *EcoRI/BlnI* (“B”) and separated by PFGE, was visualised with probe p13E-11. The patient inherited a 65 kb 4-type repeat array from his father, which is reduced by 50% of his PBLs to the FSHD range of 10 kb (both alleles have a reduced intensity [arrows]). Note that this patient inherited from his mother two 4-type repeat arrays and, therefore, has a “4 on 10” repeat array configuration. “Y” alleles are marked with an asterisk (*). The son of this patient inherited both grandmaternal alleles and is therefore not affected.

(B) Somatic mosaicism in a female carrier of *de novo* FSHD kindred (FSHD9). DNA, digested by *EcoRI/HindIII* (“H”) or *EcoRI/BlnI* (“B”) and separated by PFGE, was visualised with probe p13E-11. The patient carries a short *BlnI*-insensitive repeat array of 10 kb (arrow), indicating FSHD. This allele (arrow) is also weakly present in the mother, who carries on chromosomes 4 and 10 four-type repeat arrays, identified on the basis of their *BlnI* insensitivity. The “Y” chromosomal co-hybridising fragment is marked with an asterisk (*).

and possibly much more, of *de novo* families carry somatic mosaicism for the disease allele. Therefore, we propose that the FSHD rearrangement is predominantly mitotic.

4.5.2 Interchromosomal repeat interference

In 6 (45%) of 13 of the mosaic individuals in whom we could score all alleles, we observed one or more 4-type repeat arrays on chromosome 10. In the Dutch population, this type of repeat array configuration is present in only 10% of the individuals [9]. The reverse configuration, equally present in 10% of controls, was never found in association with mosaicism. This provides strong support for the causal involvement of the extra 4-type repeat in the partial

deletion of one of them and, in the case of chromosome 4, leading to FSHD. Studies of minisatellite repeat structures and a similar megasatellite repeat (*RNU2*) in the human genome indicate that intrachromosomal recombination, such as unequal sister-chromatid exchange (USCE) or intrachromatid gene conversion, dominates over interchromosomal recombination processes [13, 19]. A high rate of intrachromosomal recombination effects homogenisation and hence concerted evolution of repetitive multigene families, which is thought to maintain the integrity of each repeat gene [18]. In FSHD, a similar process might well play a role in the partial deletion of the D4Z4 repeat array rearrangements. Since USCE should result in a reciprocal allele in the mosaics, which we have never observed, these deletions may arise by gene conversion.

On the other hand, the numerical excess of 4-type repeats on chromosome 10 in mosaic individuals indicates that supernumerary 4-type repeats may physically facilitate the partial deletion on either chromosome 4 or 10 (which would go undetected). In this model, the deletion arises independent of its inter- or intrachromosomal nature. Since recombination depends highly on homology between both alleles [15], normal 10-type repeat units may have diverged sufficiently to suppress their contribution to D4Z4 repeat deletions. For rDNA repeat arrays, recombination between non-homologous chromosomes has been reported and was attributed to the close proximity of these repeat alleles in the nucleolus [1].

4.5.3 Allele sizes

In most of the mosaic individuals, the original allele from which the disease allele has arisen could be identified by the lower hybridisation intensity. Strikingly, in most of the cases (8 of 11), it was the shortest D4Z4 repeat allele that was reduced to an FSHD-sized repeat array. This finding may be due to ascertainment bias wherein mosaicism of relatively large alleles remains unnoticed, since they will not be reduced to arrays smaller than 35 kb. Indeed, the size distribution of the original alleles was significantly smaller than the size distribution of 78 standard chromosome 4 alleles in the control population ($P < 0.05$). This may indicate that deletions from relatively large alleles may in general have no pathological consequences.

As already observed by Zatz *et al.* [29], the FSHD allele in asymptomatic individuals is relatively short. The average repeat length in mosaic individuals in this study is 3.1 units, whereas non-mosaic patients carry, on average, 3.9 units. This may be explained in part by the association between the residual repeat size and severity and age at onset of the disease. In general, small alleles result in a more severe phenotype [22, 27], and it may well be that, in mosaic individuals, the tolerance for short alleles is smaller than that for larger FSHD alleles. In this scenario, high proportions of cells carrying larger FSHD alleles may be required, to elicit a disease phenotype. The resulting low proportion of original alleles may go undetected by PFGE.

Alternatively, in addition to the mosaic rearrangement described here, which exhibits relatively short arrays, and undefined mutational mechanism may exist that results in relatively larger FSHD arrays. If such an alternative exists, the high proportion of severely affected patients among *de novo* cases may be explained by a mechanistic difference rather than by a clinical ascertainment bias.

4.5.4 Sex and affection status

Somatic mosaicism has been observed in 15-20% of the unaffected parents of *de novo* FSHD patients. In these studies, a female predominance of mosaic carriers (15 females : 6 males) was reported and was then attributed to a higher mutation rate in females than in males [14, 20, 29]. In agreement, we found parental mosaicism in 14% of our *de novo* families (3 females and 2 males).

In addition to mosaicism in parents, we found a high frequency (26%) of somatic mosaicism in *de novo* FSHD patients who belong to different families as the mosaic carrier parents. In these mosaic patients we find, in contrast, an excess of mosaic males over mosaic females (7 of 2). The proportion of cells carrying the deletion allele varied from 15 to 19% on the basis of the signal intensities of the different alleles. Typically, in mosaic female patients, the proportion of the deleted allele was higher than in mosaic male patients (*Table 4.3*). Combining the female excess among unaffected mosaic carrier parents with the male mosaic excess among the patients themselves, we propose that the female predominance of mosaic asymptomatic carriers is not due to a higher mutation rate in females, but rather to a higher clinical tolerance for mosaic disease alleles in females compared with males. Consistent with this proposal, males are more severely affected than females [21, 23, 24, 29].

Although the extent of somatic mosaicism depends on the timing of the deletion event and on tissue-specific selection, we established a relationship between the residual repeat size and the proportion of cells carrying this disease allele, on one hand, and the severity and age at onset on the other ($P \approx 0.02$). This suggests that the mutation may already occur early in embryogenesis, before divergence of muscle and lymphocyte lineages. Nevertheless, it will be important to analyse the mosaicism in muscles of mosaic patients and parents.

4.5.5 Implications for FSHD diagnosis

For autosomal dominant diseases, recurrence risks in parents of *de novo* patients are considered very low, whereas disease carriers have 50% probability of having affected offspring. Since we have detected somatic mosaicism in $\geq 40\%$ of the *de novo* families, this is not necessarily true for *de novo* FSHD families. According to *Table 4.3*, since asymptomatic mosaic carriers may have as much as 40% of cells carrying the disease allele, such carriers may have a $\leq 20\%$ risk that their

offspring will be affected. In contrast, the mosaic FSHD patient in family 5 may have as little as 10% risk of having affected offspring. Detailed analysis of somatic and germline mosaicism in *de novo* FSHD families will be required, to obtain more accurate figures for genetic counselling.

In conclusion, we have shown (1) that the D4Z4 repeat reduction associated with FSHD arises in ~40% of the *de novo* cases, mitotically, in either parent or patient; (2) that the basic mechanism of the repeat reduction likely involves inter- or, possibly, intrachromosomal gene conversion; (3) that this is facilitated by the presence of D4Z4 repeat arrays on chromosome 10, which thus predisposes for the deletion; (4) that somatic mosaicism occurs in both males and females, but that males are more often affected than females; therefore, the proposed sex difference in the occurrence of somatic mosaicism may in fact be a sex-dependent clinical threshold for the mosaic disease allele; and (5) that a relationship exists between, on one hand, a combination of the residual repeat size and proportion of cells carrying the disease allele and, on the other hand, the age at onset and the severity of the disease. This chromosomal copy number-dependent repeat interference may well turn out to be a basic mechanism for genome instability involving multicopy repetitive elements.

4.6 Acknowledgements

We thank Dr. M de Visser and Dr. H Meijers Heijboer, for clinical evaluation of their patients, and Dr. N de Wind, for his constructive comments. This study was funded by the Prinses Beatrix Fonds, The Netherlands Organization for Scientific Research (NWO), The Muscular Dystrophy Association (USA), The Dutch FSHD Foundation, and the FSH Society, Inc.

4.7 Electronic database information

Accession number and URL for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim> (for FSHD [MIM 158900])

4.8 References

1. Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, and Zimmer E (1980) Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. *Proc Natl Acad Sci USA* 77(12): 7323-7327.
2. Bakker E, Broeckhoven van C, Bonten EJ, Vooren van de MJ, Veenema H, Hul van W, Ommen van GJ, Vandenberghe A, and Pearson PL (1987) Germline mosaicism and Duchenne muscular dystrophy mutations. *Nature* 329(6139): 554-556.
3. Bakker E, Wielen van der MJ, Voorhoeve E, Ippel PF, Padberg GW, Frants RR, and Wijmenga C (1996) Diagnostic, predictive, and prenatal testing for facioscapulohumeral muscular dystrophy: diagnostic approach for sporadic and familial cases. *J Med Genet* 33(1): 29-35.
4. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, Wielen van der M, Rasmussen K, and Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2: 39-44.
5. Becker J, Schwaab R, Moller-Taube A, Schwaab U, Schmidt W, Brackmann HH, Grimm T, Olek K, and Oldenburg J (1996) Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies. *Am J Hum Genet* 58(4): 657-670.
6. Cacurri S, Piazzo N, Deidda G, Vigneti E, Galluzzi G, Colantoni L, Merico B, Ricci E, and Felicetti L (1998) Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric *KpnI* repeat units implicated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 63: 181-190.
7. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, and Felicetti L (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3: 155-167.
8. Deidda G, Cacurri S, Piazzo N, and Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet* 33(5): 361-365.
9. 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.
10. 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.
11. Evans DG, Wallace AJ, Wu CL, Trueman L, Ramsden RT, and Strachan T (1998) Somatic mosaicism: a common cause of classic disease in tumor-prone syndromes? Lessons from type 2 neurofibromatosis. *Am J Hum Genet* 63(3): 727-736.
12. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, Deutekom van JC, Francis F, Sharpe PT, Hofker M, *et al.* (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3(8): 1287-1295.
13. Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, and Armour JAL (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* 6: 136-145.
14. Köhler J, Rupilius B, Otto M, Bathke K, and Koch MC (1996) Germline mosaicism in 4q35 facioscapulohumeral muscular dystrophy (FSHD1A) occurring predominantly in oogenesis. *Hum Genet* 98(4): 485-490.

15. Lambert S, Saintigny Y, Delacote F, Amiot F, Chaput B, Lecomte M, Huck S, Bertrand P, and Lopez BS (1999) Analysis of intrachromosomal homologous recombination in mammalian cell, using tandem repeat sequences. *Mutat Res* 433(3): 159-168.
16. Lee JH, Goto K, Matsuda C, and Arahata K (1995) Characterization of a tandemly repeated 3.3 kb *KpnI* unit in the facioscapulohumeral muscular dystrophy (FSHD) gene region on chromosome 4q35. *Muscle Nerve* 2: S6-S13.
17. 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, *et al.* (1998) Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet* 7(8): 1207-1214.
18. Liao D (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 64(1): 24-30.
19. 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.
20. Lunt PW (1998) 44th ENMC International Workshop: Facioscapulohumeral Muscular Dystrophy: Molecular Studies 19-21 July 1996, Naarden, The Netherlands. *Neuromusc Disord* 8(2): 126-130.
21. 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.
22. 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.
23. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
24. 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.
25. Passos-Bueno MR, Bakker E, Kneppers AL, Takata RI, Rapaport D, Dunnen den JT, Zatz M, and Ommen van GJ (1992) Different mosaicism frequencies for proximal and distal Duchenne muscular dystrophy (DMD) mutations indicate difference in etiology and recurrence risk. *Am J Hum Genet* 51(5): 1150-1155.
26. Sippel KC, Fraioli RE, Smith GD, Schalkoff ME, Sutherland J, Gallie BL, and Dryja TP (1998) Frequency of somatic and germ-line mosaicism in retinoblastoma: implications for genetic counseling. *Am J Hum Genet* 62(3): 610-619.
27. Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, and Figlewicz D (1996) Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39(6): 744-748.
28. 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.
29. 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.
30. Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campiotto S, Cerqueira A, Wijmenga C, Padberg G, and Frants R (1995) High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. *Am J Hum Genet* 56(1): 99-105.

5

Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy

*PGM van Overveld¹, RJFL Lemmers¹, LA Sandkuijl¹,
L Enthoven², ST Winokur³, F Bakels¹, GW Padberg⁴,
G-JB van Ommen¹, RR Frants¹ and SM van der
Maarel¹*

Originally published in:
Nature Genetics
(2003) 35: 315-317

- ¹ Department of Human Genetics, Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
- ² Division of Medical Pharmacology, Leiden/Amsterdam Center for Drug Research, Leiden University Medical Center, Leiden, The Netherlands.
- ³ Department of Biological Chemistry, College of Medicine, University of California, Irvine CA, USA.
- ⁴ Department of Neurology, University Medical Center St. Radboud, Nijmegen, The Netherlands.

5.1 Abstract

The autosomal dominant myopathy facioscapulohumeral muscular dystrophy (FSHD1, OMIM 158900) is caused by contraction of the D4Z4 repeat array on 4qter. We show that this contraction causes marked hypomethylation of the contracted D4Z4 allele in individuals with FSHD1. Individuals with phenotypic FSHD1, who are clinically identical to FSHD1, but have an unaltered D4Z4, also have hypomethylation of D4Z4. These results strongly suggest that hypomethylation of D4Z4 is a key event in the cascade of epigenetic events causing FSHD1.

5.2

Epigenetic changes of DNA have been investigated as causes of monogenic disorders, tumorigenesis and aging and are suspected to be important for common multifactorial diseases also. Monogenic diseases associated with epigenetic phenomena are caused either by mutations in chromatin remodelling factors [1] or by position effect variegation mechanisms mostly involving regulatory elements [19]. An epigenetic role is also prominent in imprinting disorders [16].

FSHD1, which progressively and variably affects muscles of the face, shoulder and upper arm [15], has such a suspected epigenetic aetiology. We previously mapped FSHD1 to 4qter and showed that it is caused by contraction of the polymorphic D4Z4 repeat array [4, 20] (*Supplementary Figure 5.1*). In healthy individuals, D4Z4 consists of 11-150 units on both chromosomes, whereas individuals with FSHD1 carry one 4q array of 1-10 units. About 5% of individuals with FSHD1 do not have a contraction of D4Z4 and are considered to have phenotypic FSHD [10].

Several observations suggest an epigenetic aetiology in FSHD1. First, the subtelomere of chromosome 10q contains a nearly identical polymorphic D4Z4 repeat, but size reductions of this repeat are not pathogenic. Second, exchanges between the homologous repeat arrays on 4q and 10q are frequently observed, but these exchanged repeats are only pathogenic when the contracted form resides on chromosome four [3]. Finally, D4Z4 contractions on 4qter are necessary but not sufficient to cause FSHD1, because the actual pathogenicity is only associated with one of two alleles (4qA) of the 15-20 kb 4qter moiety located immediately distal to D4Z4 [8].

The presentation of FSHD1 requires that there is a contracted D4Z4 repeat located on chromosome 4q and in *cis* with only one of the two telomeric alleles (4qA). Therefore, it is

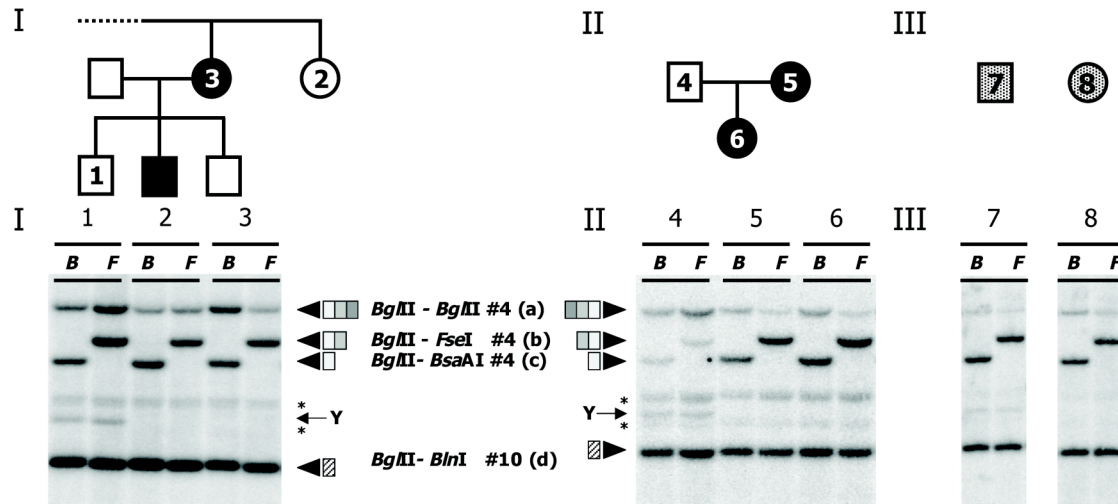


Figure 5.1 Methylation-sensitive digestions after Southern blotting and hybridisation with probe p13E-11. In the left panel, the pedigree of family I with two family members with classical FSHD is shown. Individual 1 is an unaffected monosomic (explained in *Supplementary Figure 5.1*) person with a D4Z4 repeat array of 73 kb. Individuals 2 and 3 are both monosomic and inherited the same pathogenic D4Z4 4qA repeat array of 22 kb from one of their parents. Individual 2 is a minimally affected gene carrier, and individual 3 has FSHD. In the middle panel, the pedigree of family II with phenotypic FSHD is shown. The father (individual 4) is a healthy disomic person. His wife (individual 5) and his daughter (individual 6) are both disomic and have clinical characteristics of FSHD but do not carry a contracted D4Z4 array on chromosome 4. Note the marked reduction of signal intensity of the undigested chromosome 4-derived fragments in individuals with FSHD. In the right panel (III), two unrelated individuals with ICF show pronounced hypomethylation of D4Z4 (individuals 7 and 8). Methylation-sensitive digestions (“B”, *BsaAI*; “F”, *FseI*) for each individual are shown in different lanes beneath the pedigrees, and arrows at the sides of the figures indicate the methylated and unmethylated fragments according to the colour codes used in *Supplementary Figure 5.1*. “Y” indicates cross-hybridisation with the sequence of the Y chromosome in male individuals, and asterisks mark cross-hybridisation with other sequences.

unlikely that contraction of the D4Z4 repeat array itself directly results in loss of all or part of a putative gene mutated in FSHD1. Thus, FSHD1 may be due to an epigenetic phenomenon that causes the transcriptional deregulation of genes close to D4Z4.

To explain the epigenetic basis for FSHD1, we examined the DNA methylation, the most common modification of mammalian DNA, known to be involved in development, X-chromosome inactivation, imprinting and gene silencing [18]. We examined two CpG methylation-sensitive restriction sites (*BsaAI* and *FseI*) in the first (proximal) unit of the D4Z4 repeat array on chromosome 4q35 (see *Supplementary subjects and methods* and *Supplementary Figure 5.1*). First, we determined methylation of the proximal D4Z4 unit for the 4q-type repeat exclusively (see examples in *Figure 5.1*). We then showed by pulsed-field gel electrophoresis that these results were representative for the entire array (*Supplementary Figure 5.2*).

Using this test, we found significant hypomethylation of D4Z4 in individuals with FSHD1 on both sites (*Figure 5.2*). Considering the presence of one wild type and one mutated allele in disomic individuals with FSHD1, hypomethylation was highly significant on both sites ($P < 0.001$).

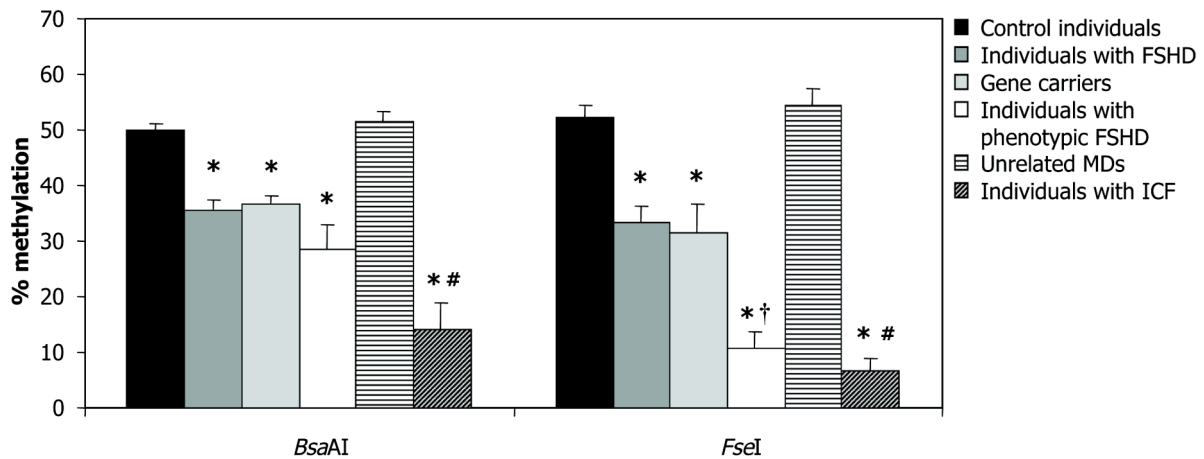


Figure 5.2

Bar diagram of the methylation of D4Z4. The methylation at the *Bsa*AI and *Fse*I methylation-sensitive restriction sites in the proximal D4Z4 repeat unit is shown for control individuals ($n = 79$), individuals with FSHD ($n = 40$), obligate gene carriers ($n = 10$), individuals with phenotypic FSHD (without contraction of D4Z4; $n = 5$), unrelated individuals with non-FSHD muscular dystrophy (MD; $n = 14$) and individuals with ICF ($n = 2$). See *Supplementary subjects and methods* for detailed information on all groups. Data are presented as mean $\pm$ S.E.M. * $P < 0.05$ versus control individuals; † $P < 0.05$ versus individuals with FSHD; # $P < 0.05$ versus individuals with FSHD and obligate gene carriers.

Non-penetrant gene carriers, clinically unaffected individuals carrying the mutated allele, were similarly hypomethylated at their shortened D4Z4 allele ($P < 0.001$; **Figure 5.2**). We found no tissue-specific methylation differences between muscle and blood lymphocytes of three individuals with FSHD1 and three control individuals. Two biopsy samples from clinically affected and unaffected muscle from the same individuals with FSHD1 also show no differences in methylation (data not shown). We also found no effect of age or ageing in lymphocyte DNA from unaffected individuals and controls samples twice with a time interval of several years (data not shown). Thus, D4Z4 methylation seems to be established early and transmitted stably.

In about 5% of affected individuals, FSHD1 shows neither linkage to 4qter nor contraction of D4Z4 and is thought to be caused by defects in other unidentified loci, one of which maps to chromosome 15 [17]. These individuals with phenotypic FSHD1 are clinically indistinguishable from those with 4q-linked FSHD1. Analysis of five individuals with phenotypic FSHD1 showed pronounced hypomethylation for both sites on both chromosomes four ($P < 0.001$), strongly supporting a central role for D4Z4 hypomethylation in the aetiology of FSHD1 (**Figures 5.1** and **5.2** and *Supplementary Figure 5.3*).

To further determine whether D4Z4 hypomethylation is specific to FSHD1, we analysed DNA of 14 unrelated individuals with muscular dystrophy and of individuals with autosomal recessive immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome. ICF syndrome (OMIM 242860) presents with immunodeficiency, facial anomalies, mental retardation

and developmental delay and is caused by mutations in the gene DNA methyltransferase 3B (*DNMT3B*) [21]. Individuals with ICF have hypomethylation at D4Z4 [7].

Unrelated individuals with muscular dystrophy had methylation comparable to control individuals, but individuals with ICF had even lower methylation than individuals with FSHD1 ($P < 0.05$ for both sites; *Figures 5.1* and *5.2*). In contrast to FSHD1, in which hypomethylation is restricted to the disease chromosome, hypomethylation was detected on chromosomes 4q and 10q in ICF (*FseI*, $P < 0.01$; *BsaAI*, $P = 0.07$ (a trend)).

Recent findings suggest that reduction of a specific repressor complex, associated with D4Z4 contraction, causes improper activation of 4q35 genes in a distance-dependent manner [5]. In contrast, a recent study based on histone H4 acetylation levels suggests that 4qter has properties of unexpressed euchromatin and that a differential long-distance looping mechanism seems to be the probable mechanism [6]. But neither model, spreading or looping, explains the etiological link between 4q allele-specificity of D4Z4 contraction and transcriptional upregulation. Our results showing chromosome-specific hypomethylation may provide this missing link between DNA changes and transcriptional derepression in FSHD1. Whatever the exact nature of this mechanism, our results showing D4Z4 hypomethylation in individuals with 4q-linked and phenotypic FSHD1 strongly supports a central role of D4Z4 demethylation in the pathogenic pathway of FSHD1.

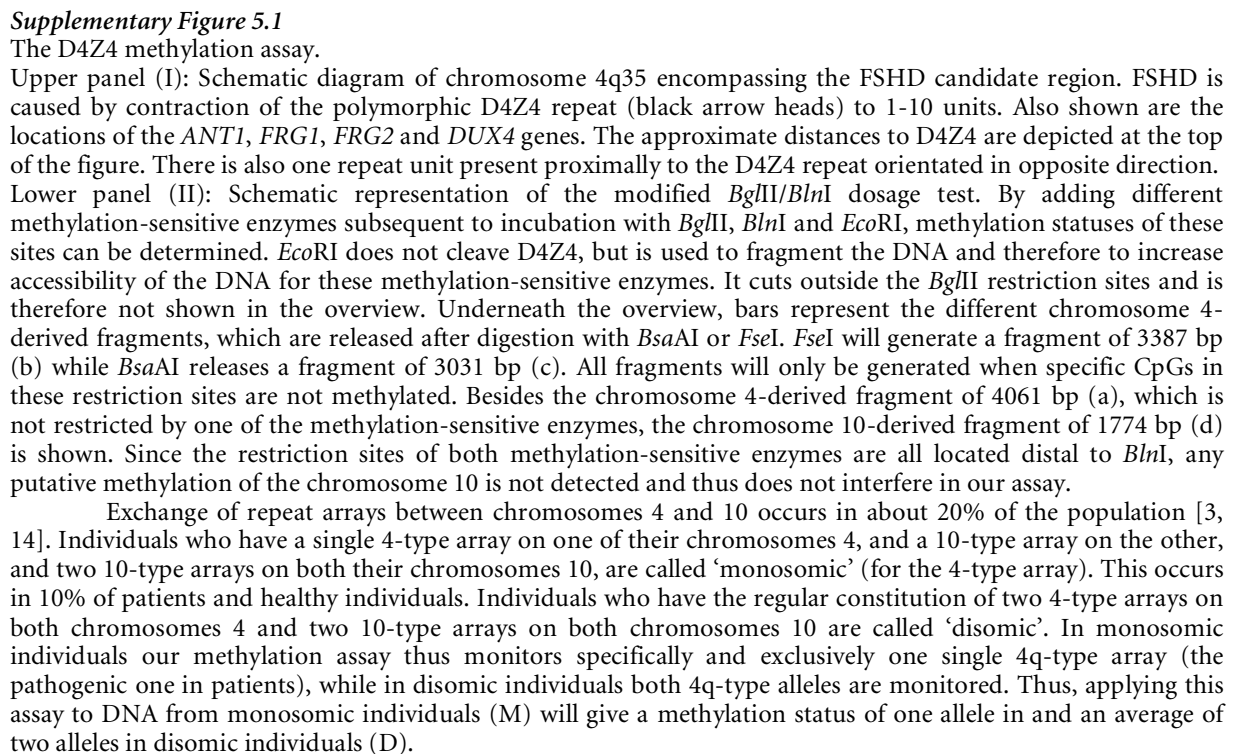
5.3 Supplementary subjects and methods

5.3.1 Individuals

All individuals included were previously analysed after informed consent for their D4Z4 allele sizes and constitution of repeat arrays ([11, 12, 14] and unpublished results).

Genomic DNA was isolated from PBLs of 79 control individuals, both family members and unrelated individuals (19 monosomic and 60 disomic), 40 FSHD patients (5 monosomic and 35 disomic), 5 disomic phenotypic FSHD patients, 10 non-penetrant obligate gene carriers (2 monosomic and 8 disomic), 14 individuals with an unrelated muscle disease (BMD ($n = 3$), CAPN3 ($n = 2$), DMD ($n = 3$), DM1 ($n = 1$), OPMD ($n = 3$) and γ -SARC ($n = 2$); 2 monosomic and 12 disomic) and 2 disomic ICF patients. DNA was extracted essentially as described by Miller *et al.* [13].

Phenotypic FSHD patients were collected from three different families. Kindred 1 consists of two healthy parents and two children affected with FSHD, both disomic with a 4qA D4Z4 repeat array of 96 kb inherited from the father and a 170 kb 4qA fragment from the mother.



159

repeat array from her mother, a 132 kb large 4qB fragment from her father. Neither of the maternal grandparents is affected.

5.3.2 Methylation analysis of proximal chromosome 4 repeat unit

Four µg of PBL-DNA was co-digested with restriction enzymes *EcoRI*, *BglII* (MBI Fermentas), and *BlnI* (Amersham Pharmacia Biotech). Subsequently, 12 units of CpG methylation-sensitive enzyme *BsaAI* (New England Biolabs Inc.) were added. For CpG methylation-sensitive enzyme *FseI*, after digestion with *EcoRI/BglII/BlnI*, five µg of PBL-DNA was precipitated with 1/10 3M NaAc pH 5.3, 2.5× 100% EtOH and the pellet was dissolved in 30 µl Tris/EDTA pH 7.5. Thereafter, an *FseI* digestion (12 units, New England Biolabs Inc.) was performed. All digestions were incubated overnight according to the manufacturer's instructions.

DNA was separated on a linear 0.8% agarose gel (UltraPURE agarose, GibcoBRL Lifetechnologies Ltd). Electrophoresis was performed in 1× TAE supplemented with ethidium bromide. After electrophoresis, DNA was transferred to a Hybond-XL membrane (Amersham Pharmacia Biotech). The probe p13E-11 (D4F104S1) [20], and an empty vector control probe were labelled by random priming with ³²P-dCTP using the megaprime DNA labelling system (Amersham Pharmacia Biotech). Hybridisations were performed as described previously [11].

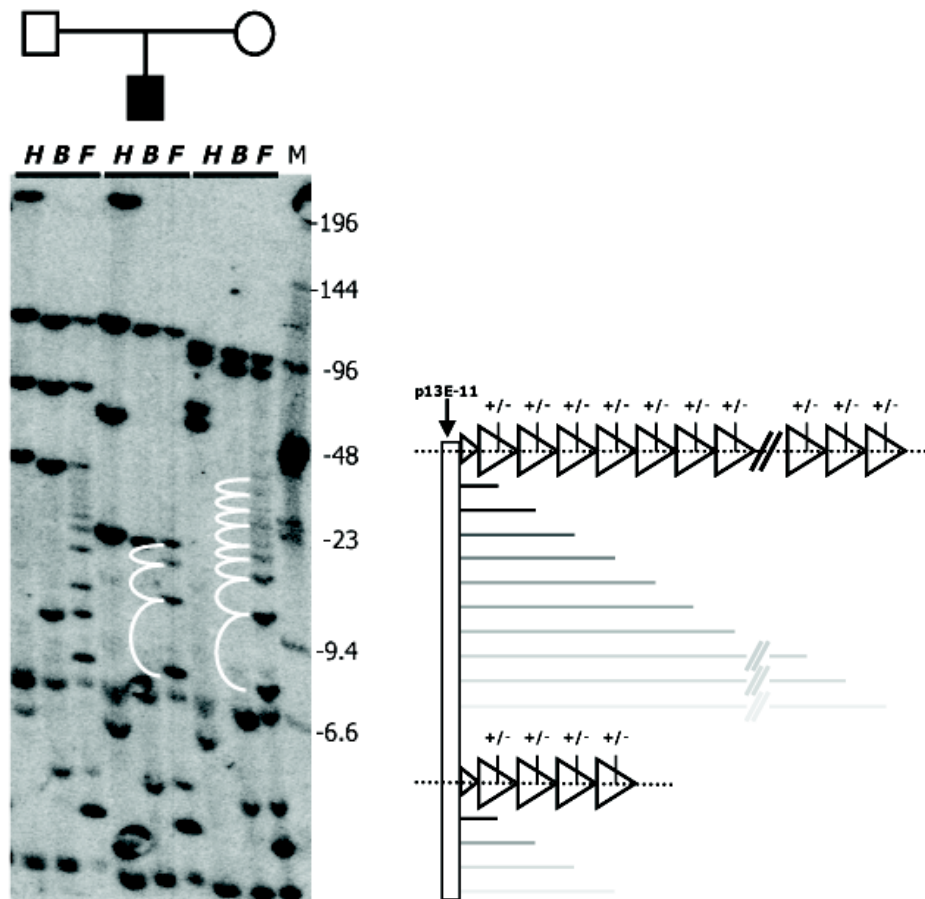
Next, phosphorimager patterns were analysed using ImageQuant software (Molecular Dynamics). After correction for background, methylation of chromosome 4 alleles was calculated by comparing undigested and digested fragments.

Several control experiments were performed to confirm complete digestion by methylation-sensitive enzymes. Since methylation-insensitive isoschizomers are not available for either of these enzymes, equimolar amounts of a plasmid containing one restriction site for each methylation-sensitive enzyme were added. This allows endogenous verification of complete digestion by hybridisation with an empty vector probe. Although this does not monitor the presence of impurities bound to the genomic DNA, identical results were achieved in the methylation PFGE assays. Moreover, the methylation assay was applied five times independently to DNA of one FSHD patient and two controls without significant differences ($P = 0.74$; with an average deviation of <3%).

Finally, six FSHD patients and two controls were tested, sampled twice with an interval of at least five years without observing an effect of aging (*BsaAI*, $P = 0.21$ and *FseI*, $P = 0.74$), and with an average variation in methylation within each individual <2%.

5.3.3 Methylation analysis of whole chromosome 4 repeat array

Five µg of PBL-DNA embedded in agarose (InCert agarose, Cambrex) was equilibrated overnight in the proper restriction buffer and double digested with 20 units *EcoRI/HindIII* (MBI Fermentas),



Supplementary Figure 5.2

Example of methylation analysis by PFGE of one classical FSHD kindred.

To assess whether the methylation status of the proximal unit reflects that of the entire array, DNA from 15 individuals from 6 FSHD families was digested with *EcoRI*, *BlnI* and *FseI* and separated by pulsed-field gel electrophoresis (PFGE). After hybridisation, extended, regular ladders of digested D4Z4 arrays were detected in the *FseI* lanes, characteristic for equal partial methylation of all chromosome 4-type repeat units within one array. Moreover, a clear inter-individual and inter-allelic variation of methylation at the *FseI* site was observed. Demonstrating the hypomethylation of FSHD alleles, the FSHD-specific ladders were far shorter than the control ladders, while also the completely undigested disease fragments typically showed less residual signal intensity in FSHD samples than in normal samples. The four monosomic individuals tested by both pulsed-field and proximal repeat unit tests, showed an average deviation of <5% between both methylation assays.

In this figure, the left panel gives a PFGE example of three individuals: two parents with a *de novo* FSHD kindred digested with *EcoRI/HindIII* ("H"), *EcoRI/BlnI* ("B") and with *EcoRI/BlnI/FseI* ("F"), respectively. The father carries three 4-type arrays of 48 kb, 75 kb and 105 kb and one 10-type array of 200 kb (trisomic). The mother carries 4-type arrays of 90 kb and 96 kb and 10-type arrays of 60 kb and 70 kb (disomic). The affected son inherits from the father the 105 kb 4-type and 200 kb 10-type array and from the mother the 60 kb 10-type array and an apparently maternal rearranged FSHD allele of 23 kb (disomic). In the *FseI* lane, variable ladders of D4Z4 repeat arrays (indicated in white) are hybridising due to the equal partial methylation of each consecutive D4Z4 unit. The marker lane (in kb) is indicated with (M).

In the right panel a schematic representation of the generation of ladders of D4Z4 fragments by partial methylation of the *FseI* restriction site (+/-) is drawn. Triangles represent D4Z4 units and the probe region p13E-11 is indicated proximal to the array. Partial methylation of each consecutive unit will lead to a gradual fading of the signal intensity with increased array size (gray scales). Due to the hypomethylation of FSHD arrays, these *FseI* ladders will fade sooner than those derived from normal arrays.

or with *EcoRI/BlnI*. Subsequently, plugs digested with *EcoRI/BlnI* were washed in Tris/EDTA pH 7.5 and equilibrated in *FseI* restriction buffer. Next, 20 units of *FseI* were added. All digestions were incubated overnight and performed according to the manufacturer's instructions.

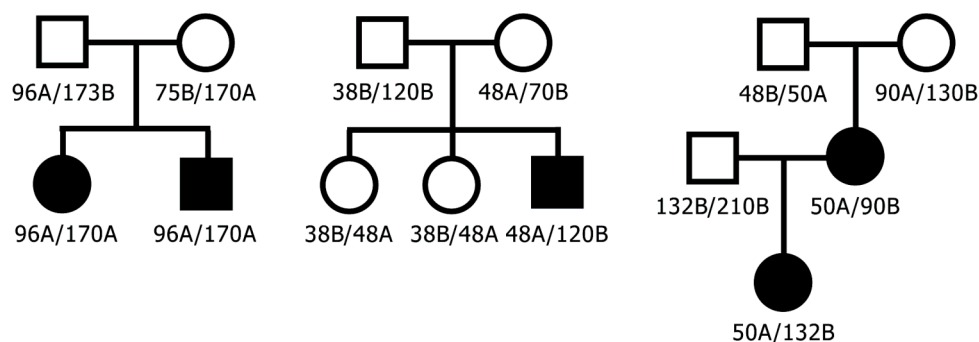
After digestion, DNA was separated by PFGE on a 0.8% agarose gel (MP agarose, Boehringer Mannheim) as described previously [3, 9, 11] with the exception that DNA was transferred to Hybond-XL membranes. Thus, 4q-type alleles were separated from 10q-type alleles by the *BlnI* sensitivity of the latter [2]. Phosphorimager patterns were again analysed using ImageQuant software.

5.3.4 Statistical analyses

Main effects of age (at time of blood sampling; in years) and gender were assessed with ANOVA. Methylation statuses were compared between healthy family members and healthy unrelated controls by independent sample T-tests. To determine whether methylation status was due to allele constitution, we compared the two most divergent allele constitutions (monosomic patients versus disomic controls) in a non-parametric two independent samples test (Mann-Whitney *U* test, asymmetric). To determine whether methylation status on either chromosome 4 or 4 and 10 is due to pathology or repeat array constitution, we used independent sample T-tests.

Statistical analyses of all individuals showed no differences in methylation due to age or gender, nor between healthy family members and unrelated controls. In subsequent analyses, all data were pooled across these variables.

The fractions of healthy and disease chromosomes 4 in patients and control individuals were used in a linear regression analysis to determine the contribution of these alleles on percentage of methylation at each restriction site. Monosomic FSHD patients have a fraction of zero (no healthy alleles), disomic patients have a fraction of 1/2 (one FSHD and one healthy allele). Monosomic control individuals have a fraction of 1 (one healthy allele out of one) and disomic



Supplementary Figure 5.3

Pedigrees of the three phenotypic FSHD kindreds. The affected patients are indicated in black. None of the affected individuals shows a contraction of the D4Z4 repeat on chromosome 4. Normal segregation of the D4Z4 repeats is observed in all individuals. The sizes of the D4Z4 repeat (in kb) and the allelic origin of the repeat, 4qA (A) or 4qB (B), are indicated below each individual.

control individuals have a fraction of 2 (two healthy alleles out of two). The data from monosomic and disomic control individuals were pooled into fraction 1, as both groups only contain healthy alleles.

5.4 Acknowledgements

We dedicate this work to the memory of Lodewijk Sandkuijl, who unexpectedly passed away during the preparation of this manuscript. We thank R ten Hove for contributing to the development of the methylation assays and C Wijmenga and D Smeets for providing material from individuals with ICF. FSHD research is made possible by the Prinses Beatrix Fonds, the Muscular Dystrophy Association USA, the FSH Society, the Stichting FSHD, the Shaw Family and the National Institutes of Health.

5.5 References

1. Amir RE, Veyver van den IB, Wan M, Tran CQ, Francke U, and Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* 23(2): 185-188.
2. Deidda G, Cacurri S, Piazzo N, and Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet* 33(5): 361-365.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.

8. Lemmers RJLF, Kievit de P, Sandkuijl L, 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.
9. 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, *et al.* (1998) Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet* 7(8): 1207-1214.
10. Lunt PW (1998) 44th ENMC International Workshop: Facioscapulohumeral Muscular Dystrophy: Molecular Studies 19-21 July 1996, Naarden, The Netherlands. *Neuromusc Disord* 8(2): 126-130.
11. Maarel van der SM, Deidda G, Lemmers RJ, 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.
12. 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.
13. Miller SA, Dykes DD, and Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3): 1215.
14. 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.
15. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
16. Rand E and Cedar H (2003) Regulation of imprinting: A multi-tiered process. *J Cell Biochem* 88(2): 400-407.
17. Randolph-Anderson BL, Stajich JM, Graham FL, Pericak-Vance MA, Speer MC, and Gilbert JR (2002) Evidence consistent with linkage to 15q of a non-chromosome 4 linked FSHD family. *Am J Hum Genet* 71(4; Supplement): 530.
18. Robertson KD and Wolffe AP (2000) DNA methylation in health and disease. *Nat Rev Genet* 1(1): 11-19.
19. Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, and Higgs DR (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 34(2): 157-165.
20. 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.
21. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, and Viegas-Pequignot E (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402(6758): 187-191.

6

Residual D4Z4 repeat size and D4Z4 methylation levels separate FSHD into two severity classes

*PGM van Overveld¹, L Enthoven², E Ricci³,
L Felicetti³, M Jeanpierre⁴, ST Winokur⁵, RR Frants¹,
GW Padberg⁶ and SM van der Maarel¹*

Submitted

- 1 Department of Human Genetics, Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
- 2 Division of Medical Pharmacology, Leiden/Amsterdam Center for Drug Research, Leiden University Medical Center, Leiden, The Netherlands.
- 3 Center for Neuromuscular Diseases, UILDM-Rome Section & Institute of Neurology, Università Cattolica; Rome, Italy.
- 4 U129 INSERM & Service de Biochimie Génétique, ICGM, CHU Cochin, Paris, France.
- 5 Department of Biological Chemistry, College of Medicine, University of California, Irvine CA, USA.
- 6 Department of Neurology, University Medical Center St. Radboud, Nijmegen, The Netherlands.

6.1 Abstract

Facioscapulohumeral muscular dystrophy (FSHD) progressively affects the facial, shoulder and upper arm muscles, and is associated with contractions of the polymorphic D4Z4 repeat array in 4q35. Recently, we demonstrated that FSHD alleles are hypomethylated at D4Z4. To study potential relationships between D4Z4 hypomethylation and residual repeat size and clinical severity, we compared the clinical severity score with D4Z4 methylation in unrelated FSHD patients. We show that correcting the clinical severity score for age at examination improves the parameter to define clinical severity and provide further support for a significant hypomethylation of FSHD alleles. However, a linear relationship between repeat size and severity of the disease cannot be established by this method. Interestingly, our data indicate that FSHD can be separated into two clinical severity classes: patients with residual D4Z4 repeat sizes of 10-20 kb are always severely affected and show pronounced D4Z4 hypomethylation. In contrast, patients with D4Z4 repeat sizes of 20-38 kb show a large interindividual variation in clinical severity and D4Z4 hypomethylation. As the vast majority of familial FSHD cases are represented in this interval and considering the overt variation in clinical severity in these familial cases, it is thus imperative to develop comprehensive allele-specific assays monitoring total D4Z4 methylation to investigate whether interindividual variation in D4Z4 methylation can be translated into a prognostic factor for clinical severity.

6.2 Introduction

Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD1A, MIM 158900) progressively affects the facial, shoulder and upper arm muscles. This myopathy displays a high inter- and intrafamilial variability in age at onset and progression of the disease [21]. The major locus for FSHD has been mapped to the subtelomere of chromosome 4q (4q35) [26, 28] where it is associated with contractions of a polymorphic repeat array consisting of 3.3 kb repeated elements, designated D4Z4 [7, 27]. In healthy individuals the D4Z4 array consists of 11-100 units, while patients with FSHD carry a repeat array of 1-10 units [27]. The subtelomere of chromosome 10 also harbours a polymorphic repeat array, which is 98% homologous to D4Z4 [1, 5]. Despite the presence of translocated 10-derived repeat units on chromosome 4, and *vice versa*, in ~20% of individuals [4, 6, 11, 19], only short repeat arrays on chromosome 4 cause FSHD.

The D4Z4 repeat is very GC-rich, which is characteristic for CpG islands. Each unit encodes a putative open reading frame (ORF) containing two homeobox domains (*DUX4*). A promoter region precedes this ORF, which contains a TACAA box 150 bp upstream of the ATG start codon [8]. Sequence analysis of D4Z4 further revealed a 316 bp region with homology to the sequence of *LSau*, a GC-rich DNA repeat, and also a region with high similarity to the sequence of a low-copy repeat hhspm3. Both these repetitive elements are associated with heterochromatin [9]. FISH analysis showed an extensive dispersion of D4Z4 to heterochromatic regions of the genome [2, 9, 13, 29]. This heterochromatic association, and the high GC content of the repeat, may indicate that normally *DUX4* is transcriptionally inactive. Indeed, no transcripts of *DUX4* have been identified so far *in vivo*.

The heterochromatic properties of D4Z4 and the lack of evidence for active transcription emanating from D4Z4 have led to the hypothesis that FSHD is caused by an epigenetic disease mechanism [3]. In this model contraction of the D4Z4 repeat results in a local change in chromatin structure and consequently the transcriptional deregulation of one or more FSHD genes. These genes were originally proposed to be located in, or close to the D4Z4 repeat array, but recent evidence suggests that they may also be located at considerable distance through a long distance looping model [10] or even at different chromosomes through a nuclear disorganisation model as 4qter occupies a distinct peripheral nuclear territory [16].

Recently, we presented evidence for an epigenetic factor involved in FSHD by demonstrating that FSHD alleles are hypomethylated at D4Z4 [20]. We applied a methylation assay on DNA isolated from peripheral blood lymphocytes (PBLs) and muscle of monosomic and disomic individuals. Individuals carrying a standard allele constitution of 4-derived repeat units on chromosome 4 and 10-derived repeat units on chromosome 10 are referred to as 'disomic' for D4Z4, while individuals carrying one array of 10-derived repeat units at one of their chromosomes 4 are referred to as 'monosomic' for D4Z4. The methylation assay quantifies the methylation of two CpG dinucleotides within the most proximal unit of the D4Z4 array on chromosome 4 by use of two different methylation sensitive restriction enzymes. Due to the extensive spreading of D4Z4 through the genome, methylation analysis of D4Z4 is very complex. Currently, this is the only assay with sufficient specificity to demonstrate significant hypomethylation of D4Z4 in FSHD alleles despite the fact that only 2 of the 290 CpG dinucleotides in the D4Z4 unit are tested. As we can discriminate between chromosomes 4 and 10 in our methylation assay, only monosomic individuals, who are observed in only 7% of the population [6, 17, 19, 23, 25], allow the analysis of the methylation of D4Z4 in a single chromosome 4. In contrast, disomic individuals, who are much more common (80% of the

population), can only be analysed for methylation at both chromosomes 4 simultaneously which prohibits precise methylation measurements of the disease allele.

DNA analysis of monosomic individuals thus provides the unique possibility to examine the methylation of single healthy or disease alleles, without interference or masking of any other homologous locus in the genome. To obtain more information on the degree of hypomethylation at the FSHD allele and to investigate a possible relationship of methylation with repeat array size and severity, and the clinical implications, we analysed D4Z4 methylation in 40 independent chromosome 4 alleles.

6.3 Subjects and methods

6.3.1 Patients and control individuals

All 40 individuals (21 patients and 19 control individuals) included in this survey were analysed previously for their allele sizes and constitution of repeat arrays on chromosomes 4 and 10 ([14, 15, 19] and unpublished results). Five of these patients were already included in a previous study [20]. Only monosomic individuals with a 4-type array on chromosome 4, a translocated 10-type array on chromosome 4 and two 10-type arrays on chromosome 10 were included in this study to compare the methylation status of the proximal D4Z4 repeat unit with repeat array size and with two different parameters that define severity of the disease.

All 21 patients, 12 probands and 9 non-proband affected family members, were clinically evaluated and the expression of their phenotype was scored according to the clinical severity score list (CSS) formulated by Ricci *et al.* [22]. Since FSHD is a slowly progressive myopathy, it is very difficult to determine the precise onset in a patient [21]. Age at onset is usually determined in retrospect and often depends on the recollection and perception of the patient and on a possible family bias. We therefore corrected the CSS for age at examination, given that the duration of this progressive disease determines the CSS at the time of examination:

$$\text{age-corrected CSS} = ((\text{CSS} \times 2) / \text{age at examination}) \times 1000$$

The severity score was multiplied by two to generate whole numbers and we then multiplied the outcome of this calculation with 1000 to improve interpretation of the results and visualisation in graphs.

6.3.2 DNA analysis

Genomic DNA was isolated from peripheral blood lymphocytes (PBLs) from all monosomic FSHD patients and control individuals, both family members and unrelated individuals. DNA was extracted essentially as described by Miller *et al.* [18].

For analysis of D4Z4 by the methylation-sensitive restriction enzymes *Bsa*AI and *Fse*I (New England Biolabs), we applied a method illustrated in **Figure 6.1** and as described previously [20]. To monitor complete digestion, we added equimolar plasmid DNA to the genomic DNA prior to digestion with the methylation-sensitive enzyme of interest and after a *Bgl*II/*Bln*I digestion, all according to the manufacturer's instructions. DNA was then separated by linear gel electrophoresis and after Southern blotting the membranes were hybridised with radioactive labelled probe p13E-11 (D4F104S1) [27] and subsequently with an empty plasmid vector to confirm complete digestion of the exogenous DNA control. Next, membranes were exposed to phosphor-imager screens and signal intensities were analysed with ImageQuant software (Molecular dynamics).

6.3.3 Statistical analyses

Methylation was analysed by ANOVA with disease status (control individuals and FSHD patients) as a fixed factor. Possible relationships between (1) repeat array size versus FSHD severity, (2) repeat array size versus methylation and (3) methylation versus FSHD severity were estimated using linear regression analyses and were tested by Pearson correlation. The mean of a set of data is given as mean $\pm$ S.E.M. Analyses were performed on all available data.

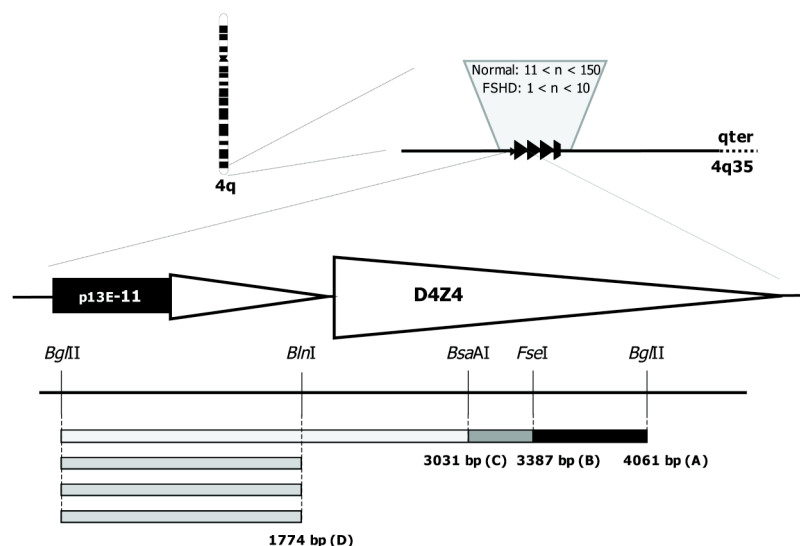


Figure 6.1

Schematic overview of the modified *Bgl*II/*Bln*I dosage test developed for analysis of the proximal D4Z4 repeat unit. Bars represent the different fragment sizes, which are released when PBL-DNA from monosomic individuals is digested with *Bsa*AI or *Fse*I subsequent to *Bgl*II and *Bln*I. *Fse*I will generate a fragment of 3387 bp ("B") and *Bsa*AI of 3031 bp ("C"). Besides the chromosome 4 fragment of 4061 bp ("A"), generated by *Bgl*II, when both *Bsa*AI and *Fse*I do not cut, also the chromosome 10 fragment of 1774 bp ("D") is shown that is released by the use of *Bgl*II and *Bln*I. Both methylation-sensitive restriction sites are located distal to *Bln*I, so any putative methylation of the chromosome 10 will not be detected with this assay. Monosomic individuals have a single 4-type array on one of their chromosomes 4, and a 10-type array on the other, and two 10-type arrays on both their chromosomes 10. Thus, with this method only one 4-type array will be analysed.

6.4 Results

6.4.1 Repeat array size in relation to FSHD severity

As expected, all monosomic FSHD patients displayed smaller chromosome 4 repeat arrays compared to the control group (FSHD: $24.14 \text{ kb} \pm 1.09 \text{ kb}$ versus control individuals: $91.81 \text{ kb} \pm 5.94 \text{ kb}$).

We first tested whether we could confirm previous findings that showed a correlation between the severity of the phenotype and D4Z4 fragment size in FSHD patients [12, 24]. We evaluated this in our monosomic patients with two different definitions for clinical severity; comparing repeat array size with the clinical severity scale (CSS) according to Ricci *et al* [22] and with an age-corrected CSS in which we corrected for the age at examination as FSHD is a progressive disease. This first comparison revealed a significant relationship ($r = -0.45$; $P = 0.05$) and confirmed that severity of the disease increases with decreasing residual repeat size. Moreover, when the CSS was corrected for age at examination, age-corrected CSS, the inverse relationship between repeat array size and severity became even stronger ($r = -0.52$; $P < 0.05$).

At close examination, we observed considerable variation in the clinical severity within the 20-38 kb interval, while in the 10-20 kb interval all patients were severely affected based on age-corrected CSS.

6.4.2 Comparison of FSHD patients with control individuals

Next, we compared the methylation of the proximal D4Z4 unit on chromosome 4 in 21 independent FSHD alleles and 19 independent control alleles. Corroborating and extending on our previous study [20], we observed a significant hypomethylation in FSHD alleles ($P < 0.001$) on

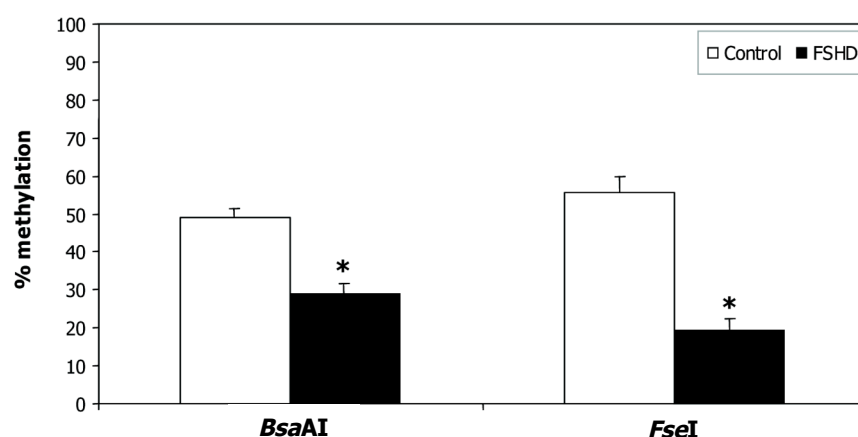


Figure 6.2

Bar diagram of methylation percentages in monosomic individuals. Methylation at *Bsa*AI and *Fse*I methylation-sensitive restriction sites within the proximal D4Z4 repeat unit is shown for 19 control individuals and 21 FSHD patients. Data are presented as mean $\pm$ S.E.M. *: $P < 0.001$ versus control individuals.

both restriction sites (FSHD: *Bsa*AI 29.05% $\pm$ 2.43% and *Fse*I 19.47% $\pm$ 2.77% versus control individuals: *Bsa*AI 48.88% $\pm$ 2.23% and *Fse*I 55.68% $\pm$ 4.21%). All results are presented in **Figure 6.2**.

When the patient group was analysed in more detail, we did not observe any differences between probands and non-proband affected family members for both methylation-sensitive enzymes, repeat array size, age at examination, CSS or age-corrected CSS.

6.4.3 Repeat array size, methylation and severity

Since we were interested in a possible relationship of methylation on D4Z4 and both severity and residual repeat array size, we performed a linear regression analysis. The results for both enzymes are plotted in **Figure 6.3**. Although initially, statistical analysis displayed a very significant relationship ($P < 0.001$), this is solely due to the clustering into two groups that were formed by the patients on the one hand and the control individuals on the other hand. When these two groups were analysed separately, the relationship was abolished (FSHD patients: *Bsa*AI $P = 0.24$ and *Fse*I $P = 0.45$; control individuals: *Bsa*AI $P = 0.54$ and *Fse*I $P = 0.18$), indicating that there is no apparent relation between methylation and residual repeat array size by this method. Accordingly, we did not observe a linear relation between methylation of the D4Z4 repeat array and phenotypic severity of the disease, whether for CSS or for age-corrected CSS.

However, separating the patient group in a 10-20 kb and 20-38 kb interval, we again observed a marked difference in methylation status of D4Z4, although the groups are too small to test statistically. Patients in the 10-20 kb interval were on average 18% methylated at *Bsa*AI and 11% methylated at *Fse*I, with an average age-corrected CSS of 186. In contrast, in the 20-38 kb interval average intermediate methylation levels were detected of 31% for *Bsa*AI and 21% for *Fse*I, with an average age-corrected CSS of 108. Moreover, in analogy with the considerable variation in clinical severity in the 20-38 kb interval, methylation values in this interval were also highly variable.

6.5 Discussion

Recently, we demonstrated that an epigenetic mechanism likely plays a role in FSHD by demonstrating hypomethylation of D4Z4 in 4q-linked and non 4q-linked FSHD [20]. However, only 5 monosomic Dutch FSHD patients were then compared with all other analysed groups (including 79 control individuals), due to the rarity of the monosomic allele constitution in the FSHD population. These five patients displayed a significant hypomethylation for restriction sites *Bsa*AI and *Fse*I ($P < 0.01$) in comparison with monosomic and disomic healthy individuals. As

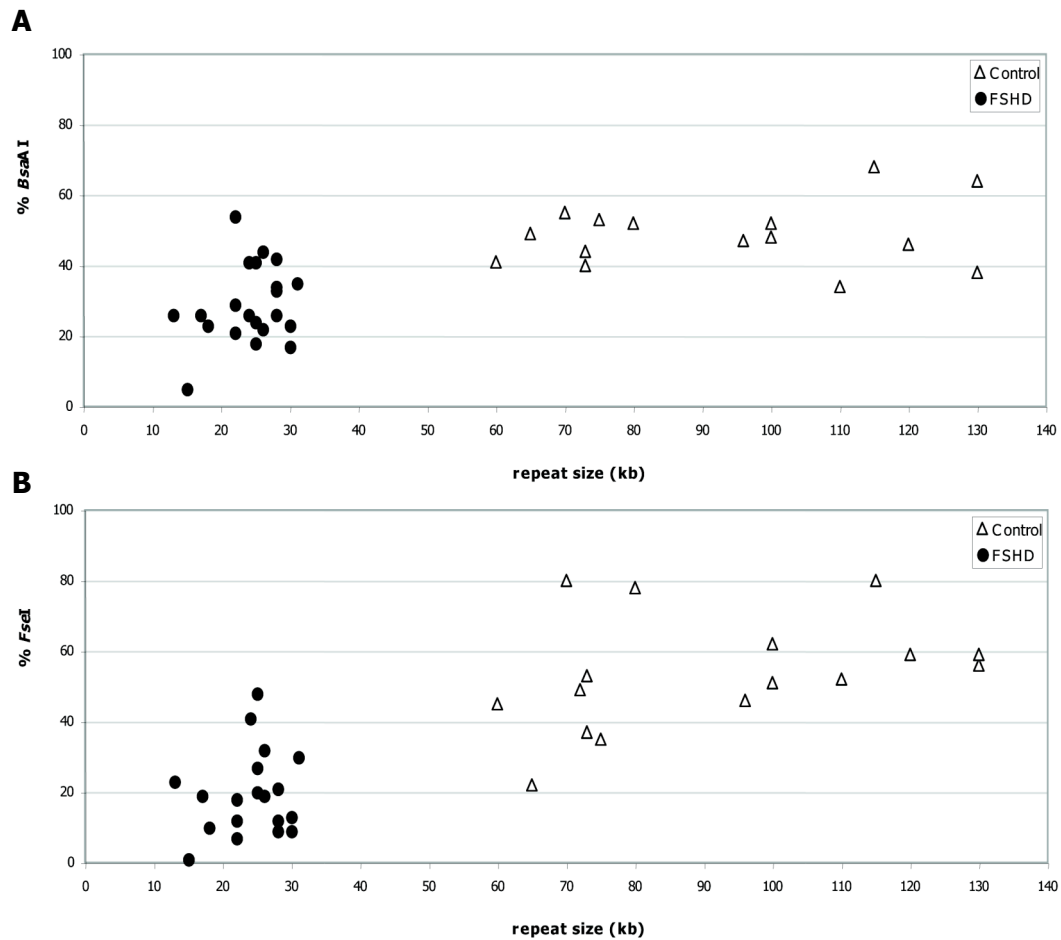


Figure 6.3

Scatter plots of methylation-sensitive enzymes *Bsa*AI (A) and *Fse*I (B) versus repeat size. Methylation status per enzyme (%) is plotted against the residual fragment size (kb). FSHD patients are depicted as black circles and control individuals as open triangles. Per enzyme all available data are plotted.

disomic patients displayed intermediate methylation levels between monosomic patients and control individuals, we concluded that in these patients the reduced methylation of the disease allele was masked by the high methylation of the healthy allele [20]. The observed average methylation in disomic patients can therefore not adequately be related to repeat array size or disease severity.

In contrast, when applying our restriction enzyme-based assay on DNA from monosomic individuals only, it enabled us to examine the characteristics of a single chromosome 4q35 locus without interference of any other homologous locus. We therefore collected 16 additional monosomic FSHD patients to study potential relationships between methylation and residual repeat size and clinical severity. This survey of 21 monosomic patients corroborates and extends our previous study by revealing a significant reduction in methylation of D4Z4 in FSHD alleles at the sites tested ($P < 0.001$), as depicted in *Figure 6.2*. This figure also shows that the CpG dinucleotides at both restriction sites were similarly reduced for their methylation status in FSHD

alleles, indicating that these sites are representative for the changes in methylation at the whole repeat array.

Various research groups have established a correlation between the severity of the FSHD phenotype and residual repeat fragment size [12, 24]. Some studies have used age at onset as clinical parameter in this comparison. We used the Clinical Severity Scale (CSS) by Ricci *et al.* [22], since determination of disease onset depending on the patient's recollection is not a reliable parameter. In our sample set we could confirm the previously observed relationship between residual repeat array size and severity of the disease as defined by CSS. Furthermore, when we corrected the CSS for the age at examination, this relationship became stronger. It therefore seems that this age-corrected CSS is a more accurate method to define severity in FSHD. Although a significant relation between the residual repeat size and either CSS or age-corrected CSS was observed, the R^2 indicates a weak predictive validity ($R^2 = 0.2$, $R^2 = 0.3$, respectively). Evaluating the individual data allows a division of FSHD patients into two severity classes: within the 20-38 kb interval there is large variation in the clinical severity, while, in contrast, patients with a residual repeat size of 10-20 kb are all severely affected based on age-corrected CSS. A similar pattern with a similar predictive validity of the correlation between fragment size and clinical severity was previously observed by Ricci *et al.* [22]. Both studies thus indicate that patients with a small residual fragment size (and concomitantly a high age-corrected CSS) are solely responsible for the observed linear relation. Although it is generally accepted that small residual repeat arrays coincide with a high CSS, our data indicate that there is no linear relationship between both parameters for the FSHD patient population as a whole.

We next determined if a relationship existed between D4Z4 methylation and repeat array size or between clinical severity and methylation. Although we do observe a significant hypomethylation on the 4q35 locus in individuals with repeat array sizes of 10-38 kb, we could not establish a linear relationship between methylation and repeat sizes, nor between methylation and disease severity, scored by CSS or age-corrected CSS. However, when separating the patients again according to the two severity classes, patients in the 10-20 kb interval show much more pronounced hypomethylation of D4Z4 than patients in the 20-38 kb interval, where the hypomethylation is also much more variable.

Although both groups are too small to perform statistical tests, our results indicate that D4Z4 methylation may in part determine the FSHD severity. Therefore, a comprehensive methylation test assaying the average methylation of all CpG dinucleotides within the entire array rather than the analysis of two single CpG dinucleotides within the most proximal D4Z4 unit may prove more suitable for establishing phenotype-(epi)genotype relationships. However, due to the extensive spreading of D4Z4 copies throughout the genome, such test is currently not available. Finally, if D4Z4 methylation is an important parameter determining clinical severity,

differential D4Z4 methylation between affected individuals within a single family may explain the overt intrafamilial variation in clinical severity. It is therefore imperative to develop comprehensive allele-specific D4Z4 methylation assays to address this issue and to investigate whether D4Z4 methylation levels can be used as a prognostic factor for clinical severity.

6.6 Acknowledgements

This work was supported by grants from The Netherlands Organisation for Scientific Research (NWO) (917.56.338), the Prinses Beatrix Fonds, the Stichting Spieren voor Spieren, the Muscular Dystrophy Association USA, and the Shaw family.

6.7 References

1. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, Wielen van der M, Rasmussen K, and Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2: 39-44.
2. Bengtsson U, Altherr MR, Wasmuth JJ, and Winokur ST (1994) High resolution fluorescence in situ hybridization to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. *Hum Mol Genet* 3(10): 1801-1805.
3. Bickmore WA and van der Maarel SM (2003) Perturbations of chromatin structure in human genetic disease: recent advances. *Hum Mol Genet* 12 Spec No 2: R207-R213.
4. Cacurri S, Piazzo N, Deidda G, Vigneti E, Galluzzi G, Colantoni L, Merico B, Ricci E, and Felicetti L (1998) Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric *KpnI* repeat units implicated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 63: 181-190.
5. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, and Felicetti L (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3: 155-167.
6. 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.
7. 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.
8. Gabriels J, Beckers MC, Ding H, Vriese de A, Plaisance S, Maarel van der SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al.* (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 236(1): 25-32.
9. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, Deutekom van JC, Francis F, Sharpe PT, Hofker M, *et al.* (1994) Analysis of the tandem repeat locus D4Z4

- associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3(8): 1287-1295.
10. 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.
 11. 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, *et al.* (1998) Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet* 7(8): 1207-1214.
 12. 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.
 13. Lyle R, Wright TJ, Clark LN, and Hewitt JE (1995) The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* 28(3): 389-397.
 14. Maarel van der SM, Deidda G, Lemmers RJ, 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.
 15. 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.
 16. Masny PS, Bengtsson U, Chung SA, Martin JH, Engelen van B, Maarel van der SM, and Winokur ST (2004) Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum Mol Genet* 13(17): 1857-1871.
 17. 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).
 18. Miller SA, Dykes DD, and Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3): 1215.
 19. 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.
 20. Overveld van PGM, Lemmers RJLF, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, Ommen van GB, Frants RR, and Maarel van der SM (2003) Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* 35(4): 315-317.
 21. Padberg GW (1982) Facioscapulohumeral disease. Thesis, Leiden University.
 22. Ricci E, Galluzzi G, Deidda G, Cacurri S, Colantoni L, Merico B, Piazza 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 *KpnI* repeats at the 4q35 locus and clinical phenotype. *Ann Neurol* 45(6): 751-757.
 23. 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.
 24. Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, and Figlewicz D (1996) Evidence for anticipation and association of deletion size with

- severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39(6): 744-748.
25. 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.
 26. Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, and Padberg GW (1990) Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* 336(8716): 651-653.
 27. 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.
 28. Wijmenga C, Padberg GW, Moerer P, Wiegant J, Liem L, Brouwer OF, Milner EC, Weber JL, Ommen van GB, Sandkuijl LA, *et al.* (1991) Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and *in situ* hybridization. *Genomics* 9(4): 570-575.
 29. Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, *et al.* (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosome Res* 2(3): 225-234.

7

Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q

*G Jiang¹, F Yang¹, PGM van Overveld²,
V Vedanarayanan³, SM van der Maarel² and M Ehrlich¹*

Originally published in:
Human Molecular Genetics
(2003) 12: 2909-2921

- ¹ Human Genetics Program & Department of Biochemistry, Tulane Medical School, New Orleans LA, USA.
- ² Department of Human Genetics, Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.
- ³ Department of Neurology, University of Mississippi Medical School, Jackson MS, USA.

7.1 Abstract

Facioscapulohumeral muscular dystrophy (FSHD) is a unique dominant disorder involving shortening of an array of tandem 3.3 kb repeats. This copy number polymorphic repeat, D4Z4, is present in arrays at both 4q35 and 10q26, but only 4q35 arrays with one to 10 copies of the repeat are linked to FSHD. The most popular model for how the 4q35 array-shortening causes FSHD is that it results in a loss of postulated D4Z4 heterochromatinisation, which spreads proximally, leading to overexpression of FSHD genes in *cis*. This would be similar to a loss of position-effect variegation (PEV) in *Drosophila*. To test for the putative heterochromatinisation, we quantitated chromatin immunoprecipitation with an antibody for acetylated histone H4 that discriminates between constitutive heterochromatin and unexpressed euchromatin. Contrary to the above model, H4 acetylation levels of a non-repeated region adjacent to the 4q35 and 10q26 D4Z4 arrays in normal and FSHD lymphoid cells were like those in unexpressed euchromatin and not constitutive heterochromatin. Also, these control and FSHD cells displayed similar H4 hyperacetylation (like that of expressed genes) at the 5' regions of 4q35 candidate genes *FRG1* and *ANT1*. Contrary to the loss-of-PEV model and a recent report, there was no position-dependent increase in transcript levels from these genes in FSHD skeletal muscle samples compared with controls. Our results favour a new model for the molecular genetic aetiology of FSHD, such as differential long-distance *cis* looping that depends upon the presence of a 4q35 D4Z4 array with less than a threshold number of copies of the 3.3 kb repeat.

7.2 Introduction

Facioscapulohumeral muscular dystrophy (FSHD) has a very unusual molecular genetic aetiology. It involves a 3.3 kb DNA repeat, D4Z4, in a tandem array. In FSHD patients, the array is shortened below a threshold copy number on only one homologue at the disease-associated 4q35. While the dominantly inherited genetic defect in FSHD has been identified, the way in which it causes the characteristic type of muscular dysfunction is mysterious [35]. In almost all FSHD patients there are only 1-10 tandem copies of the D4Z4 repeat in a subtelomeric region of one chromosome 4 (Chr 4) homologue [30]. Unaffected individuals have 11-100 copies on both Chr 4 homologues. Below this threshold copy number (11 copies) for D4Z4 repeats at one 4q35 allelic region, there is about 95% penetrance of the disease by age 20. The disease is generally much worse (earlier onset and greater clinical severity) when the short 4q D4Z4 array is in the smaller size range (one to four repeats) for FSHD [35, 46, 51].

Copy number polymorphic arrays of D4Z4 are also found in the subtelomeric region of 10q (10q26) [3, 11]. The homology between the canonical D4Z4 repeats at 4q35 (GenBank AF117653) and those at 10q26 (GenBank AY028079) is 99% and includes an open reading frame (ORF), *DUX4*, within the D4Z4 repeat [4, 24] and an upstream sequence (*Figure 7.1*) that functions as a promoter in a reporter gene assay [18]. The ORF could encode a protein containing two homeobox-type sequences. However, no polyadenylation signal is present downstream of this ORF, and while there is evidence for transcription of partially homologous sequences from the acrocentric chromosomes, *in vivo* transcripts exactly homologous to *DUX4* in 4q or 10q D4Z4 repeats have not been detected [4, 13, 24]. Homology between 4q35 and 10q26 is also seen 42 kb proximally to D4Z4 repeat arrays (>95% homology) and distally for approximately 15-25 kb in the region preceding the telomeric (TTAGGG)_n [19]. A short D4Z4 array at 4q35 appears to be only indirectly causing FSHD by *cis* interactions because the almost identical D4Z4 repeats in equally short arrays on 10q26 do not result in a disease phenotype [3], despite the extensive homology within, proximal, and distal to the D4Z4 arrays on 4q35 and 10q26. Furthermore, in FSHD patients, who have one short 4q35 array, the phenotype is unaffected by the exact copy number of D4Z4 at the two 10q26 allelic regions or one normal 4q35 allelic region, even though each of these regions can vary by over an order of magnitude in their D4Z4 copy numbers.

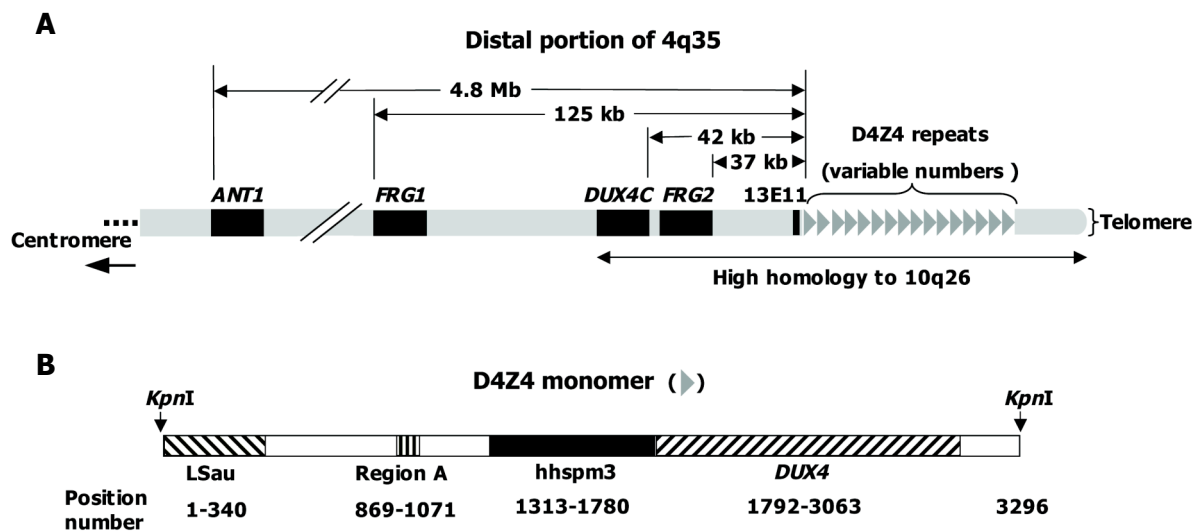


Figure 7.1

Schematic for the distal end of 4q35.

(A) Map of 4q35 region (not drawn to scale). Genes or putative genes (*ANT1*, *FRG1*, *DUX4C*, *FRG2*) and the non-gene region 13E11 (formally named p13E-11) are indicated, and the distance between these regions and the beginning of the first D4Z4 repeat is noted above each.

(B) Map of the D4Z4 monomer (GenBank AF117653). The unique *KpnI* site is shown as half sites at the end of the monomer. Several of the subregions mentioned in the text and their positions relative to the *KpnI* cleavage site are indicated.

It has been hypothesised that the vicinity of the D4Z4 array in 4q35 in unaffected individuals is heterochromatic, because it normally has many large tandem repeats containing two subsequences, LSau and *DUX4*, with highly homologous copies in known heterochromatic regions (the short arms of the acrocentric chromosome and 1qh) [24, 38]. Furthermore, the 4q35 region harbouring the array appears as a dark band on Giemsa staining, and is very close to telomeres [62]. However, whether human telomeric regions themselves are heterochromatic is unclear [64]. We found that unaffected individuals have one property indirectly associated with condensation of this DNA repeat, a high level of 5-methylcytosine [53]. However, DNA regions can be highly methylated without residing in constitutive heterochromatin [37].

The most frequently invoked model for how short arrays of D4Z4 at one allelic 4q35 region lead to the FSHD syndrome is derived from the putative heterochromatic structure of normal D4Z4 arrays and is based upon position-effect variegation (PEV) in *Drosophila* [58] and telomere silencing in yeast [23]. In PEV and yeast telomere silencing, there is a decrease in the percentage of cells expressing a euchromatic gene because that gene has been placed in the vicinity of heterochromatin by a rearrangement [43]. PEV can also result from multimerisation of a DNA sequence, e.g. three to seven copies of a 10 kb *Drosophila* transposon, with the severity of the silencing correlated with transgene copy number and proximity to constitutive heterochromatin [14, 50]. This *cis*-acting repression in PEV in *Drosophila* had been assumed to display a gradient of heterochromatin spreading and consequent decreased expression inversely proportional to the proximity of the gene to the endogenous heterochromatin region [56]. Telomeric silencing at broken chromosome ends in *Saccharomyces cerevisiae* has been demonstrated to involve linear spreading of repressive chromatin structures originating from the telomeric C₁₋₃A repeats by means of extended binding of heterochromatic proteins from the distal end proximally for about 2-4 kb inward from the C₁₋₃A repeats [44, 47].

The adaptation of the *Drosophila* PEV and yeast telomere silencing models to FSHD involves the assumption that the long D4Z4 arrays are normally heterochromatic and this heterochromatinisation in unaffected individuals spreads from the array to genes important to FSHD on 4q35 [17, 24, 29, 62]. According to this model, the putative normal heterochromatinisation at 4q35 is lost when there is only a short D4Z4 array on one Chr 4 homologue, as in FSHD patients. The consequence of the loss of this heterochromatin-spreading is predicted to be an inappropriate increase in expression of some critical gene(s) in *cis* in affected skeletal muscle of patients. We will refer to this type of *cis*-spreading of repressive structures as the loss-of-PEV hypothesis, which is predicated upon the loss of a genetically programmed repression rather than the gain of rearrangement-associated repression, as is usually the case for PEV. In considering this hypothesis, it should be noted that three essential starting points of the model are unknown, namely, whether the D4Z4 array at Chr 4 in unaffected

individuals is highly condensed, whether this condensation spreads, and whether FSHD patients have too little of this condensation.

Histone H3 methylation in the N-terminal tail at lysine 9 (K9) has been linked to heterochromatinisation in PEV and in normal gene control [9, 33, 45]. In contrast, hyperacetylation of a number of positions in the N-terminal tails of core histones and hypermethylation of H3 at K4 are characteristic of expressed gene or promoter regions or those poised for expression [21, 25, 41]. In *S. cerevisiae* there is a gradient of histone H4 K16 acetylation ranging from hypoacetylation within 1 kb of the telomere to hyperacetylation about 5-30 kb from the telomere [27]. Furthermore, in human cells, constitutive heterochromatin can be distinguished from unexpressed euchromatin by its stronger hypoacetylation of core histones [26, 42]. By chromatin immunoprecipitation (ChIP) assays on uncultured blood samples and cultured cells from unaffected individuals, as well as on somatic cell hybrids (SCHs), we looked for the type of histone hypoacetylation seen in constitutive heterochromatin in the following chromatin regions: normal long D4Z4 arrays; a non-repeated region present very close to the proximal end of the D4Z4 array on 4q and 10q; two gene or gene-like regions on both 4q35 and 10q26 (*FRG2* and *DUX4C*); and two candidate FSHD genes (*FRG1* and *ANT1*) found only on 4q35 (**Figure 7.1**). Several of these regions were also examined in analogous samples from FSHD patients to test for the predicted increase in histone acetylation that should accompany a decrease in heterochromatinisation according to the loss-of-PEV model. Also, we have re-examined the relative expression of *FRG1* and *ANT1* in control and FSHD skeletal muscle biopsy samples using quantitative real-time PCR (Q-PCR) because their expression was reported to be very much higher in FSHD samples [17]. Furthermore, in that study, genes closer to the D4Z4 array on 4q35 were reported to be more overexpressed than more distant genes in FSHD muscle, which is also in accordance with the loss-of-PEV hypothesis. In contrast, our results from both Q-PCR and ChIP contradict this hypothesis for FSHD and indicate that some other molecular genetic mechanism links FSHD symptoms to the presence of a 4q35 D4Z4 array having less than a threshold number of copies of the 3.3 kb repeat.

7.3 Subjects and methods

7.3.1 Cell lines and tissues

Normal human LCLs [15] and skin fibroblast cell strains derived from newborn foreskin were grown under standard conditions. Somatic cell hybrids GM11687 (mouse-human), GM14193 (Chinese hamster-human), and GM11448 (Chinese hamster-human) contained the following human chromosomes: Chr 4 only, del(4)p16.2 plus Chr 5, or del(4)(4qter>4p16.1) plus Chr 5

and Chr 8, respectively. To obtain PBMC samples, EDTA-treated human peripheral blood samples from healthy individuals or FSHD patients (about half of each from females) were used to generate a mononuclear cell fraction by density gradient centrifugation (Lymphocyte Separation Medium, Cappel). Unless otherwise noted, FSHD muscle samples were biopsy or orthopaedic scapular fixation tissue from moderately affected deltoid or biceps skeletal muscle and were from patients 13-79 years of age, with the exception of FM917, which was from unaffected muscle. Analogous disease-control muscle biopsy samples were from patients with unrelated neuromuscular diseases involving myopathic changes or denervation atrophy. Control and FSHD muscle samples showed normal standard histochemical reactions and histological examination revealed at most a very small percentage of fibres with evidence of active regeneration or necrosis. D4Z4 repeat copy numbers were determined from 5 µg of DNA from LCLs or PBMC samples embedded in agarose (InCert agarose, Cambrex) and digested with 20 U of *EcoRI* and *HindIII*, or *EcoRI* and *BlnI* and subject to PFGE and blot hybridisation with a 13E11 probe as described previously [30]. The 4q-type alleles were resolved from the 10q-type alleles by the *BlnI* sensitivity of the latter [11]. All human samples were obtained from individuals who signed consent forms for this study that were approved by the Institutional Review Board of Tulane Medical Center and the University of Mississippi Medical Center in Jackson.

7.3.2 Chromatin immunoprecipitation assays

About 0.5×10^6 cells were treated with 1% formaldehyde, sonicated in 1 ml of protease inhibitor-containing buffer, and then chromatin was immunoprecipitated essentially according to the manufacturer's specifications (Upstate Biotechnology) with a 1:100 dilution of an H4 acetylation-specific antibody prepared to H4 N-terminal peptide acetylated at lysine 5, 8, 12, and 16. The input DNA for comparison to the immunoprecipitates was an aliquot of the supernatant from each centrifuged sonicate (DNA size range, about 200-600 bp). The pre-clearing before addition of antibody and the collection of the immunoprecipitates was done for 3 hours at 4°C with constant agitation using 60 µl of salmon sperm DNA/protein A-agarose beads added to the 1 ml samples. The purified immunoprecipitated DNA was dissolved in 50 µl of 10 mM Tris-HCl (pH 8.0) / 0.1 mM EDTA. PCR primers were from 4q35 (*Table 7.1*) or standards and gave products in the size range of about 130-300 bp. The primers (5' to 3') for the standards and the annealing temperatures were as follows: satellite 2 (Sat2) from Chr 1, CATCGAATGGAAATGAAAGGAGTC and ACCATTGGATGATTGCAGTCAA, 58°C; satellite (Sat) from Chr 1, TCATTCCCACAAACTGCGTTG and TCCAACGAAGGCCACAAGA, 54°C; Chr 4 Sat, CTGCACTACCTGAAGAGGAC and GA-TGGTTCAACACTCTTACA, 52°C; pre-proinsulin exon 3 (pro-insulin), CCTGCAGAAGCGTGGCATT and CACAGACGGCA-

Table 7.1 4q35 PCR primers for chromatin immunoprecipitation.^a

General location of DNA sequence	DNA sequence GenBank accession no.	Primer sequences (5' to 3')	Product size (bp)	Q-PCR annealing temperature (°C)
D4Z4 subregions	region A; AF117653; 869-1017 ^b	ACGACGGAGGCGTGATTT AGTGTGGCCGGTTTGAA	203	58
	hhspm3 (putative <i>DUX4</i> promotor) immediately upstream of <i>DUX4</i> ; AF117653; 1598-1732 ^b	GGGCTCACCGCCATTCAT TGCACCTCAGCCGGACTGT	135	65
	<i>DUX4</i> double homeobox region; AF117653; 1859-2153 ^b	ACGGAGACTCGTTTGGAC TGGAAGCGATCCTTCTC	295	52
proximal sequences to the D4Z4 array present on 4q35	13E11 (p13E-11), 651 bp ^c proximal to the D4Z4 array at 4q35 and 10q26; AF117653	GACAGCAATAGTCCAGGCT CAGTTTGTGTCTGCTGAGAAG	218	58
	<i>DUX4C</i> upstream, putative promotor region of <i>DUX4C</i> ; -0.1 kb relative to TSP; AF146191	GCCCTTAGAAGACCTACC GGCTTCCAGTTTCCATAG	210	52
	<i>FRG2</i> promotor, 37 kb proximal to the D4Z4 array; -77 relative to TSP; AF146191	GGCTTACTCTTATTGGCTGAA TCTGGACACCATGACAGG	141	60
	<i>FRG1</i> , 125 kb proximal to the D4Z4 array; +57 relative to TSP; AF146191	TCTACAGAGACGTAGGCTGTCA CTTGAGCAGAGCTTGGTAG	178	65
	<i>ANT1</i> promotor, 4.8 Mb proximal to the D4Z4 array, -140 relative to the TSP; J04982	CAAGAACTCTCCACCGGC CTGTCCCTGCAACTTGGG	173	60
	<i>ANT1</i> 3' exon, last exon, 5305 bp downstream of the TSP; J04982	GTGCATTAAGTGGTCTTTATT TGTGGTTAATAGACTATTCTTA	234	60

a: Other primers are given in *Subjects and methods*. The *FRG1* [12] and *ANT1* primers [17] were previously described. TSP: transcription start point according to GenBank annotation except for *DUX4C*, whose putative TSP was inferred from that of *DUX4* and *FRG2*, which is annotated only in the homologous 10q *FRG2* sequence (AY028079). b: the positions of the PCR product with these primers is given relative to the *KpnI* site of a D4Z4 monomer. These entire subregions are shown in *Figure 7.1.B*. c: Where indicated, a 13E11 sequence that is 175 bp proximal to the start of the D4Z4 array was examined with the following primers and annealing temperature: CACGGACAAGGCCAGAGTT and TGCCTGTGAGTTTCGAATGC, 60°C.

CAGCAG, 58°C; G- globin exon 2, TCTACCCATGGACCCAGAGGT and CCACATGCAGCTTGTCACAGT, 58°C; AFP exon 1, GTTTCTCGTTGCTTACACAAAG and AGGCCAATAGTTTGTCTCACT, 59°C; albumin exon 4, GTTGCAACTCTTCGTGAAAC and TCACATCAACCTCTGGTCTC, 58°C; ADH5, intron 1, GCATAATTGAGCCTACGCC and GCAGAGGTGTTTGTACGTG, 59°C; and GAPD primers, 62°C [16].

7.3.3 Real-time quantitative PCR and determining the specificity of PCR primers

Q-PCR (Bio-Rad iCycler) was performed using SYBR green dye fluorescence and *Taq* polymerase (SYBR Green PCR Master Mix, Applied Biosystems; iCycler Optical analysis software, version 3, BioRad). Each 20 µl reaction contained 350 nM primers and 2 µl of undiluted immunoprecipitated DNA or a 1:100 dilution of input DNA. The specificity of the product was demonstrated by the presence of one peak in a melting curve and spot-checking by gel electrophoresis. To quantitate the amount of product DNA from the threshold cycle number, a standard curve for each primer-pair was generated for each PCR set from serial 2-fold dilutions of a reference mixture of sonicated human DNA. This curve was used to determine, in arbitrary

units, the relative amounts of input DNA and IP DNA. The slope of the standard curve for each test reaction was -3.3 ± 0.4 and the correlation coefficient was ≥ 0.99 .

To determine the specificity of primers for subregions of 4q35, a panel of human-rodent SCHs (Mini Mapping Panel 2, Version 3; Coriell Institute) was amplified using 50 ng of DNA and 25 pmol primers in a standard 25 μ l mixture with *Taq* DNA polymerase (Qiagen, 1 U per 25 μ l reaction) for 35 cycles.

7.3.4 RNA isolation and RT-PCR

For isolation of total RNA from muscle biopsy samples, about 5 mg of tissue that had been snap-frozen and stored at -80°C , was converted to a powder in liquid N_2 . RNA was isolated from approximately 5 mg tissue samples or 3×10^6 LCL cells, fibroblasts, or PBMC by standard techniques [8]. RNA (3 μ g) was digested with 4 U of RNase-free DNaseI (Invitrogen) in a 40 μ l reaction mixture for 15 minutes at room temperature and then for 30 minutes at 37°C . After addition of EDTA and incubation at 65°C for 10 minutes, 25 μ l was used for reverse transcription (RT) in a standard 50 μ l reaction mixture with M-MLV reverse transcriptase (300 U, Invitrogen) and random hexamer primers at 37°C for 60 minutes followed by heating to 95°C for 5 minutes. All RNA samples were confirmed to give no PCR product without the prior RT reaction. Q-PCR on 2 μ l of the cDNA samples was done as described except that the RT-PCR primers and annealing temperatures were as follows: *ANT1* exons 2 and 3, AGCGTGATTTCATGGTC and GCATCATCATTCTACTACG, 56°C ; *HPRT*, 52°C and β -actin, 60°C [31], and the *GAPD* primers referenced above. The 18S rRNA primers were used together with a competing oligonucleotide (3:7 molar ratio) to dampen the signal so that it was more like that from mRNAs (QuantumRNA 18S internal standards; Ambion). The same primers from the 5' exon of *FRG1* were used for RT-PCR and for ChIP assays (*Table 7.1*). For RT-PCR reactions the mean of triplicate amplifications, which usually differed by less than 20%, are shown. The standard curve for each primer pair was generated for each PCR set as described above except that a mixture of cDNAs was used as the standard.

7.4 Results

7.4.1 The chromosomal distribution of 4q35 sequences related to FSHD

To use histone hypoacetylation of different subregions of 4q35.2 as a marker for heterochromatin [26, 42], we first had to define the specificity of the primers for real-time quantitative PCR (Q-PCR) to be used in ChIP assays. This empirical determination of primer

specificity is particularly important because the 4q35 region of interest has much homology with the acrocentric p arms and pericentric regions [2, 24, 36] not adequately represented in DNA databases, as well as with 10q26. By amplification of template DNA from a panel of rodent-human SCHs, each containing a different single human chromosome, primer pairs from the following D4Z4 array-proximal sequences (*Table 7.1* and *Figure 7.1.A*) were evaluated: 13E11 (D4F104S1), a 0.8 kb non-genic sequence [24] previously called p13E-11 [63], which is 111 bp from the 4q and 10q D4Z4 array; the *FRG2* promoter; and the presumed promoter of the putative *DUX4C* gene (GenBank AF146191, AF117653, and AY028079). *DUX4C* is highly homologous to a putative gene, *DUX4* [18, 24, 29], present within each D4Z4 repeat, but the upstream regions for these putative genes are not homologous, and *DUX4C* is located as an isolated sequence 42 kb proximal to the D4Z4 array and in the opposite orientation [19, 36]. The Chr 10 hybrid gave the same size PCR products as did the Chr 4 hybrid with 13E11, *FRG2* and *DUX4* primers (data not shown), in accordance with the extensive homology between the 42 kb region proximal to the D4Z4 array on chromosomes 4 and 10 that contains these sequences (*Figure 7.1.A*) [5, 10, 19, 62]. All the other chromosomes had sequences amplified by the *DUX4C* primers. Chromosomes 18, 21, 22, X and Y, as well as 4 and 10, were amplified by the *FRG2* primers. The 13E11 primers amplified only Chr 4 and Chr 10 in accordance with its use as a hybridisation probe for molecular diagnosis of FSHD by pulsed-field gel electrophoresis (PFGE) of *EcoRI/HindIII* and *EcoRI/BlnI* digests [11, 54, 60]. Other than the D4Z4 arrays, only one fragment of 9.5 kb from an invariant Y chromosome sequence cross-hybridises with 13E11, and this fragment is apparently not amplified with our 13E11 primers. We also searched for a ~125-300 bp sequence in the D4Z4 repeat that might amplify only Chr 4 and 10. The six tested subregions of D4Z4 were as follows: *DUX4*, the above-mentioned sequence with the potential to encode a double homeobox gene, although it lacks a poly A site [4, 18, 24]; the putative *DUX4* promoter (hhspm3, homologous to sequences displaying sperm-specific DNA hypomethylation) [18, 24, 68]; the LSau subrepeat homology region, which is also present in the short arms of the acrocentric chromosomes and some pericentromeric regions [38]; region A, which is upstream of hhspm3; and two sequences in the vicinity of the *KpnI* site in D4Z4 (*Figure 7.1.B*). None of these subsequences proved specific in PCR, which is consistent with the cross-hybridisation of D4Z4 probes to the acrocentric chromosomes [2, 24]. However, primers for region A and hhspm3 amplified only acrocentric chromosomes and Chr 11 or Chr Y in addition to Chr 4 and Chr 10.

7.4.2 The D4Z4 array-adjacent 13E11 region at 4q35 and 10q26 showed levels of histone H4 acetylation typical of unexpressed euchromatin rather than of constitutive heterochromatin

The existence of homologous sequences to D4Z4 in many chromosomes precludes testing human cells for histone hypoacetylation, a marker for constitutive heterochromatin [26, 42], within the 4q and 10q D4Z4 arrays. However, we could analyse histone acetylation very close to the proximal end of the D4Z4 array using the above-mentioned 13E11 primers. Preliminary ChIP assays were done with eight commercially available antibodies to acetylated or methylated histones and the immunoprecipitates were analysed by Q-PCR for the relative amounts of precipitation of euchromatic and heterochromatic standards. We found that an antibody specific for acetylation of histone H4 (H4 Ac Ab) was the most informative for distinguishing unexpressed euchromatin standards (genes inactive in the cell types being analysed) and constitutive heterochromatin (satellite DNA-rich heterochromatin). For example, an antibody to histone H3 dimethylated at K9 gave no significant difference between unexpressed

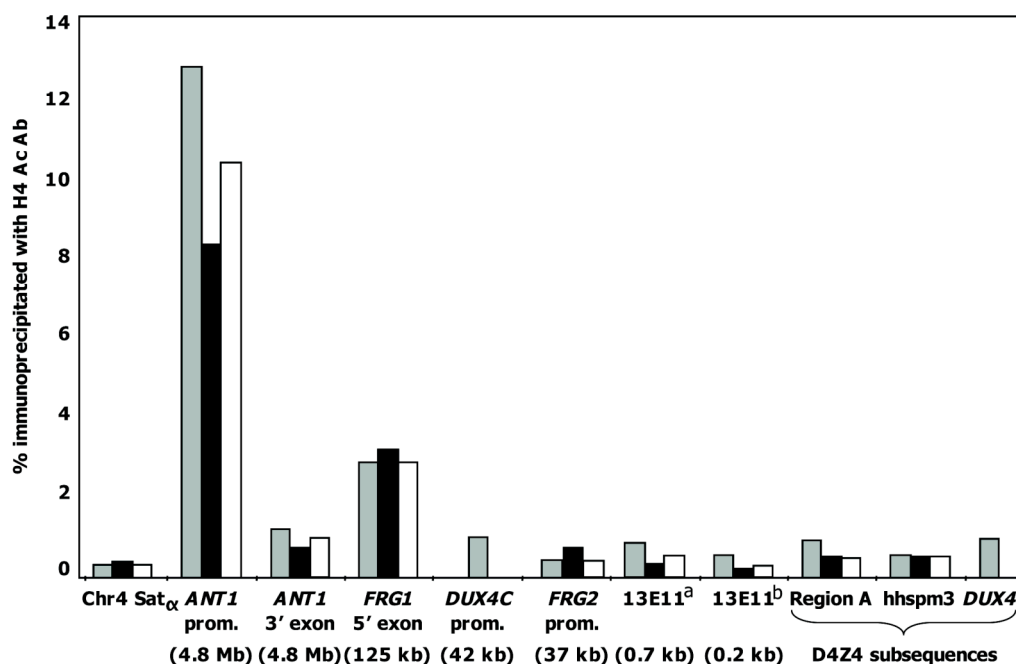


Figure 7.2

Comparison of histone H4 acetylation in different 4q35 regions in Chr 4-containing SCHs by ChIP with H4 Ac Ab. Gray, black and white bars represent the percentage immunoprecipitation of the given sequence in SCH GM11687, GM14193 and GM11448, containing the following human chromosomes: Chr 4; Chr 4 and 5; and Chr 4, 5, and 8, respectively. Q-PCR for the different immunoprecipitates were done on the same day with the same standard DNA mixture for quantitation and so are comparable. The copy numbers of the D4Z4 repeats in the 4q35 region of these SCHs are 40, 35 and 39, respectively. The results for *DUX4* and *DUX4C* subsequences are given only for GM11687, which does not have non-Chr 4 sequences that can be amplified by the corresponding primer pairs. In parentheses are shown the approximate distances of 4q35 proximal sequences from the D4Z4 array; 13E11^a and 13E11^b, two different 13E11 subsequences situated the indicated distances from the beginning of the D4Z4 array. The immunoprecipitation of chromatin from these SCHs was done in a second experiment with similar results (data not shown).

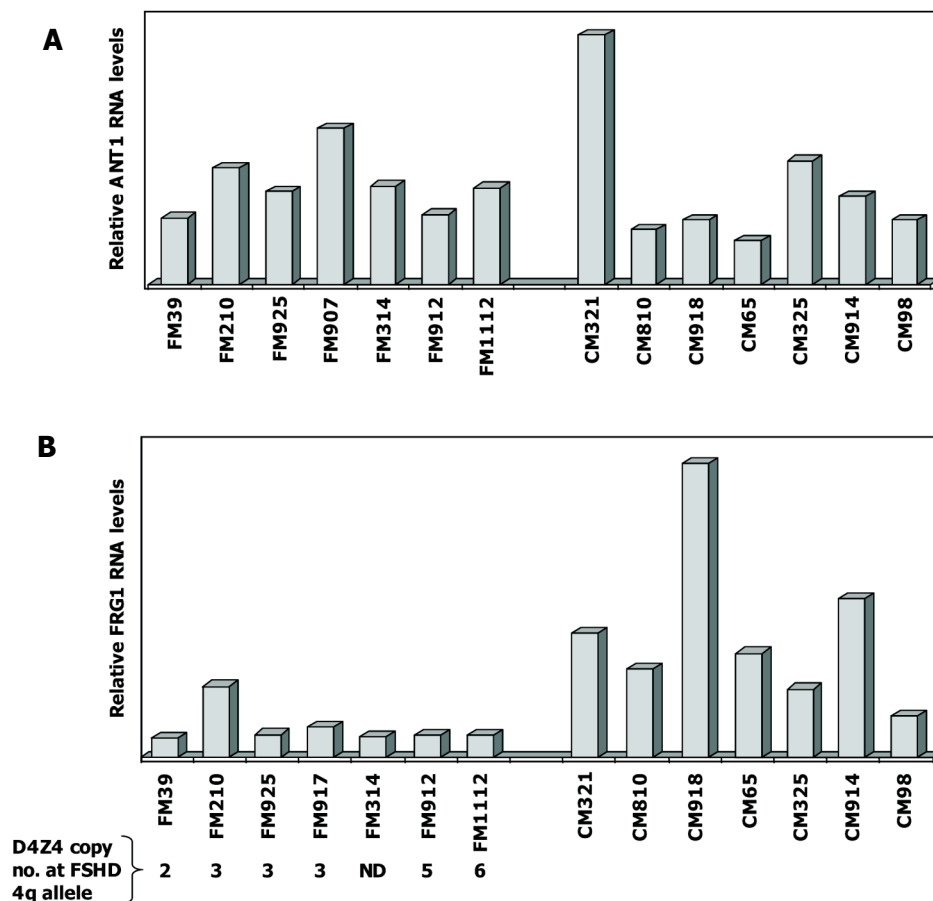


Figure 7.3

Relative *ANT1* and *FRG1* expression in FSHD and disease-control muscle samples determined by Q RT-PCR.

(A) *ANT1* RNA levels normalised with respect to 18S rRNA.

(B) *FRG1* RNA levels normalised with respect to 18S rRNA.

The D4Z4 copy number on the FSHD-causing 4q35 allele is indicated underneath each bar for the FSHD samples. The 4q35 D4Z4 copy number should be >10 for each Chr 4 homologue in the control samples. ND: not determined. The D4Z4 copy numbers were 4, 9, 10 and 36 for the 4q and 10q arrays of cells from this patient, but for technical reasons the position assignment could not be made. The units for RNA levels are arbitrary but all PCR determinations were made on the same day, quantitated with the same standard curve, and so are comparable.

euchromatin and constitutive heterochromatin in five ChIP assays using PCR primers for seven DNA sequences. This antibody did give an average of 5-fold less immunoprecipitation of expressed gene standards than unexpressed gene standards (data not shown). In contrast, in more than 20 ChIP assays with H4 Ac Ab on lymphoblastoid cell lines (LCLs), peripheral blood mononuclear cells (PBMC), and diploid fibroblasts, an average of about 2.5-fold less constitutive heterochromatin immunoprecipitated compared with unexpressed euchromatin, and the differences between these two types of chromatin were statistically significant ($P < 0.001$; *Table 7.2* and data not shown).

In H4 Ac Ab immunoprecipitates obtained from control PBMC, LCLs, and diploid fibroblast cell strains, the relative amounts of 13E11 DNA and DNA from heterochromatic and

euchromatic standards were evaluated by Q-PCR. The assays shown in *Table 7.2* involved the 13E11 primers that amplify the above-described DNA sequence 651 bp from the beginning of the D4Z4 array on both Chr 4 and Chr 10 (*Table 7.1* and *Figure 7.1.A*). If the 13E11 chromatin had behaved like constitutive heterochromatin, the ratio of the percent immunoprecipitation of this 13E11 sequence to that of constitutive heterochromatin regions should have been about 1 and the ratio to that of unexpressed euchromatic gene regions should have been about 0.4 for these control samples. Instead, the average ratios comparing the three groups of control samples were 3.7 ± 0.2 and 1.4 ± 0.2 , respectively (*Table 7.2*). Furthermore, FSHD PBMC and LCLs and samples with low copy numbers of D4Z4 at 10q arrays displayed no more H4 acetylation of the 13E11 region than did the controls (*Table 7.2*). We also designed another set of 13E11 primers that amplified a sequence only 175 bp (instead of 651 bp) away from the start of the array. From two control PBMC samples and three control LCLs, these primers gave the following average ratios of immunoprecipitated 13E11 to immunoprecipitated heterochromatin standards or unexpressed euchromatin standards: 2.2 ± 0.7 and 0.8 ± 0.3 , respectively. Therefore, this region, which is very close to the D4Z4 array, is also more like unexpressed euchromatin than constitutive heterochromatin in its histone acetylation in these control human cells.

7.4.3 Chromosome 4-containing somatic cell hybrids exhibited unexpressed euchromatin-like H4 acetylation levels at D4Z4 subsequences

In rodent-human SCHs containing as the only human Chr 4 (GM11687), Chr 4 and 5 (GM14193, or Chr 4, 5 and 8 (GM11448), we could analyse hhspm3, the D4Z4 subregion immediately upstream of the putative *DUX4* gene [18, 24], and another D4Z4 subregion about 0.7 kb proximal, which we call region A (*Figure 7.1*), without interference from homologous human sequences. Rodent genomes do not have D4Z4 homologous sequences [24] and they were not amplified with any of the tested D4Z4 subregion primers. In the Chr 4-only SCH, we could also study *DUX4*, the putative gene in D4Z4, as well as the presumed promoter of the putative *DUX4C* gene, a sequence 42 kb proximal to the D4Z4 array (*Figure 7.1.B*). In addition, we tested in all three SCHs the two aforementioned 13E11 subsequences and the promoter region of *FRG2* (*Figure 7.1.A*). These SCHs contain long D4Z4 arrays (35-40 D4Z4 repeats) at 4q35. The three D4Z4 subsequences in these SCHs displayed a low level of immunoprecipitation with H4 Ac Ab, but not as low as that of the Chr 4 Sat centromeric region (*Figure 7.2*). In contrast, considerable amounts of the *ANT1* promoter and *FRG1* exon 1 were in the H4 Ac Ab immunoprecipitate. Interestingly, the *FRG2* promoter and the putative *DUX4* and *DUX4C* promoters gave a low extent of immunoprecipitation suggesting that, unlike the *ANT1* and *FRG1* promoter or 5' region, they are not transcribed or poised for

transcription in these hybrids. In the region spanning the D4Z4 array to 42 kb proximal, there was not a gradient of H4 acetylation with the lowest levels in the D4Z4 array, as would be expected if there were spreading of heterochromatinisation seeded at long D4Z4 arrays.

7.4.4 FSHD and control samples showed similar expressed gene-like histone H4 acetylation levels in FRG1 and ANT1 at 4q35

We looked for evidence in control cells of histone-hypoacetylated repressive chromatin structures at the FSHD 4q35-specific candidate genes *ANT1* and *FRG1* using the same immunoprecipitates analysed for H4 acetylation in the 13E11 region. *FRG1* exon 1 and the *ANT1* promoter in control PBMC, LCLs, and fibroblasts resembled expressed genes in their high degree of immunoprecipitation with the H4 Ac Ab (**Table 7.3**). Similarly, hyperacetylation of the *ANT1* promoter and *FRG1* exon 1 was seen in Chr 4 SCHs but the ratio of percent immunoprecipitation of *FRG1* exon 1 to that of the *ANT1* promoter was higher for lymphoid cells than for the SCHs (**Table 7.3** and **Figure 7.2**). In both control human cells and the SCHs, there was less H4 acetylation in the last *ANT1* exon than at the *ANT1* promoter, as has been found for some expressed genes [40]. When comparing H4 acetylation in any of these regions in FSHD and control PBMC or LCLs, no consistent differences were observed between FSHD and non-FSHD samples nor was there a discernible effect of the exact number of tandem D4Z4 copies on this H4 acetylation. We will extend our ChIP analyses to FSHD and control myoblasts (which have just become available) and to chromatin proteins other than histones and include well-characterized myoblast cultures induced to differentiate *in vitro* as a model for *in vivo* muscle fibres.

7.4.5 FSHD skeletal muscle samples did not show higher levels of FRG1 or ANT1 transcripts than controls

We found that muscle biopsy samples were not amenable to ChIP assays to test predictions about histone acetylation in 4q35 subregions. To examine another prediction of the loss-of-PEV hypothesis, *FRG1* and *ANT1* transcript levels in seven FSHD skeletal muscle biopsies and seven disease-control biopsies were analysed by Q-PCR. First we demonstrated, as previously reported [17], that there were much higher average levels of *ANT1* RNA in control muscle samples than in four control PBMC samples or three LCLs assayed the same day and normalised to 18S rRNA (19, 0.09, and 0.7, respectively, in arbitrary units). We observed more similar levels of *FRG1* RNA from control muscle biopsies, LCLs, and PBMC samples (2.7, 2.7, and 0.4, respectively). In the lymphoid samples from FSHD patients compared with the analogous samples from controls, no significant differences were seen in normalised *ANT1* or

FRG1 transcript levels (data not shown). That *FRG1* and *ANT1* RNAs were expressed in LCLs and PBMC is consistent with the H4 hyperacetylation in their 5' gene regions (*Table 7.3*).

It was recently reported that there were, on average, about 10-fold higher *ANT1* transcript levels in FSHD versus control skeletal muscle biopsies by end-point RT-PCR with normalization to the glyceraldehyde phosphate dehydrogenase (*GAPD*) RNA levels [17]. However, by real-time RT-PCR, we found no consistent differences in steady-state *ANT1* RNA levels normalised to those for 18S rRNA when FSHD and disease-control muscle were compared (*Figure 7.3.A*; averages of 6.7 ± 2.0 and 6.6 ± 4.9 , arbitrary units $\pm$ SD, respectively). Ribosomal RNA has been considered the best internal standard RNA for quantitating transcript levels in several recent reports [22, 48], although there is still controversy about which constitutively expressed mammalian RNA quantitation standard is optimal. The FSHD sample with the lowest D4Z4 copy number (two copies, FM39) in the deletion-associated 4q35 region did not show higher levels of *ANT1* RNA and *FRG1* RNA compared with the FSHD samples having the highest deletion-associated 4q D4Z4 copy numbers (six copies, FM1112; 5 copies, FM912), as would have been predicted by the loss-of-PEV hypothesis and the above-mentioned study [17]. Because it has been suggested that averaging the results from several standards may be better than using just one [55], we employed three constitutively expressed mRNA standards in addition to 18S rRNA. The average values for *ANT1* RNA normalized to *GAPD* RNA were 10.4 ± 4.8 (FSHD) and 5.9 ± 4.9 (control); to *HPRT* RNA, 2.9 ± 1.4 (FSHD) and 3.1 ± 2.2 (control); and to α -actin, 6.6 ± 3.9 (FSHD) and 4.4 ± 1.9 (control). The mean ratios derived from averaging the levels normalised to each of the four RNA standards were 6.7 ± 3.0 and 5.0 ± 1.5 for FSHD and control skeletal muscle biopsies, respectively. There was no significant difference between the two groups of muscle samples ($P > 0.8$ for normalization to 18S rRNA or *HPRT*, $P > 0.1$ for normalisation to α -actin and *GAPD*, respectively).

For *FRG1* RNA, the FSHD skeletal muscle samples unexpectedly [17] displayed lower levels than did the control samples. The average *FRG1* RNA levels (arbitrary units $\pm$ SD) normalized with respect to 18S rRNA were 1.2 ± 0.7 and 5.0 ± 3.3 for FSHD and control muscle samples, respectively (*Figure 7.3.B*). When the values for *FRG1* RNA were normalised to *GAPD* RNA levels, they were 6.7 ± 2.3 and 14 ± 9 , respectively, and for normalisation to *HPRT* RNA levels, they were 0.5 ± 0.4 and 2.1 ± 1.6 , respectively. The differences in the average *FRG1* RNA level for FSHD muscle samples versus control samples were statistically significant whether the samples were normalised with respect to 18S rRNA, *GAPD* RNA, or *HPRT* RNA ($P = 0.01$, 0.05, or 0.02, respectively).

Table 7.2 ChIP analysis of histone H4 acetylation at 13E11, a non-gene region adjacent to the 4q and 10q D4Z4 arrays.^a

Sample type	Sample ID	D4Z4 copy number in 4q- or 10q-type arrays ^b	% IP heterochromatin standards				% IP unexpressed gene standards				% IP expressed gene standards		Relative % IP standards ^c		IP 13E11	Relative % IP 13E11 vs. standards ^c	
			Chr 1 4q-type	Chr 1 10q-type	Chr 4	Chr 4	G- γ globin	Pro- insulin	AFP	Alb.	GAPD	ADH5	unexpressed genes vs. heterochromatin	expressed genes vs. heterochromatin		13E11 vs. 13E11	unexpressed genes vs. 13E11
Control PBMC	CB56	13, 18	13, 20	0.39	0.79	0.32	0.79	2.00	1.20	1.80	8.60	11.00	2.8	20.0	1.0	2.0	0.7
	CB56	13, 18	0.28	0.13	0.57	0.28	0.88	ND	0.99	5.60	5.30	2.1	17.0	1.7	5.2	2.4	
	CB523	34, 39	0.59	0.58	0.63	0.86	1.70	1.20	0.90	7.80	7.30	2.0	13.0	2.1	3.5	1.8	
	CB914	21, 34	0.29	0.21	0.39	0.70	0.67	1.60	1.10	2.70	12.00	3.3	25.0	1.4	4.7	1.4	
			average $\pm$ SD										2.6 $\pm$ 0.6	19 $\pm$ 5		3.8 $\pm$ 1.4	1.6 $\pm$ 0.7
FISHD PBMC	FB1112	6, 18	16, 25	0.09	0.21	0.32	0.57	0.71	0.56	0.42	3.80	4.40	2.7	20.0	0.6	3.0	1.1
	FB210	3, 27	0.22	0.37	0.48	0.65	0.73	0.47	0.32	3.40	3.30	1.5	9.4	0.3	0.9	0.9	
	FB314	ND ^d	0.23	0.36	0.34	1.10	0.94	0.83	0.94	4.90	4.20	3.1	15.0	1.1	3.5	1.2	
	FB925	3, 25	0.28	0.52	0.42	0.90	1.60	0.74	0.83	4.80	5.00	2.4	12.0	0.9	2.2	0.9	
			average $\pm$ SD										2.4 $\pm$ 0.7	14 $\pm$ 5		2.4 $\pm$ 1.2	0.9 $\pm$ 0.3
Control LCLs	MG	19, 22, 31	7 ^b	0.80	0.70	0.44	1.40	1.80	2.30	1.30	10.00	12.00	2.6	17.0	2.7	4.2	1.6
	ND	24, 39	19, 37	0.70	0.97	0.60	2.40	2.60	1.30	1.90	11.00	9.20	2.8	13.0	1.4	1.8	0.7
	423	19, 27	7, 15	0.57	0.65	0.67	1.70	0.80	1.40	0.99	16.00	11.00	1.9	21.0	2.8	4.4	2.3
			average $\pm$ SD										2.4 $\pm$ 0.5	17 $\pm$ 4		3.5 $\pm$ 1.4	1.5 $\pm$ 0.8
FISHD LCLs	305	6, 36	9, 9	0.80	0.97	1.20	2.00	2.00	1.70	1.60	15.00	11.00	1.8	13.0	1.5	1.5	0.8
	WJ 1952	6, 27	12, 13	1.10	1.50	1.30	4.00	2.30	ND	2.60	11.00	10.00	2.2	8.1	2.0	1.5	0.7
	GM 17724	6, 18	16, 25	1.40	1.30	0.98	3.90	4.30	1.50	1.20	14.00	9.40	2.2	10.0	3.4	2.8	1.3
			average $\pm$ SD										2.1 $\pm$ 0.3	10 $\pm$ 3		1.9 $\pm$ 0.7	0.9 $\pm$ 0.3
Control fibroblasts	FB1	30 ^e	27, 19, 18	1.10	1.50	2.00	5.00	2.60	ND	2.60	29.00	17.00	2.2	15.0	2.0	2.3	0.6
	FB2	ND	ND	0.87	0.62	0.81	1.60	1.10	1.90	2.20	7.20	8.90	2.2	11.0	2.0	2.6	1.2
	FB3	ND	ND	0.41	0.64	0.57	1.50	1.50	2.20	1.70	9.10	15.00	3.1	22.0	3.3	6.1	1.9
			average $\pm$ SD										2.5 $\pm$ 0.5	16 $\pm$ 6		3.7 $\pm$ 2.1	1.2 $\pm$ 0.7

a: ChIP analysis was done on the indicated samples of peripheral blood mononuclear cells (PBMC), lymphoblastoid cell lines (LCLs), or diploid fibroblast cell strains. The average of duplicate real-time PCR determinations of DNA in the immunoprecipitated (IP) chromatin is given; duplicates generally varied by <20%. SD: standard deviation; ND: not determined. The CB56 PBMC sample was obtained twice and immunoprecipitated on two different days; the results from the duplicate immunoprecipitations are shown. Note that the percentage IP of different DNA sequences in a given sample can be compared, but for different samples, only the relative percentage IP of different sequences should be compared unless they were immunoprecipitated on the same day and concurrently amplified. b: The number of 10q-type D4Z4 arrays (containing a *Bhl* site in each monomer) and 4q-type (not containing a *Bhl* site in each monomer) was determined by PFGE. Sometimes a 10q-type array is found on 4q, and *vice versa*, even in the unaffected population due to the high frequency of recombination between 4q35 and 10q26. Therefore, while the total number of 4q-type arrays is four, the number of 4q-type arrays can vary from zero to four, as can the number of 10q-type arrays. FISHD is linked only to a short D4Z4 array (1-10 copies) at 4q35. c: The relative percentage IP for a given immunoprecipitate refers to the % IP for the first indicated sequence divided by the average percentage IP for the second set of sequences. d: This sample presented technical problems: the lengths of the D4Z4 arrays were 4, 9, 10 and 36 tandem copies of the repeat, but it is uncertain which of these are located on 4q and which are on 10q.

Table 7.3 ChIP analysis of histone H4 acetylation in FSHD candidate genes *FRG1* and *ANT1*, 0.1 and 4.8 Mb proximal to the D4Z4 array on 4q35.^a

Sample type	Sample ID	4q-type D4Z4 copy numbers	Heterochromatin and gene standards				<i>FRG1</i> exon 1		<i>ANT1</i> promoter		<i>ANT1</i> 3' exon	
			average % IP heterochromatin standards	average % IP unexpressed gene standards	average % IP expressed gene standards	relative % IP unexpressed vs. expressed genes	% IP	relative % IP vs. unexpressed genes	% IP	relative % IP vs. unexpressed genes	% IP	relative % IP vs. unexpressed genes
Control PBMC	CB56	13, 18	0.50 ± 0.25	1.4 ± 0.5	9.8 ± 1.4	7.0	18.0	13.0	8.6	6.1	5.6	4.0
	CB56	13, 18	0.33 ± 0.20	0.7 ± 0.4	5.5 ± 0.2	7.9	12.0	17.0	12.0	17.0	4.9	7.0
	CB523	34, 39	0.60 ± 0.03	1.2 ± 0.4	7.6 ± 0.4	6.3	17.0	14.0	7.9	6.6	2.6	2.2
	CB914	21, 34	0.30 ± 0.10	1.0 ± 0.4	7.4 ± 6.6	7.4	6.7	6.7	1.8	1.8	0.9	0.9
			average ± SD			7.2 ± 0.7		12.7 ± 4.4		7.9 ± 6.5		3.5 ± 2.6
FSHD PBMC	FB1112	6, 18	0.21 ± 0.11	0.57 ± 0.12	4.1 ± 0.3	7.2	4.4	7.7	2.7	4.7	1.4	2.5
	FB210	3, 27	0.36 ± 0.13	0.54 ± 0.18	3.4 ± 0.1	6.3	5.3	9.8	1.6	3.0	0.8	1.5
	FB314	ND ^b	0.31 ± 0.10	0.95 ± 0.1	4.6 ± 0.5	4.8	5.1	5.4	7.9	8.3	1.5	1.6
	FB925	3, 25	0.41 ± 0.10	1.0 ± 0.4	4.9 ± 0.1	4.8	9.2	9.0	13.0	13.0	1.7	1.7
			average ± SD			5.8 ± 1.2		8.0 ± 1.9		7.2 ± 4.3		1.8 ± 0.4
Control LCLs	MG	19, 22, 31	0.65 ± 0.20	1.7 ± 0.5	11 ± 1.1	6.5	8.0	4.7	7.2	4.2	2.2	1.3
	ND	24, 39	0.76 ± 0.20	2.1 ± 0.6	10 ± 1.4	4.8	9.5	4.5	6.3	3.0	1.2	0.6
	423	19, 27	0.63 ± 0.05	1.2 ± 0.4	13 ± 3.6	11.0	12.0	10.0	ND	ND	2.6	2.2
			average ± SD			7.4 ± 3.2		6.4 ± 3.1		3.6 ± 0.9		1.3 ± 0.8
FSHD LCLs	305	6, 36	0.99 ± 0.20	1.8 ± 0.2	13 ± 2.6	7.2	8.7	4.8	7.0	3.9	3.3	1.8
	WJ 1952	7, 27	1.3 ± 0.1	3.0 ± 0.9	11 ± 0.8	3.5	16.0	5.3	6.3	2.1	3.1	1.0
	GM 17724	6, 18	1.2 ± 0.2	2.7 ± 1.6	12 ± 1.1	4.3	9.1	3.4	7.0	2.6	2.9	1.1
			average ± SD			5.0 ± 2.0		4.5 ± 1.0		2.9 ± 0.9		1.3 ± 0.5
Control fibroblasts	Fib1	30	1.5 ± 0.4	3.4 ± 3.5	23 ± 13	6.8	12.0	3.5	14.0	4.1	3.9	1.1

a: The percentage immunoprecipitation of the indicated *FRG1* and *ANT1* regions was determined by Q-PCR of the same immunoprecipitates as described in *Table 7.2*, whose legend gives the abbreviations and the method for calculating the relative percentage IP. The percentage IP for the standards is summarised from *Table 7.2* for comparison with the percentage IP of *FRG1* and *ANT1* regions. For the average percentage IP from different standards and the average relative percentage IP from different immunoprecipitations, the standard deviation is given. b: For this patient, the D4Z4 arrays contained 4, 9, 10 and 36 repeat units, but it is uncertain which of these are on 4q and on 10q for technical reasons.

7.5 Discussion

Almost all cases of FSHD show linkage to short D4Z4 arrays in the subtelomeric region of 4q although there is no association with short D4Z4 arrays at subtelomeric 10q [54] despite their near sequence identity and similar copy-number polymorphism as well as their homology proximally and distally (*Figure 7.1*) [11, 20, 31]. Given the threshold effect of the D4Z4 copy number at 4q35 on the disease, the dominant nature of FSHD is most easily explained by inappropriate upregulation of one or more 4q35 disease-causing genes rather than by haploinsufficiency. No genes have been found in the ~15-25 kb region between the 4q35 D4Z4 array and the telomeric hexanucleotide repeat, although the region is only partly sequenced [19]. Therefore, the FSHD gene (or genes) is likely to be proximal, rather than distal, to the 4q35 D4Z4 array. The 161 kb region proximal to the D4Z4 array at 4q35 is very gene poor. The high homology between 4q and 10q for 42 kb proximal to the D4Z4 array and recent evidence from patients with deletions extending this far [32] make it unlikely that this region is involved in FSHD. Because the closest identified gene proximal to this region is *FRG1*, 125 kb from the D4Z4 array at 4q35, the FSHD 4q35 target gene is probably quite far from the D4Z4 array. The loss-of-PEV hypothesis, which involves differential heterochromatinisation, was proposed to explain how a shortened D4Z4 array exerts this *cis* effect at a distance. However, no subregion of 4q35 has been tested for a specific attribute of heterochromatin. We examined histone hypoacetylation as a marker for constitutive heterochromatin [26, 42] (*Table 7.2*), in the vicinity of D4Z4 arrays at sequences that could be studied in human cells (13E11, *FRG1*, and *ANT1*) or, because of homology elsewhere, only in somatic cell hybrids (D4Z4, *FRG2*, and *DUX4C*; *Figure 7.1.A*).

In PBMC, LCLs, and fibroblasts from unaffected individuals, 13E11 chromatin only 651 or 175 bp proximal to the long D4Z4 arrays exhibited histone H4 acetylation levels typical of unexpressed euchromatin rather than of constitutive heterochromatin (*Table 7.2*). This makes it highly unlikely that D4Z4 arrays are sources of *cis*-spreading heterochromatinisation. Although, we were unable to immunoprecipitate constitutively expressed gene standards from muscle biopsies to study the tissue involved in FSHD, several factors suggest that the conclusions from the above cell populations are applicable to skeletal muscle. Firstly, constitutive heterochromatin, e.g. centromeric and juxtacentromeric heterochromatin, does not show tissue-specific differences in heterochromatinisation. PEV is associated with constitutive heterochromatin [65], which displays tight packing and distinct nucleosome spacing [1, 34, 49, 57], rather than with the less highly condensed structure of transcriptionally inactive euchromatin. Secondly, we did not see tissue-specific differences between LCLs (rapidly dividing, immortalized B-cell lines), PBMC (mostly quiescent T cells), and diploid skin fibroblasts in the

extent of H4 acetylation at 13E11. Furthermore, there were similar steady-state concentrations of transcripts from *FRG1*, the closest FSHD candidate gene to the D4Z4 array, in control skeletal muscle samples and LCLs. That *FRG1* RNA was present at considerable levels in control skeletal muscle indicates that there is not a region of repressive heterochromatin spreading far from the normal-sized D4Z4 arrays at 4q35 in skeletal muscle, the target tissue for FSHD. Also inconsistent with the loss-of-PEV model was our finding that in lymphoid cells and fibroblasts from unaffected individuals, the *FRG1* exon 1 and *ANT1* promoter regions were not hypoacetylated in H4, and no gradient of H4 acetylation proportional to the distance from D4Z4 was seen (**Table 7.3**). Moreover, there was no increase in H4 acetylation in PBMC or LCL samples from affected versus unaffected individuals in the *FRG1* and *ANT1* regions (**Table 7.3**).

ChIP assays of H4 acetylation at the *FRG2* promoter, the putative *DUX4C* promoter, 13E11, and D4Z4 monomer subregions in human-rodent SCHs also did not provide support for the loss-of-PEV hypothesis for FSHD. Rodent-human SCHs have been shown to often maintain the chromosome structure of the respective regions in human cells [6, 67]. In hybrids containing a normal Chr 4 or Chr 4 plus one or two other non-interfering human chromosomes, the long D4Z4 array did not exhibit the lowest amount of H4 acetylation (**Figure 7.2**). Also, there was not a distance-dependent gradient of H4 hypoacetylation originating from the D4Z4 array and extending to proximal genes at 4q35. While our data on H4 acetylation at long D4Z4 arrays in SCHs argue against this region being heterochromatic, and therefore being capable of losing this heterochromatin structure upon contraction of the array in FSHD, it is quite possible that there are other changes in histone modification, not associated with a heterochromatin-to-euchromatin transition that take place upon shortening of the D4Z4 array from >10 to 1-10 copies of the repeat.

A linear gradient of heterochromatinisation initiating at constitutive heterochromatin or transgene repeats had been postulated as the basis for PEV in *Drosophila* (reviewed in [56]). However, this model of linear propagation of heterochromatinisation is currently out of favour because it is difficult to reconcile with the *trans* effects (transvection) seen in some cases of *Drosophila* PEV; with PEV operating over very large distances (up to several Mb) from constitutive heterochromatin; and with the complexity of some of the PEV-generating rearrangements [50, 56]. Nonetheless, heterochromatinisation is implicated in most cases of *Drosophila* PEV [50, 65]. That we did not find evidence for heterochromatin-like hypoacetylation about 1 kb from normal-length D4Z4 arrays or within these arrays argues against a loss-of-PEV mechanism for the effect of reduced D4Z4 copy number at 4q35.

Nonetheless, evidence consistent with the loss-of-PEV hypothesis for FSHD was recently reported by Gabellini *et al.* [17] from RT-PCR analysis of levels of *ANT1*, *FRG1*, and

FRG2 RNA from several 4q35 genes. Of these genes, *ANT1* [12, 17, 61] is the best candidate because it encodes an adenine nucleotide translocator that has been implicated in a myopathy (progressive external ophthalmoplegia), has been reported to be predominantly expressed in heart and skeletal muscle [17], and is specific for 4q35. However, it is very distant from the D4Z4 array (*Figure 7.1.A*). In the study of Gabellini and co-workers, which did not involve real-time quantitative RT-PCR, *ANT1*, *FRG1*, and *FRG2* transcript levels were an average of about 10-, 25-, and 65-fold higher, respectively, in FSHD versus control skeletal muscle biopsies. Also, Gabellini *et al.* [17] noted that the ratio of expression of the above genes in FSHD skeletal muscle relative to control muscle was higher for the genes closer to the 4q35 D4Z4 array. That finding was considered in agreement with a gradient of spreading of repression from a normally long D4Z4 array. Lastly, they observed that the lower the copy number of the D4Z4 disease-associated repeat, the higher the *FRG2* transcript level determined by RT-PCR after digestion of the PCR product with restriction endonucleases that allowed homoduplexes of 4q35 and 10q26 RT-PCR products to be distinguished. However, in skeletal muscle samples examined by Q-PCR, we did not find any significant FSHD-associated increase in the level of either *FRG1* or *ANT1* transcripts upon normalisation with four different internal RNA standards. For *FRG1*, we obtained the opposite result, a modest but significant decrease in transcript levels (*Figure 7.3*). By RNA-based single-strand conformation polymorphism analysis, van Deutekom *et al.* [12] also found no evidence for increased levels of *FRG1* transcripts from the disease-associated allele in skeletal muscle biopsies of FSHD patients. The report of Gabellini *et al.* [17] involved three FSHD muscle biopsies as compared with seven in the present study and only one normalisation standard, GAPD, instead of four in our study. Most importantly, they used end-point PCR without providing evidence from reactions with different dilutions of the cDNA or different numbers of cycles to show that they were in the linear response range with respect to template cDNA concentration for each test and reference gene. Real-time PCR, as in the present study, avoids the pitfalls of misleading comparisons of product yields due to the exponential nature of PCR and the possible plateauing of the reaction for the test sequence or normalisation standards. A caveat in any biochemically based expression analysis of FSHD versus control muscle is that we cannot discount the possibility that there is overexpression of a 4q35 gene in a small percentage of nuclei in the FSHD muscle fibres or satellite cell precursors, which starts a disease-inducing signal transduction cascade. Nonetheless, our data strongly argue against the previous model of an FSHD-linked loss of postulated heterochromatin spreading from a D4Z4 array to proximal 4q35 genes.

We propose that the intrachromosomal communication at an FSHD-causing D4Z4 array and an FSHD target gene in *cis* occurs by looping rather than by a loss of progressive spreading of heterochromatin. Because of the dearth of identified genes in the vicinity of the

array [20, 31], this looping may be of a very long-distance nature involving a gene more than 160 kb away. Such abnormal looping interactions with a short D4Z4 array at 4q35 might inappropriately upregulate transcription of the 4q35 target gene by direct chromatin-to-chromatin delivery of a positive transcription factor, by locally altering the structure of chromatin at the target gene's promoter, and/or by influencing the association of the target gene region with the nuclear scaffold or nuclear membrane. There is evidence for transcription regulatory looping interactions between enhancers and promoters [52], locus control regions and the rather distant genes that they positively control [7, 52], and polycomb response elements and associated promoters up to 100 kb away [39]. We further hypothesise that disease-associated long-distance loops involving the D4Z4 repeat array form only when a specific interaction between D4Z4 repeats at least 33 kb apart, i.e. intra-array looping, does not sequester the array. The intra-array looping would sequester both moderate-length and very long D4Z4 arrays. While there is no exact precedent for the type of alternative loop formation that we are proposing, there may be some parallels in the way certain insulators affect promoter-enhancer interactions via inferred chromatin loops [28, 59] as well as in LCR interactions with alternate promoters [7].

There could be constraints on the higher-order organisation of chromatin in the D4Z4 array that could account for the threshold effect of D4Z4 copy number at 4q35 in FSHD. The nature of the higher-order structure of chromatin is still poorly understood but the extraordinarily high GC content of the D4Z4 repeat (73%) compared with the overall GC content of human DNA (42%) might facilitate the hypothesised intra-array interactions between sufficiently distant tandem copies of the D4Z4 repeat. As a corollary of this model of alternative looping structures, it is predicted that D4Z4 arrays with >10 copies efficiently form the stable intra-array interactions, those with 5-10 copies of the D4Z4 repeat at 4q35 sometimes establish less stable intra-array interactions, and arrays with 1-4 copies almost never form these interactions. This would explain the above-mentioned increased severity of symptoms and decreased age at diagnosis associated with very short D4Z4 arrays at 4q35 [35, 46, 51]. Because all examined D4Z4 patients have at least one copy of the D4Z4 repeat, we propose that it is the D4Z4 sequence itself in a short array at 4q35 which activates inappropriate expression by long-distance looping interactions. This D4Z4 interaction could involve the *DUX4* putative promoter region, which may not be active itself *in vivo* (hhspm3; **Figure 7.2**) but apparently binds four nuclear proteins [17, 18]. These are the transcription factor Sp1, the transcription activator/repressor YY1, the DNA helicase-containing nucleolin, and the chromatin architectural protein HMGB2, all of which might enable D4Z4 to upregulate transcription of a *cis* promoter as a result of looping interactions. Moreover, analysis of stable transfectant clones of a mouse myoblast cell line suggests that several tandem D4Z4 repeats can influence

expression of genes located elsewhere in the genome [66]. In summary, a model involving alternative looping (intra-array or between the D4Z4 array and a distant *cis* transcription control region) now appears to be a more attractive hypothesis for how the copy number of tandem 3.3 kb D4Z4 repeats at 4q35 controls the disease phenotype than does a PEV-like model involving constitutive heterochromatin.

7.6 Acknowledgements

We thank M Ray, the Coriell Institute, Dr. A Romeo and F Brown for help in obtaining samples and K Jackson for sharing unpublished results. This research was supported in part by FSH Society grant FSHS-MB-06 and NIH grant R21 AR48315.

7.7 References

1. Allshire RC, Javerzat JP, Redhead NJ, and Cranston G (1994) Position effect variegation at fission yeast centromeres. *Cell* 76(1): 157-169.
2. Altherr MR, Bengtsson U, Markovich RP, and Winokur ST (1995) Efforts toward understanding the molecular basis of facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2: 32-38.
3. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, Wielen van der M, Rasmussen K, and Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2: 39-44.
4. Beckers M, Gabriels J, Maarel van der SM, Vriese de A, Frants RR, Collen D, and Belayew A (2001) Active genes in junk DNA? Characterization of *DUX* genes embedded within 3.3 kb repeated elements. *Gene* 264(1): 51-57.
5. Cacurri S, Piazzo N, Deidda G, Vigneti E, Galluzzi G, Colantoni L, Merico B, Ricci E, and Felicetti L (1998) Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric *KpnI* repeat units implicated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 63: 181-190.
6. Carrel L, Cottle AA, Goglin KC, and Willard HF (1999) A first-generation X-inactivation profile of the human X chromosome. *Proc Natl Acad Sci USA* 96(25): 14440-14444.
7. Carter D, Chakalova L, Osborne CS, Dai YF, and Fraser P (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet* 32(4): 623-626.
8. Chomczynski P and Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162(1): 156-159.
9. Coffee B, Zhang F, Warren ST, and Reines D (1999) Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat Genet* 22(1): 98-101.
10. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, and Felicetti L (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3: 155-167.

11. Deidda G, Cacurri S, Piazzo N, and Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet* 33(5): 361-365.
12. 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.
13. Ding H, Beckers MC, Plaisance S, Marynen P, Collen D, and Belayew A (1998) Characterization of a double homeodomain protein (DUX1) encoded by a cDNA homologous to 3.3 kb dispersed repeated elements. *Hum Mol Genet* 7(11): 1681-1694.
14. Dorer DR and Henikoff S (1997) Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* 147(3): 1181-1190.
15. Ehrlich M, Buchanan KL, Tsien F, Jiang G, Sun B, Uicker W, Weemaes CM, Smeets D, Sperling K, Belohradsky BH, *et al.* (2001) DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. *Hum Mol Genet* 10(25): 2917-2931.
16. Fulmer-Smentek SB and Francke U (2001) Association of acetylated histones with paternally expressed genes in the Prader-Willi deletion region. *Hum Mol Genet* 10(6): 645-652.
17. 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.
18. Gabriels J, Beckers MC, Ding H, Vriese de A, Plaisance S, Maarel van der SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al.* (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* 236(1): 25-32.
19. 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.
20. Geel van M, Heather LJ, Lyle R, Hewitt JE, Frants RR, and Jong de PJ (1999) The FSHD region on human chromosome 4q35 contains potential coding regions among pseudogenes and a high density of repeat elements. *Genomics* 61(1): 55-65.
21. Gilbert SL and Sharp PA (1999) Promoter-specific hypoacetylation of X-inactivated genes. *Proc Natl Acad Sci USA* 96(24): 13825-13830.
22. Goidin D, Mamessier A, Staquet MJ, Schmitt D, and Berthier-Vergnes O (2001) Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal Biochem* 295(1): 17-21.
23. Hecht A, Strahl-Bolsinger S, and Grunstein M (1996) Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* 383(6595): 92-96.
24. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, Deutekom van JC, Francis F, Sharpe PT, Hofker M, *et al.* (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3(8): 1287-1295.
25. Hwang KK, Eissenberg JC, and Worman HJ (2001) Transcriptional repression of euchromatic genes by *Drosophila* heterochromatin protein 1 and histone modifiers. *Proc Natl Acad Sci USA* 98(20): 11423-11427.
26. Johnson CA, O'Neill LP, Mitchell A, and Turner BM (1998) Distinctive patterns of histone H4 acetylation are associated with defined sequence elements within both heterochromatic and euchromatic regions of the human genome. *Nucleic Acids Res* 26(4): 994-1001.
27. Kimura A, Umehara T, and Horikoshi M (2002) Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nat Genet* 32(3): 370-377.

28. Kuhn EJ, Viering MM, Rhodes KM, and Geyer PK (2003) A test of insulator interactions in *Drosophila*. *EMBO J* 22(10): 2463-2471.
29. 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.
30. Lemmers RJLF, Kievit de P, Geel van M, Wielen van der MJ, Bakker E, Padberg GW, Frants RR, and Maarel van der SM (2001) Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. *Ann Neurol* 50(6): 816-819.
31. Lemmers RJLF, Kievit de P, Sandkuijl L, 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.
32. Lemmers RJLF, Osborn M, Haaf T, Rogers M, Frants RR, Padberg GW, Cooper DN, Maarel van der SM, and Upadhyaya M (2003) D4F104S1 deletion in facioscapulohumeral muscular dystrophy: Phenotype, size, and detection. *Neurology* 61(2): 178-183.
33. Litt MD, Simpson M, Gaszner M, Allis CD, and Felsenfeld G (2001) Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 293(5539): 2453-2455.
34. Lundgren M, Chow CM, Sabbattini P, Georgiou A, Minaee S, and Dillon N (2000) Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* 103(5): 733-743.
35. Lunt PW (2000) Facioscapulohumeral muscular dystrophy: diagnostic and molecular aspects: 44-60. In: *Neuromuscular diseases: from basic mechanisms to clinical management*. Vol. 18. Deymeer F (Ed.) Karger, Basel.
36. Lyle R, Wright TJ, Clark LN, and Hewitt JE (1995) The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* 28(3): 389-397.
37. Magewu AN and Jones PA (1994) Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. *Mol Cell Biol* 14(6): 4225-4232.
38. Meneveri R, Agresti A, Marozzi A, Saccone S, Rocchi M, Archidiacono N, Corneo G, Della VG, and Ginelli E (1993) Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. *Gene* 123(2): 227-234.
39. Mihaly J, Hogga I, Barges S, Galloni M, Mishra RK, Hagstrom K, Muller M, Schedl P, Sipos L, Gausz J, *et al.* (1998) Chromatin domain boundaries in the Bithorax complex. *Cell Mol Life Sci* 54(1): 60-70.
40. Myers FA, Evans DR, Clayton AL, Thorne AW, and Crane-Robinson C (2001) Targeted and extended acetylation of histones H4 and H3 at active and inactive genes in chicken embryo erythrocytes. *J Biol Chem* 276(23): 20197-20205.
41. Noma K, Allis CD, and Grewal SI (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* 293(5532): 1150-1155.
42. O'Neill LP and Turner BM (1995) Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J* 14(16): 3946-3957.
43. Pirrotta V and Rastelli L (1994) White gene expression, repressive chromatin domains and homeotic gene regulation in *Drosophila*. *Bioessays* 16(8): 549-556.
44. Pryde FE and Louis EJ (1999) Limitations of silencing at native yeast telomeres. *EMBO J* 18(9): 2538-2550.
45. Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, and Reuter G (2002) Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* 21(5): 1121-1131.
46. Song M, Goto K, Lee JH, Matsumura T, Sahashi K, and Arahata K (2000) Facioscapulohumeral muscular dystrophy (FSHD). *Neuroscience News* 3: 28-33.

47. Strahl-Bolsinger S, Hecht A, Luo K, and Grunstein M (1997) SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev* 11(1): 83-93.
48. Sturzenbaum SR and Kille P (2001) Control genes in quantitative molecular biological techniques: the variability of invariance. *Comp Biochem Physiol B Biochem Mol Biol* 130(3): 281-289.
49. Sun FL, Cuaycong MH, and Elgin SC (2001) Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin. *Mol Cell Biol* 21(8): 2867-2879.
50. Talbert PB and Henikoff S (2000) A reexamination of spreading of position-effect variegation in the white-rough region of *Drosophila melanogaster*. *Genetics* 154(1): 259-272.
51. Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, and Figlewicz D (1996) Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39(6): 744-748.
52. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, and Laat de W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10(6): 1453-1465.
53. 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.
54. Upadhyaya M and Cooper DN (2002) Molecular diagnosis of facioscapulohumeral muscular dystrophy. *Expert Rev Mol Diagn* 2(2): 160-171.
55. Vandesompele J, Preter de K, Pattyn F, Poppe B, Roy van N, Paepe de A, and Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3(7): research 0034. 1-11.
56. Wakimoto BT (1998) Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*. *Cell* 93(3): 321-324.
57. Wallrath LL and Elgin SC (1995) Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev* 9(10): 1263-1277.
58. Weiler KS and Wakimoto BT (1995) Heterochromatin and gene expression in *Drosophila*. *Annu Rev Genet* 29: 577-605.
59. West AG, Gaszner M, and Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev* 16(3): 271-288.
60. 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.
61. Wijmenga C, Winokur ST, Padberg GW, Skraastad MI, Altherr MR, Wasmuth JJ, Murray JC, Hofker MH, and Frants RR (1993) The human skeletal muscle adenine nucleotide translocator gene maps to chromosome 4q35 in the region of the facioscapulohumeral muscular dystrophy locus. *Hum Genet* 92(2): 198-203.
62. Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, *et al.* (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosome Res* 2(3): 225-234.
63. Wright TJ, Wijmenga C, Clark LN, Frants RR, Williamson R, and Hewitt JE (1993) Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. *Hum Mol Genet* 2(10): 1673-1678.
64. Wright WE, Tesmer VM, Liao ML, and Shay JW (1999) Normal human telomeres are not late replicating. *Exp Cell Res* 251(2): 492-499.

65. Yan CM, Dobie KW, Le HD, Konev AY, and Karpen GH (2002) Efficient recovery of centric heterochromatin P-element insertions in *Drosophila melanogaster*. *Genetics* 161(1): 217-229.
66. Yip DJ and Picketts DJ (2003) Increasing D4Z4 repeat copy number compromises C2C12 myoblast differentiation. *FEBS Lett* 537(1-3): 133-138.
67. Yoshioka H, Shirayoshi Y, and Oshimura M (2001) A novel in vitro system for analyzing parental allele-specific histone acetylation in genomic imprinting. *J Hum Genet* 46(11): 626-632.
68. 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.

8

Genetic and epigenetic studies of the FSHD-associated D4Z4 repeat

Discussion and future perspectives



Table of contents

8.1 Subtelomeric exchanges between 4q35 and 10q26	213
8.1.1 Exchange frequencies	213
8.1.1.1 Comparisons between different control and patient populations	213
8.1.1.2 Discussion and perspectives for future research	215
8.1.2 Size distributions of both repeat arrays	217
8.1.2.1 Analysis of 4q and 10q size distribution	217
8.1.2.2 Discussion and perspectives for future research	217
8.2 Mosaicism for the FSHD-associated region	220
8.2.1 Detection of mosaicism and the relevance for FSHD development	220
8.2.2 Discussion and perspectives for future research	221
8.3 Epigenetic modifications of the 4q35 region	222
8.3.1 D4Z4 and the link with the ICF syndrome	223
8.3.2 DNA methylation patterns of D4Z4	223
8.3.2.1 FSHD-associated hypomethylation of D4Z4	223
8.3.2.2 Discussion and perspectives for future research	224
8.3.3 Histone modifications	229
8.3.3.1 H4 acetylation of D4Z4	229
8.3.3.2 Discussion and perspectives for future research	230
8.4 References	231

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 (<i>n</i>)	% Exchange	% 4 on 10	% 10 on 4
Van Deutekom <i>et al.</i> 1996	Dutch	50	20 ^a	10	10
Van Overveld <i>et al.</i> 2000	Dutch	208	21 ^a	12	9
Matsumura <i>et al.</i> 2002	Japanese	153	28	20	8
	Korean	124	30	26	4
	Chinese	114	19	14	5
Su <i>et al.</i> 2003	Chinese	52/107 ^b	19	10	9
Wang <i>et al.</i> 2003	Chinese	50	16	8	8
Wu <i>et al.</i> 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 (<i>n</i>)	% Exchange	% 4 on 10	% 10 on 4
van der Maarel <i>et al.</i> 2000	Dutch	8 ^b	38	3	0
Matsumura <i>et al.</i> 2002	Japanese	56	32	23	9
Su <i>et al.</i> 2003	Chinese	70	19	13	6
Wang <i>et al.</i> 2003	Chinese	45	20	15	5
Wu <i>et al.</i> 2004	Chinese	55	17	15	2
Goto <i>et al.</i> 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)	<i>de novo</i> (n)	Familial (n)	<i>de novo</i> (n)		Familial (n)	
					4 on 10	10 on 4	4 on 10	10 on 4
van der Maarel <i>et al.</i> 2000	Dutch	8 ^a	8	0	3	0	0	0
Wang <i>et al.</i> 2003	Chinese	45	16	29	7	0	0	2
Wu <i>et al.</i> 2004	Chinese	55	16	39	6	0	2	1
Goto <i>et al.</i> 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 *BglII/BlnI* 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, Filipinski *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 than 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 contracted 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-derived 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 *BlnI*. 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 *BlnI* 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 *FseI*, located in the putative promoter region of *DUX4*, and to a lesser extent for *BsaAI* 350 bp proximal to *FseI*. 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. Filipinski 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.

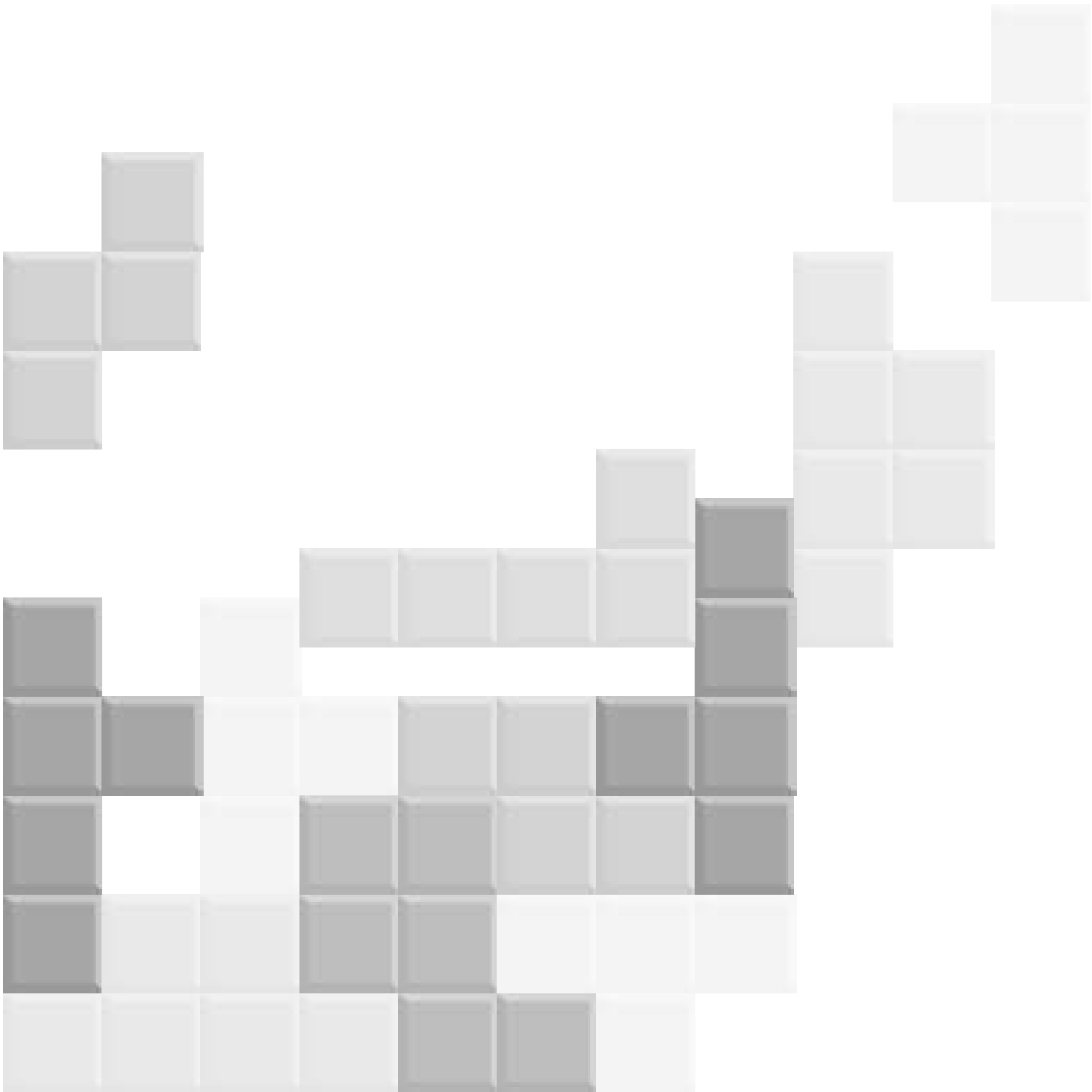
14. Gottlieb B, Beitel LK, and Trifiro MA (2001) Somatic mosaicism and variable expressivity. *Trends Genet* 17(2): 79-82.
15. Hall JG (1988) Review and hypotheses: somatic mosaicism: observations related to clinical genetics. *Am J Hum Genet* 43(4): 355-363.
16. Herman JG, Graff JR, Myohanen S, Nelkin 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 *KpnI* 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 *KpnI* 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 (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.

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 *EcoRI* 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 daardoor 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.

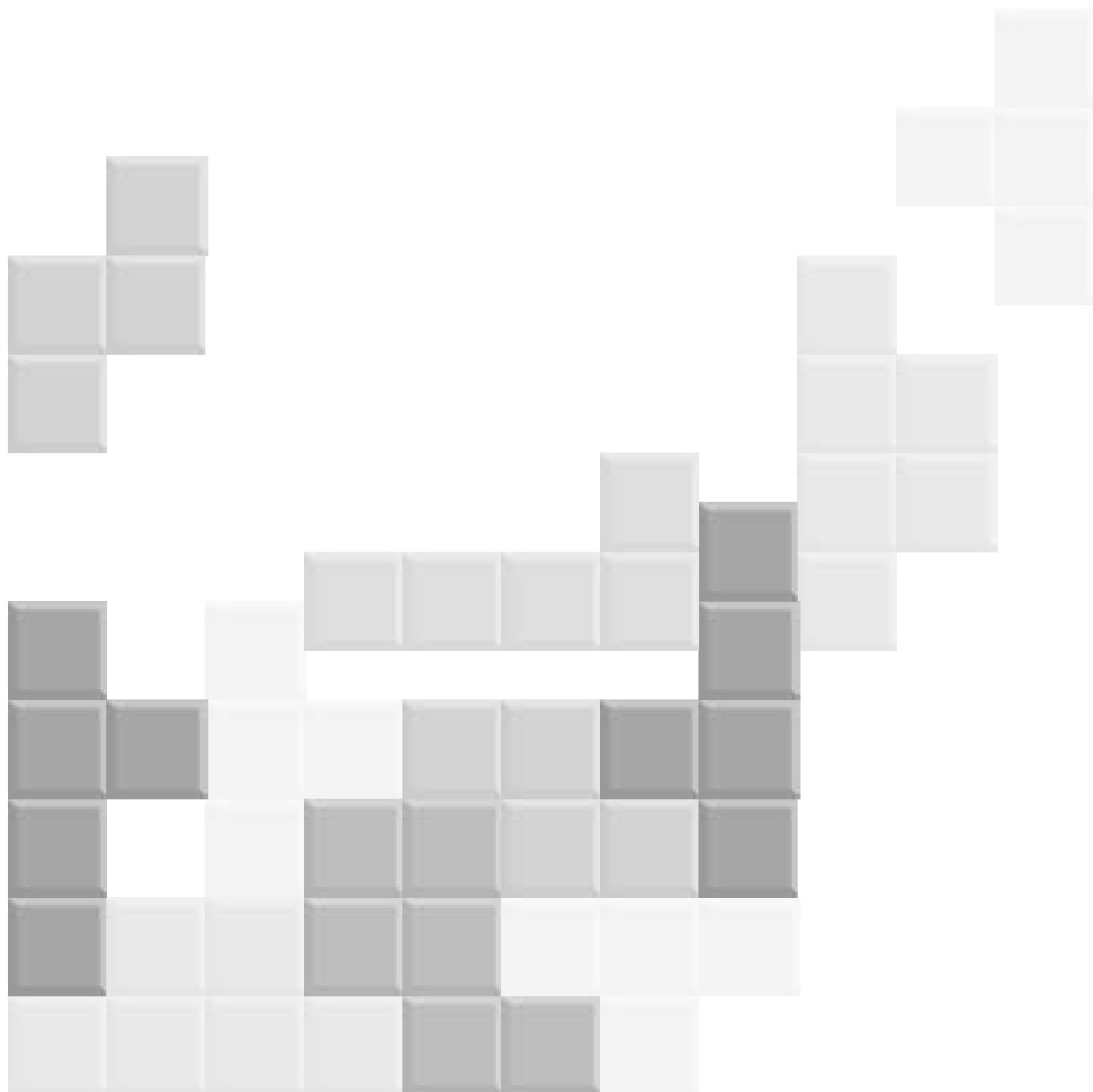
De bestudering van de acetylatie van histon nummer 4 in de sequentie van 4q35 wordt beschreven in *Hoofdstuk 7*. Deze analyse geeft aan, dat de chromatinestructuur van het gebied naast de D4Z4 repeat meer op een open, maar inactieve chromatinestructuur lijkt dan op heterochromatine. Dit suggereert, dat 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 opert, 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-mosaïeke 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.



List of abbreviations

Ab	antibody
Ac	acetylation
ADP	adenosine diphosphate
ALP	actinin-associated LIM protein gene
ANT	adenine nucleotide translocator gene 1
ATP	adenosine triphosphate
BMD	Becker muscular dystrophy
bp	basepair
cDNA	complementary DNA
cen	centromere
ChIP	chromatin immunoprecipitation
Chr	chromosome
cM	CentiMorgan
CSS	clinical severity score
del	deletion
DM	myotonic dystrophy
DMD	Duchenne muscular dystrophy
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DSB	double strand break
DUX	double homeobox protein gene
EST	expressed sequence tag
FISH	fluorescence <i>in situ</i> hybridization
FRG1	FSHD region gene 1
FRG2	FSHD region gene 2
FSHD	facioscapulohumeral muscular dystrophy
GAPD	glyceraldehyde phosphate dehydrogenase
GC+CO	gene conversion with crossover
GC-CO	gene conversion without crossover
H1	linker histone 1
H2A	histone 2A
H2B	histone 2B
H3	histone 3
H4	histone 4
HAT	histone acetyltransferase
HDAC	histone deacetylase
ICF	immunodeficiency-centromeric instability-facial anomalies syndrome
IP	immunoprecipitation
K	lysine
kb	kilobase
LCL	lymphoblastoid cell line
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LTR	long terminal repeat transposon
Mb	megabase
mRNA	messenger RNA
OPMD	oculopharyngeal muscular dystrophy
OR	olfactory receptor
ORF	open reading frame
p	as in 4p: the short arm of a chromosome
PBL	peripheral blood lymphocyte

PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PEV	position effect variegation
PFGE	pulsed-field gel electrophoresis
q	as in 4q: the long arm of a chromosome
Q-PCR	quantitative real-time PCR
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rDNA	ribosomal DNA
rRNA	ribosomal RNA
RT	reverse transcription
RTC	relative telomeric repeat content
Sat	satellite
S.E.M.	standard error of the mean
SCH	somatic cell hybrid
SD	standard deviation
SDSA	synthesis-dependent strand annealing
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
SUMO	small ubiquitin-related modifier
tel	telomere
ter	as in 4qter: the terminal region of a chromosome
TUBB4Q	tubulin beta polypeptide 4Q gene
USCE	unequal sister-chromatid exchange
VNTR	variable number of tandem repeats

List of publications

SM van der Maarel, G Deidda, RJLF Lemmers, **PGM van Overveld**, MJ van der Wielen, JE Hewitt, LA Sandkuijl, B Bakker, GJB van Ommen, GW Padberg and RR Frants. *De novo* facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *American Journal of Human Genetics* (2000) 66: 326-366.

PGM van Overveld, RJFL Lemmers, G Deidda, LA Sandkuijl, GW Padberg, RR Frants and SM van der Maarel. Frequent interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *American Journal of Human Genetics* (2000) 67 (Supplement): Abstract #171.

PGM van Overveld, RJFL Lemmers, G Deidda, LA Sandkuijl, GW Padberg, RR Frants and SM van der Maarel. Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Human Molecular Genetics* (2000) 9 (19): 2879-2884.

PGM van Overveld, LA Sandkuijl, L Enthoven, GW Padberg, RR Frants and SM van der Maarel. Hypomethylation of the D4Z4 repeat array in facioscapulohumeral muscular dystrophy alleles (a new assay to study methylation status of D4Z4 repeats in FSHD alleles). *American Journal of Human Genetics* (2002) 71 (Supplement): Abstract #1790.

M Wohlgemuth, RJFL Lemmers, EL van der Kooi, MJ van der Wielen, **PGM van Overveld**, H Dauwerse, B Bakker, RR Frants, GW Padberg, and SM van der Maarel. Possible phenotypic dosage effect in patients compound heterozygous for FSHD-sized 4q35 alleles. *Neurology* (2003) 61: 909-914.

SM van der Maarel, **PGM van Overveld**, RJFL Lemmers, LA Sandkuijl, L Enthoven, ST Winokur, GJB van Ommen, GW Padberg and RR Frants. D4Z4 hypomethylation in FSHD causes the transcriptional upregulation of 4qter genes. *American Journal of Human Genetics* (2003) 73 (Supplement): Abstract #154.

M Ehrlich, G Jiang, F Yang, **PGM van Overveld**, V Vedanarayanan and SM van der Maarel. Histone acetylation assays for heterochromatin disprove the position effect variegation hypothesis for facioscapulohumeral muscular dystrophy caused by shortening of D4Z4 repeat arrays. *American Journal of Human Genetics* (2003) 73 (Supplement): Abstract #2438.

PGM van Overveld, RJFL Lemmers, LA Sandkuijl, L Enthoven, ST Winokur, F Bakels, GW Padberg, GJB van Ommen, RR Frants and SM van der Maarel. D4Z4 hypomethylation in 4q-linked FSHD and non 4q-linked FSHD. *Nature Genetics* (2003) 35 (4): 315-317.

G Jiang, F Yang, **PGM van Overveld**, V Vedanarayanan, SM van der Maarel, and M Ehrlich. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Human Molecular Genetics* (2003) 12 (22): 2909-2921.

RJFL Lemmers, **PGM van Overveld**, LA Sandkuijl, H Vrieling, GW Padberg, RR Frants and SM van der Maarel. Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. *American Journal of Human Genetics* (2004) 75: 44-53.

RJFL Lemmers, **PGM van Overveld**, LA Sandkuijl, H Vrieling, GW Padberg, RR Frants and SM van der Maarel. Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in FSHD. *American Journal of Human Genetics* (2004) 75 (Supplement): Abstract #2490.

PGM van Overveld, RR Frants and SM van der Maarel. Mosaicism and FSHD. Chapter 12: 169-183. In: Facioscapulohumeral muscular dystrophy: clinical medicine and molecular cell biology. ME Upadyaya and DN Cooper (Eds.) 2004 Garland Science/BIOS Scientific Publishers Limited.

PGM van Overveld, L Enthoven, E Ricci, L Felicetti, M Jeanpierre, ST Winokur, RR Frants, GW Padberg, and SM van der Maarel. Residual D4Z4 repeat size and D4Z4 methylation levels separate FSHD into two clinical severity classes. *submitted*.

Curriculum vitae

Petra Grada Maria van Overveld werd geboren op 6 januari 1976 in de gemeente Roosendaal en Nispen. Ze behaalde haar gymnasium diploma in 1995 aan het St. Norbertus College te Roosendaal. In september 1995 begon zij haar studie biologie aan de Universiteit Leiden, waar ze in 1999 *cum laude* afstudeerde. Aansluitend was ze werkzaam als promovenda bij het Centrum voor Humane en Klinische Genetica, onderdeel van het Leids Universitair Medisch Centrum, onder begeleiding van Prof. Dr. Frants en Dr. Ir. van der Maarel. Het daar uitgevoerde onderzoek staat beschreven in dit proefschrift. Sinds 1 november 2004 werkt ze als postdoc bij de afdeling Urologie van het Leids Universitair Medisch Centrum, aan een project getiteld “Role of epithelial-mesenchymal transition and osteomimicry in bone metastasis from prostate carcinoma: new therapeutic opportunities”. Op 14 januari 2005 ontving ze de “Prinses Beatrix Fonds jaarprijs neuromusculaire ziekten” voor de beste publicatie van 2004 voor het artikel “Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy“, beschreven in hoofdstuk 5 van dit proefschrift.

Petra Grada Maria van Overveld was born on January 6th, 1976 in Roosendaal en Nispen. In September 1995 she graduated from the gymnasium St. Norbertus College in Roosendaal. Later that year she started her biology degree at Leiden University, which she completed *cum laude* in 1999. Subsequently she performed her PhD research at the Center of Human and Clinical Genetics from the Leiden University Medical Center, under supervision of Prof. Dr. Frants en Dr. Ir. van der Maarel. The work from this period is presented in this thesis. As of November 1st, 2004 she works as a postdoctoral researcher on a project entitled “Role of epithelial-mesenchymal transition and osteomimicry in bone metastasis from prostate carcinoma: new therapeutic opportunities” in the Department of Urology at the Leiden University Medical Center. On January 14th, 2005 she received the “Prinses Beatrix Fonds jaarprijs neuromusculaire ziekten” for the best publication of 2004 for her article “Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy“, presented in chapter 5 of this thesis.

“Het leven is een oneindige weg die we bereizen en nergens naartoe leidt.
Het enige dat telt is hoe je de weg bereist, niet waar je uitkomt”

Isabel Allende

Na mijn studie begon ik aan een lange nieuwe reis met bestemming Academiegebouw! Er kwamen natuurlijk hindernissen op mijn pad en ik was (soms) eigenwijs, maar met enthousiasme en plezier ging ik verder. Totdat het moment kwam, dat ik een deel van mijn inspiratie verloor en ik onverwacht afscheid moest nemen van zowel oma van Overveld als Lodewijk. Ik had even het gevoel dat ik verdwaalde, maar mijn reddingsvest gevuld met geborgenheid, vertrouwen, humor en geluk hielp me –zoals altijd- weer op weg. Ik heb nu mijn eindpunt bereikt en ik weet, dat ik het zonder alle liefde en vriendschap niet had gered. Jan, Trudi, Sandra, Frank, Sijsen, oma Bijman, Pietje én Adriaan, Arja, Caroline, Cynthia, Guido, Inge, Jeanet, JoukeJan, Klazina, Marjolein, Maytal, Melissa, Mirka, Peter K., Peter S., Peter-Bram, Sergiu, Silvana, Wouter, Yair en Yvonne: jullie zijn mijn reddingsvest, bedankt voor alles! Voor jou, Tinka, duizend maal dank! Zonder jou was mijn reis veel minder comfortabel geweest, want jij haalt me letterlijk altijd weer uit de knoop! Verder gaat mijn dank uit naar alle mensen bij het Centrum voor Humane en Klinische Genetica voor de gezellige reistijd en het altijd bieden van een helpende hand. Beste Frantsjes, dankzij jullie ga ik bij het horen van de woorden koffiepauze, cake, sinterklaas en kalkoen automatisch glimlachen! Iedereen bij Endocrinologie en Stofwisselingsziekten wil ik bedanken voor het warme onthaal en de steun bij het laatste stukje van mijn promotietraject. En voor jou, lieve Leo, zijn er eigenlijk geen woorden. Jij omringt me met warmte, vult mijn hart met liefde en geeft me altijd een veilige haven om naar terug te keren. Bedankt voor je optimisme, je eeuwige geduld en alle motivatie. We hebben het gehaald!

Veel liefs,

A handwritten signature in black ink that reads "Petra." with a small flourish underneath.