UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Doctoral school
Molecular and Regenerative Medicine
XXVI cycle

Investigation of genetic complexity of Facioscapulohumeral Muscular Dystrophy

PhD student:
Ana Nikolic

PhD mentor:
MD PhD Rossella Tupler

Director of the Doctoral school:
MD PhD Rossella Tupler

Contents

1.Introduction 4

1.1 The Facioscapulohumeral Dystrophy (FSHD) 3

1.2 Definition of diagnostic criteria for FSHD 3

1.2.1 Clinical diagnostic criteria 4

1.2.2 Genetic diagnostic criteria 5

1.2.3 Laboratory diagnostic criteria 5

1.3 A standardized clinical evaluation of patients affected by Facioscapulohumeral muscular dystrophy: the FSHD clinical score 6

1.3.1 Italian National Registry for FSHD (INRF) 6

1.3.2 The FSHD score 6

1.4 Identification of the molecular defect responsible for development of FSHD 8

1.4.1 Linkage analysis 8

1.4.2 Molecular test for FSHD 9

1.4.3 The identification of specific molecular signature uniquely associated with FSHD expression 11

1.4.4 Correlation studies between the size of D4Z4 allele and clinical expression in FSHD 13

1.4.5 Facioscapulohumeral muscular dystrophy in subjects carrying full-length D4Z4 alleles (FSHD2) 14

1.5 4q35 region and models for FSHD pathogenesis 16

1.5.1 D4Z4 units 16

1.5.2 Role of D4Z4 transcription in FSHD pathogenesis 17

1.5.3 Expression of proximal genes controlled by D4Z4 19

1.5.4 Epigenetic hypothesis: chromatin modifications in FSHD 21

1.6 Recent findings from different genotype-phenotype studies and clinical reports on FSHD patients and families revealed several exceptions to the general outline 25

1.6.1 Cross-sectional genotype-phenotype study of FSHD families with D4Z4 reduced allele [Ricci et al, 2013] 26

2. Aims of the thesis 32

2.1 Detailed molecular and clinical characterization of probands and their relatives carrying the shortest D4Z4 reduced allele (1-3 DRA) 32

2.2 Analysis of D4Z4 methylation status among FSHD patients and healthy controls ... 33

3. Materials and methods 34

3.1 Subject recruitment 34

3.2 Study design 35

3.2.1 Detailed molecular and clinical characterization of probands and their relatives carrying the shortest D4Z4 reduced allele (1-3 DRA).. 35

3.2.2 Analysis of D4Z4 methylation status among FSHD patients and healthy controls . 36

3.3 Molecular analysis 37

3.3.1 Standard molecular test for FSHD 37

3.3.2 BsaAI and FseI methylation assay 42

3.3.3 In-depth methylation assay: Bisulfite D4Z4 conversion and sequencing 43

3.3.4 Statistical analysis 45

4. Results 46

4.1 Detailed molecular and clinical characterization of probands carrying the shortest D4Z4 reduced allele (1-3 DRA) and their relatives 46

4.2 Analysis of D4Z4 methylation status among FSHD patients and healthy controls ... 51

4.2.1 The correlation of the level of methylation and the total number of D4Z4 repeats on 4q in FSHD patients and healthy controls 52

4.2.3 The inter-familial phenotypic variability does not correlates with the methylation status in a FSHD family members 54

4.2.4 The methylation status in the autosomal recessive family 56

4.2.5 In-depth methylation analysis 58

5. Discussion 63

5.1 Are the 4q alleles with 1-3 D4Z4 repeats always predictive for a severe form of FSHD? 64

5.2 How to explain the phenotypic variability in FSHD? 65

6. Conclusions 69

Appendix 1 71

Appendix 2 82

Appendix 3 85

References 88

1. Introduction

1.1 The Facioscapulohumeral Dystrophy (FSHD)

Facioscapulohumeral muscular dystrophy (FSHD, OMIM #158900) is the third most common form of hereditary myopathy with a prevalence of 1 in 20.000 [Mostacciuolo et al, 2009]. Duchenne de Boulogne firstly reported the FSHD syndrome: in 1862 he published one photo of an affected patient in his *Album de photographies pathologiques* [Duchenne, 1862] and in 1869 the photos of the same patient's family in the *Revue photographique des hopitaux de Paris* [Duchenne, 1869]. He was also first to describe the disease in his famous series of papers in *Archives of General Medicine* in 1868, often cited as the earliest reference of the disease [Engel and Franzini-Armstrong, 2004]. Since 1885 the disease was called "Landouzy-Dejerine form of muscular dystrophy", as Landouzy and Dejerine [Landouzy-Dejerine, 1885] presented detailed description of disease. Subsequent reports of large pedigrees suggested that disease is transmitted in an autosomal dominant trait [Pearson, 1933]. In 1950 Tyler and Stephens described a large family affected by FSHD (1248 relatives from six generations), identifying 159 affected individuals and reporting the neurological examination of 58 patients. In this report they highlighted FSHD as a progressive disorder characterized by a precocious and pronounced involvement of the brachioradialis muscles [Tyler and Stephens, 1950]. In 1954 Walton and Natrass identified the clinical criteria for the diagnosis, showing the slowly progressive nature of the disease and its limited and asymmetric muscle weakness distribution.

However, the thesis of Padberg in 1982 thoroughly reviewed the literature and identified the clinical features of the disease through the examination of a large cohort of patients. Padberg investigated a group of 107 patients, whose 73 were symptomatic. The facial weakness was present in 10% of cases, shoulder weakness in 82% and foot extensor weakness in 8%. None of his patients had pelvic girdle or calf muscle weakness [in Padberg G. Facioscapulohumeral disease [thesis]. University of Leiden, The Netherlands, 1982].

1.2 Definition of diagnostic criteria for FSHD

The International Consortium identified the clinical, genetic and laboratory criteria for FSHD diagnosis in 1991, in absence of an available diagnostic DNA test. Importantly, this work also responded to the need to select individuals that could be included in the linkage analysis [Padberg et al, 1991]. In fact, uniform and well-established clinical criteria were

considered necessary to support research activities directed towards identification of FSHD gene.

1.2.1 Clinical diagnostic criteria

Since FSHD was defined largely on clinical grounds, the following definitions were introduced: (i) "non-penetrance" refers to an obligate gene carrier without symptoms (complaints or subjective findings) or signs (objective phenomena) relating to the disease (ii) "pre-symptomatic" indicates that a person has no complaints (symptoms) related to the disease, but has muscle atrophy and weakness demonstrable by physical examination; and (iii) "symptomatic" refers to patients with complaints and objective findings related to the weakness and muscle atrophy of FSHD. It was established that presenting symptoms had to relate to weakness and wasting in facial or shoulder girdle muscles. Onset in pelvic girdle muscles was considered suggestive of alternative diagnoses, although a subsequent pelvic girdle involvement was common during the progression of FSHD. It was reported that the clinically recognizable age of onset is often very variable. The mean age of recognizable onset, at least by clinical examination, is in the second decade. Onset before the age of 5 years has been rarely reported and in these early-onset cases [Brouwer et al, 1994] the involvement of facial muscles are considered necessary for diagnosis of FSHD. Facial weakness is described affecting eye closure (orbicularis oculi) and peri-oral muscles (orbicularis oris) and it is reported in the vast majority of patients (Figure 1a).

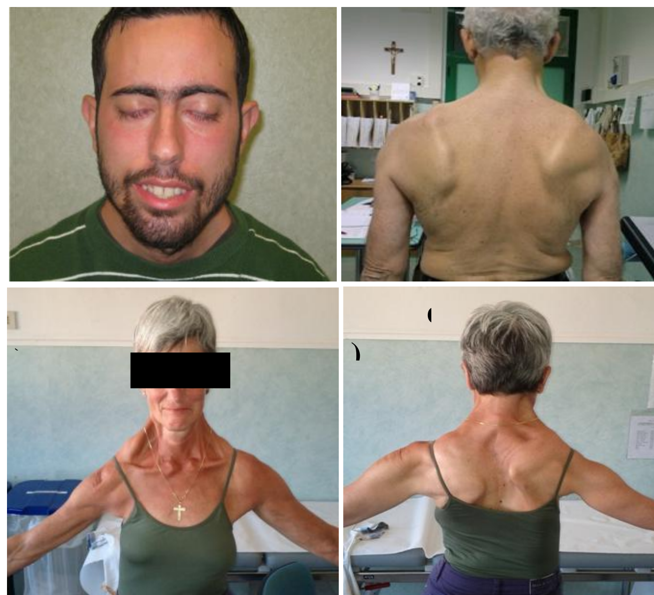

Figure 1: Phenotypic features of FSHD patients (from the left corner to the right corner) : Facial weakness affecting eye closure; weakness of the muscles of the scapula fixators, which occurs during the abduction of the arms with the raising of the trapezius muscle.

major muscles result early affected in most cases. The deltoid muscles remain unaffected for a long period of time and often have a particular pattern of atrophy, i.e. partial and proximal.

A typical feature of FSHD is considered the asymmetric muscle involvement. Symmetric weakness and atrophy at presentation is unusual and can necessitate caution before accepting the diagnosis as FSHD.

Progression of the disease is highly variable, involving abdominal and foot extensor muscles at an early stage. Also, the pelvic girdle weakness and upper arm weakness may occur at any time after the onset of shoulder girdle weakness. Neck extensor, intrinsic hand and *triceps surae* muscle weakness is described as uncommon features of disease, but occasionally observed within families. However, pre-symptomatic cases are described at any age and appear to comprise approximately 30% of all cases in large families. In the symptomatic cases, the disease is progressive tough the rate of progression is variable in the majority of cases. Rarely, there can be long periods of apparent arrest of progression.

1.2.2 Genetic diagnostic criteria

The pattern of inheritance in familial cases is considered autosomal dominant. Sporadic cases are also reported, although their frequency is unknown. Evidences for recessive inheritance were not reported.

1.2.3 Laboratory diagnostic criteria

Blood creatine kinase (CK) levels can be normal, but are often elevated, athough rarely exceed five times the upper limit of normal. Electromyography (EMG) shows a myopathic pattern; some neurogenic features, including positive sharp waves, are present occasionally. Motor and sensory nerve conduction velocities are normal. Muscle biopsies may exhibit nonspecific myopathic changes. Cellular infiltrates are frequently observed in FSHD and can be extensive.

In summary, the four main criteria were identified to define FSHD. These are: (1) onset of the disease in facial or shoulder girdle muscles; sparing of the extra-ocular, pharyngeal and lingual muscles and the myocardium; (2) facial weakness in more than 50% of the affected family members; (3) autosomal dominant inheritance in familial cases; and (4) evidence of myopathic disease in EMG and muscle biopsy in at least one affected member without biopsy features specific to alternative diagnoses. Clinical and laboratory features suggestive of alternative diagnosis are following:

- involvement of extra-ocular, masticatory, pharyngeal and lingual muscle;
- regression of symptoms and signs;
- presence of severe and diffuse contractures;
- involvement of myocardium with presence of cardiomyopathy;
- persistent high CK values above five times the upper limit

1.3 A standardized clinical evaluation of patients affected by Facioscapulohumeral muscular dystrophy: the FSHD clinical score

1.3.1 Italian National Registry for FSHD (INRF)

The Italian National Registry for FSHD (INRF) was established in 2007 by the Italian Clinical Network for FSHD (ICNF). The ICNF is composed by two diagnostic laboratories and thirteen Clinical Centers with expertise in diagnosis and management of neuromuscular disorders. The institution of the INRF was based on the collection of DNA samples accrued by the two laboratories, in Modena and Rome, which conduct molecular diagnosis of FSHD and are the referral centers for the whole Italian territory. The Clinical Centers participating to the INRF are responsible for the clinical evaluation of FSHD probands and all available and willing to participate family members. Dedicated specific software and a website for data management has been designed. The dedicated website for data management, description of the project and participating groups are available on-line at www.fshd.it.

1.3.2 The FSHD score

To define numerically the clinical severity of the disease, the Italian Clinical Network for FSHD has developed a standardized protocol that allows the quantification of muscle weakness in FSHD patients through the functional evaluation of six muscle groups specifically affected in FSHD. To this purpose, we generated a new protocol, easily used in the medical office that examines muscle groups specifically affected in FSHD by using functional criteria, allowing translation of clinical severity in quantitative terms [Lamperti et al, 2010].

The clinical form consists of three parts, named a, b, and c, that examine three aspects of the disease and have been designed to facilitate accurate study of molecularly defined FSHD subjects.Part a investigates the patient's clinical history, focusing on medical conditions and particular habits. Part b evaluates the patient's disability. Part c assesses

muscle segmental involvement by using the Medical Research Council (MRC) scale. The evaluation procedure allows to assess the strength and the function of muscular groups belonging to I) face (score from 0 to 2); II) shoulder girdle (score from 0 to 3); III) upper limbs (score from 0 to 2); IV) distal legs (0-2); V) pelvic girdle (score from 0 to 5) VI) abdominal muscles (0-1). More detailed information such as asymmetry of presentation or any observed peculiarity can be added in the section “others”. The functional examination of six different groups of muscles (Figure 2), as described in part b of the clinical form integrated with the results of part c, generates the FSHD clinical score. The total score can range from 0, when no signs of muscle weakness are present, to 15, when all muscle groups tested are severely impaired. The FSHD clinical form and the FSHD evaluation scale form, as well as a visual guide, are available online at www.fshd.it.

To date, the Italian Network for FSHD successfully use the FSHD clinical score as a tool in genotype/phenotype correlation and genetic counseling. This method represents an unique opportunity to study FSHD in standardized manner, collecting anamnestic records and information of impairment of selective muscle groups.

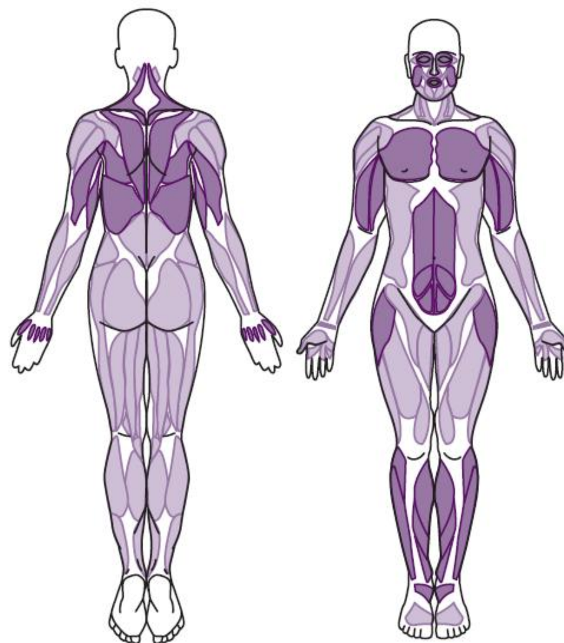

Figure 2: Schematic representation of major involved muscles in FSHD. In dark color are shown facial muscles, scapular girdle muscles, abdominal muscles, pelvic girdle muscles that are most affected in FSHD.

1.4 Identification of the molecular defect responsible for development of FSHD

1.4.1 Linkage analysis

The variability in expression, severity, age at onset and need for an accurate pre-symptomatic test gave impetus to an active search for the localization, identification, and characterization of the locus responsible for FSHD development [Lunt et al, 1989]. An International Consortium for FSHD linkage was organized for this purpose. At the first meeting of the Consortium, in 1988, an initial exclusion map for FSHD was constructed [Sarfarazi et al, 1989; Jacobsen et al, 1990]. In 1990 the FSHD gene was assigned to chromosome 4 by positional mapping in 10 Dutch families [Wijmenga et al, 1990]; confirmation of this location was performed in other families and with additional probes [Upadhyaya et al, 1990, 1991]. Wijmenga and coworkers in 1991 reported that D4S139, a Variable Number Tandem Repeat structure (VNTR) locus, was much more closely linked to FSHD. Two-point linkage analysis between FSHD and D4S139 in nine informative families showed a maximum combined lod score. D4S139 was mapped to chromosome 4q35-qter by in situ hybridization, thus firmly establishing the location of the FSHD gene in the subtelomeric region of chromosome 4q. In 1992, the members of the International Consortium for linkage analysis of the FSHD gene had pooled data for joint analyses, in an attempt to determine the precise location of the FSHD gene and the order of four DNA markers on 4q35 region. Six laboratories determined a total of 3078 genotypes in 65 families, consisting of a total of 504 affected subjects and 559 unaffected subjects. For each marker, a mean of 648 meioses were informative. D4S139 and D4S163 were identified as the closest linked markers to the FSHD locus. As result the assignment of the FSHD locus on region 4q35 in a total of 65 families was definitively established. The gene responsible for FSHD was refined to a position distal to the D4S139 locus. The cosmid clone 13E, isolated in search for homebox genes, was subsequently mapped to 4q35, also distal to D4S139 [Wijmenga et al, 1992]. The subclone p13E-11 resulted to detect in healthy subjects a polymorphic EcoRI fragment, usually larger than 28 kilobases (kb). The analysis performed by Wijmenga and coworkers [1992] showed that in healthy individuals the majority (72%) of EcoRI fragments detected by p13E-11 were larger than 28 kb, while in FSHD patients there was an overrepresentation of fragments smaller than 28 kb.

In conclusion, the hybridization of restriction enzyme EcoRI digested DNA using the p13E-11 probe identified a 3.3 kb tandemly arrayed sequence (D4Z4) located at the 4q

subtelomeric region, that resulted rearranged in almost all FSHD patients. Based on restriction fragment mapping and DNA sequencing, van Deutekom and coworkers [1993] confirmed that the rearrangements associated with FSHD result in deletion of integral number of repeat KpnI fragments, designated D4Z4 [van Deutekom et al, 1993].

Approximately half of new FSHD cases arise as a consequence of a postzygotic rearrangement of the repeat leading to somatic mosaicism for the D4Z4 repeat contraction [Lemmers et al, 2004a].

1.4.2 Molecular test for FSHD

The p13E-11 probe detects at least two pairs of EcoRI alleles, one derived from the 4q and the other from 10q. Indeed a D4Z4 repeat array 98% identical to the 4q array is present at the 10q subtelomere [van Deutekom et al, 1996]. 4q35 and 10q26 EcoRI clones can be distinguished by restriction analysis with SfiI and Styl. The accurate comparison of nucleotide sequences between 4q35 and non-4q35 EcoRI fragments led to the identification of restriction enzymes able to cut specifically in either one of the alleles; thus facilitating the interpretation of the p13E-11 hybridization patterns. The detection of sequence divergence between the KpnI tandem repeat units located at 4q and 10q showed a different distribution of restriction enzyme sites and that the restriction enzyme BlnI specifically cleaved the variable KpnI region of the 10qter p13E-11 fragments (Figure 3), leaving intact the tandem repeat units at 4q and thus allowing the direct identification of 4q35 alleles implicated in the disease [Deidda et al, 1996; Upadhyaya et al, 1997].

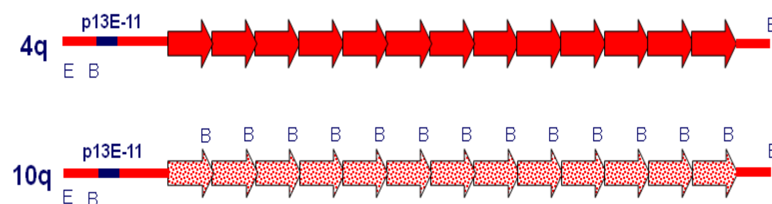

Figure 3: The p13E-11 probe detects at least two pairs of EcoRI alleles, one derived from the 4q and the other from 10q. The restriction site of enzyme BlnI specifically cleave variable KpnI region of 10qter p13E11 fragments, leaving intact the tandem repeat units at 4q and thus allowing the identification of 4q35.

Therefore, a double restriction enzyme digest with the enzymes EcoRI and BlnI was routinely used, greatly facilitating the molecular diagnosis of FSHD (Figure 5). Despite the high homology between 4q35 and 10q26 regions, FSHD has been exclusively associated with reduction of D4Z4 repeat units on chromosome 4q [Bakker et al, 1995; van Deutekom et al, 1996a; Matsumura et al, 2002].

Normal subjects carry more than 10 D4Z4 units originating from chromosome 4q; whereas array of 8 D4Z4 units or shorter are present in the majority of and familial FSHD patients [Upadhyaya et al, 1993; Wijmenga et al, 1994; Lunt et al, 1995a](Figure 4).

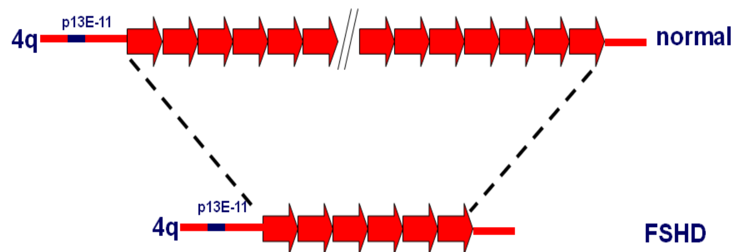

Figure 4: Deletion of a defined number of D4Z4 units on chromosome 4 in FSHD patients. Schematic representation of the D4Z4 repeats on chromosome 4 of a normal individual and a FSHD patient. Having less than 11 D4Z4 units is considered pathological.

The chromosomal context in which the D4Z4 deletions arise is considered crucial for development of FSHD. Indeed in families in which a *de novo* FSHD case appears, the *de novo* 4q DRA co-segregate with the disease from the affected parent to the offspring. In 10%

of the population, translocated 4-type repeats reside on chromosome 10q and, viceversa, translocated 10-type repeats on chromosome 4q are equally frequent [van Deutekom et al, 1996a], but only D4Z4 Reduced Alleles (DRA) located at 4q are associated with FSHD.

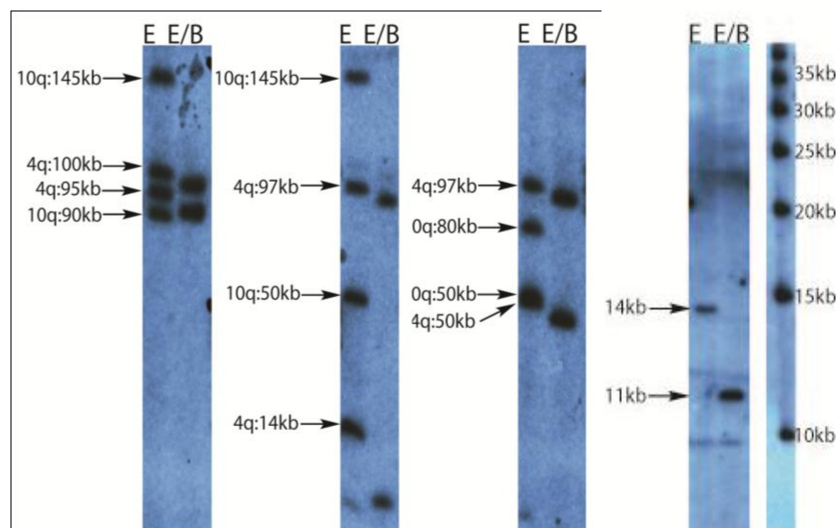

Figure 5: **Analysis of allele polymorphism of EcoRI restriction by Southern blot with probe p13E-11: A) Pulsed Field Gel Electrophoresis (PFGE) and B) Linear Gel Electrophoresis (LGE).**
The figure shows the electrophoresis of DNA samples digested with the restriction enzyme EcoRI (E) or EcoRI / BlnI (E / B) and hybridized with the probe p13E-11. One may observe the variability of the bands originated in chromosomes 4 and 10; in the lanes where samples are digested with the two enzymes EcoRI/BlnI only the bands originating from chromosome 4 are observed. The fragments are compared with the bands of the reference marker molecular weight. The fragments originated from chromosome 4 and less than 35 Kb are diagnostic for FSHD.

Moreover, it is believed that at least one D4Z4 repeat at chromosome 4q is necessary to develop FSHD, as monosomy of the very distal 4q35 region was shown not to cause disease [Tupler et al, 1996]. All these observations also suggest that the D4Z4 array is highly recombinogenic. Consistently, *de novo* mutations account for a surprisingly high percentage of FSHD patients (10-33%) [Zatz et al, 1995] and somatic mosaicism for a rearrangement of D4Z4 was found in as much as 3% of the general population.

1.4.3 The identification of specific molecular signature uniquely associated with FSHD expression

Since there are individuals with reduced D4Z4 alleles that do not have clinical signs of FSHD, it has been proposed that additional DNA sequences flanking the D4Z4 repeat array are necessary for disease development.

In 2002 a polymorphic segment of 10 kb directly distal to D4Z4 existing in two allelic forms, 4qA and 4qB, was identified [van Geel et al, 2002](Figure 6). Although both alleles are equally common in the general population, it was reported that FSHD is solely associated with the 4qA allele. Lemmers and coworkers, in 2002, analyzed 80 healthy controls and 80 unrelated individuals with FSHD for the presence of 4qA and 4qB alleles. In the controls they observed almost equal frequencies of 4qA and 4qB alleles (42% and 58% respectively) on chromosome 4, but only alleles of the 4qA type on chromosome 10. By contrast, in the 80 unrelated individuals with FSHD (44 *de novo* cases and 36 unrelated familial cases) they detected D4Z4 reduced alleles exclusively in chromosomes 4 bearing the 4qA allele, and never in those with the 4qB allele [Lemmers et al, 2002]. Subsequently, [Lemmers et al, 2004b] described three families with FSHD in which each proband of the family carried two FSHD-sized alleles and was heterozygous for the 4qA/4qB polymorphism. Segregation analysis demonstrated that FSHD-sized 4qB alleles were not associated with disease, since these were present in unaffected family members. Thus, the authors supposed that, in addition to a contraction of D4Z4, other cis-acting elements on 4qA might be necessary for the development of FSHD. Alternatively, they proposed that elements present at 4qB sub-telomeres might prevent pathogenesis of FSHD.

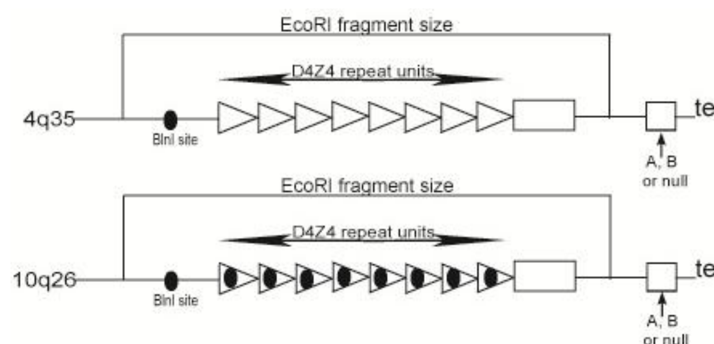

Figure 6: Schematic representation of 4qA/4qB polymorphism.

In 2007, the identification of additional sequence variations in a relatively stable Simple Sequence Length Polymorphisms (SSLP) proximal to the D4Z4 repeat was identified in the FSHD locus [Lemmers et al, 2007]. On the basis of the proximal SSLP, it was possible to distingush at least 17 genetically distinct sub-telomeric variants of chromosome 4 and 8 sub-telomeric variants of chromosome 10. It was supposed that only D4Z4 contractions in three genetically almost identical chromosome 4 sub-telomeres, the common variant 4A161 and the rare variants 4A159 and 4A168, are responsible for developing FSHD, whereas contractions in other 4q sub-telomeres were not associated with the disease. The authors presented the pedigrees of two FSHD families in which two different D4Z4 reduced allele segregate, reporting that subjects carrying the D4Z4 reduce allele with non-permissive 4qA166 haplotype did not manifest signs of disease.

Finally, it has been suggested that FSHD patients carry specific single nucleotide polymorphism (SNP) ATTAAA in the chromosomal region distal to the last D4Z4 repeat in the pLAM sequence of the 4qA alleles, that provides a PolyAdenylation Signal (PAS) [Lemmers et al, 2010]. Thus, the specific molecular signature, named 4A(159,161,168)PAS, has been proposed to define genetic background responsible for FSHD pathogenesis (Figure 7).

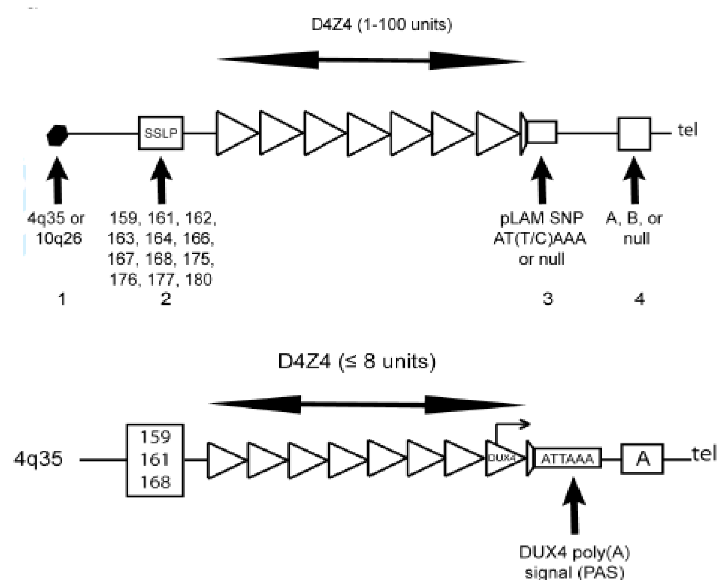

Figure 7: Schematic representation of polymorphisms at the 4q and 10q subtelomeres. **a)** The D4Z4 repeat array within the subtelomere of chromosomes 4q and 10q varies in size between 1 and 100 D4Z4 units (3.3–330 kb) and is indicated with triangles. Elements that distinguish subjects include: 1) The chromosomal localization of the D4Z4 repeat, chromosome 4q35 or 10q26. 2) The Simple Sequence Length Polymorphism (SSLP), a combination of five Variable Number Tandem Repeats, an 8 bp insertion/deletion, and two SNPs localized 3.5 kb proximal to D4Z4 and varies in length between 159 and 180 bp. 3) Single nucleotide polymorphism AT(T/C)AAA (SNP) in the pLAM region. 4) A large sequence variation (termed 4qA or B) that is distal to D4Z4. 4q chromosomes which do not hybridize to probes for A and B are termed "null" and their sequences vary from case to case. **b)** Schematic representation of the permissive chromosomal background. The ATTAAA variant creates a polyadenylation signal that stabilizes the DUX4 transcript and has been postulated to be the critical factor causing FSHD.

This specific background is characterized by (1) reduction of D4Z4 elements associated with (2) the 4qA (161/159/168) haplotype (3) and a single nucleotide polymorphism, ATTAAA, in the pLAM sequence.

1.4.4 Correlation studies between the size of D4Z4 allele and clinical expression in FSHD

Since the discovery of the FSHD molecular defect, several genotype-phenotype studies were conducted in order to evaluate if the size of the D4Z4 allele could be correlated with the clinical manifestations and to assess the impact of the molecular defect on the phenotypic expression.

Lunt and coworkers in 1995 [Lunt et al, 1995b] reported the analysis performed on 14 FSHD families and 25 clinically isolated cases, presumed to be due to new mutation, associated with D4Z4 reduced allele (respectively in two groups the range in EcoRI allele size was 19-30 and 13-24 kb). The study revealed a clear correlation between smaller fragment sizes and earlier age at onset. The median age at onset in sporadic cases resulted 6.9 years and in familial cases 18 years. Interestingly, the authors observed a difference in reported ages at onset between generations within families. They detected onset age appearing earlier in successive generations, even though it was hypothesized that this trend might be more a reflection of ascertainment bias than a biological anticipation. Their study revealed also correlation between the age of loss of ambulation and D4Z4 allele size [Lunt et al, 1995a]. The authors proposed that FSHD families could be divided broadly into three groups, necessarily with some overlap: i) new mutation cases carrying small fragment size (≤18 kb) with early onset and severe presentation of disease; ii) large ‘’classical’’ families with median onset age ranging from 8-22 years carrying alleles from 19 to 30 kb; iii) small families, often with a later onset presentation (median 15-23 years), or scapulohumeral presentation, in which a 30-38 kb D4Z4 reduced allele (DRA) segregates and non-penetrance of disease may still be observed above 20 years of age.

The subsequent study of Tawil and coworkers in 1996 [Tawil et al, 1996] confirmed the correlation by examining clinically and genetically well-defined 157 FSHD subjects. In particular, this analysis showed the presence of anticipation and that the size of the deletion and the disease severity were closely associated.

Three years later, Ricci et al. [1999], on a cohort of 165 patients with FSHD (carrying EcoRI allele size range 10-27 kb), further reported the inverse correlation between fragment

size and clinical severity. The probability of developing a severe form of disease resulted 100% in the presence of very short fragment (10-15 kb), decreased to 54% in patients carrying fragments from 16 to 20 kb and dropped to 21% or less in patients carrying fragment larger than 20 kb. The severe form of FSHD was defined when disease is expressed with severe weakness of pelvic and proximal leg muscle or both (strength less than 3 in at least one of these muscle) with inability to stand up from a chair without using support or to walk unaided, or wheelchair use.

In the genotype-phenotype correlation study performed by Tonini et al. in 2004 on 238 subjects from 106 unrelated families, it was observed that individuals with larger fragments showed a milder disease course while those who have the smaller ones were more severely affected. However, when genders are analyzed separately, this correlation was significant for females but not for males.

In 2003, Butz and coworkers conducted a systematic study of 39 unrelated FSHD patients with borderline D4Z4 repeat numbers and 102 healthy controls, in order to identify the molecular diagnostic cut-off point between FSHD cases and the control population and describe the phenotype in patients with borderline D4Z4 repeat numbers. The study revealed that there was not a definite D4Z4 diagnostic cut-off criteria separating FSHD, FSHD-like myopathies and healthy controls, without expected correlation of D4Z4 repeat number and disease severity. Therefore the authors suggested the D4Z4 cut off should be 8 repeats or 35 kb [Butz et al, 2003].

Collectively, reported studies showed a rough inverse correlation between the number of D4Z4 repeats and the severity of the disease. Alleles with 1-3 D4Z4 repeats (EcoRI fragment <20 kb) are generally associated with a severe form of disease that presents in childhood, 4-7 D4Z4 repeats (EcoRI fragment 21-34kb) with the classical form of FSHD, and 8-10 D4Z4 repeats (EcoRI fragment 35-45 kb) with a milder disease [Lunt et al, 1995a; Tawil et al, 1996; Ricci et al, 1999]. In addition, EcoRI alleles between 38-45 kb in size (9-11 D4Z4 repeats) have been detected in normal individuals and myopathic affected by FSHD or by other muscle diseases [Butz et al, 2003].

1.4.5 Facioscapulohumeral muscular dystrophy in subjects carrying full-length D4Z4 alleles (FSHD2)

The cohort of patients, termed "FSHD2" display D4Z4 alleles of normal size on both 4q alleles and they fulfill clinical diagnostic criteria for FSHD [Tawil et al, 2010], resulting

clinically indistinguishable from FSHD patients carrying D4Z4 reduced allele (also defined FSHD1).

In 2010 de Greef and coworkers have performed a cross-sectional study on 33 patients with FSHD2 from 27 families, the largest cohort described to date. The clinical presentation of FSHD2 patients appeared identical to the FSHD1. Out of 33 FSHD2 patients 20 (61%) were male. The average age at symptom onset was 26 years (range 0–60), which is almost 10 years later than in FSHD1. The reported initial symptom was scapular weakness in 61%, foot dorsiflexor weakness in 27%, facial weakness in 10%, and hip girdle weakness in 3%. A gender differences in disease severity in FSHD2 was not observed. Interestingly, notable difference between FSHD1 and FSHD2 is the mode of inheritance. In fact, the analysis showed that the majority (20/33, 67%) was sporadic, 11 were familial, and in 2 the inheritance pattern was uncertain, suggesting that the familial to sporadic ratio in FSHD2 is inverse to the ratio in FSHD1. Of the familial cases, 3 resulted dominant in inheritance (parent-child pairs) and 2 seemed recessive in inheritance (sibling pairs).

Therefore, it has been suggested that similar epigenetic and molecular mechanisms, also supposed to be pathogenic in FSHD associated with D4Z4 repeat contraction, are involved. In particular, it was observed that, unlike FSHD1, patients with FSHD2 showed loss of DNA methylation on both chromosome 4q and 10q D4Z4 repeats, suggesting that a defect in establishing or maintaining the D4Z4 repeat chromatin structure may cause FSHD2 [de Greef et al, 2009]. In the study performed by de Greef and coworker in 2010, all patients with FSHD2 (33 subjects) carried at least 1 D4Z4 repeat on the permissive haplotype 4A161. The authors studied DNA methylation levels of the D4Z4 repeat at 4 methylation-sensitive restriction sites in patients with FSHD2 and their unaffected relatives. The analysis showed that at the 4 sites tested, a significant D4Z4 hypomethylation was present in patients with FSHD2 as well as in patients with FSHD1. In particular, patients with FSHD2 showed significant D4Z4 hypomethylation on chromosomes 4q and 10q, while in patients with FSHD1 significant loss of D4Z4 methylation is proposed to be restricted only to the contracted chromosome 4q. The levels of D4Z4 hypomethylation in patients with FSHD2 were not correlated with phenotype severity. To explain the high percentage of FSHD2 sporadic cases, it was hypothesized that FSHD2 development may result from two independent events: D4Z4 hypomethylation by an unknown mechanism and presence of at least one D4Z4 repeat on the permissive 4A161 haplotype.

1.5 4q35 region and models for FSHD pathogenesis

1.5.1 D4Z4 units

The D4Z4 units are members of a large family of 3.3 kb tandem repeat loci that are located on the short arm of the acrocentric chromosomes, the pericentromeric regions (especially on chromosome 1), and the telomeric regions of the long arms of chromosomes 4 and 10 [Hewitt et al, 1994; Lyle et al, 1995].

The organization of the D4Z4 repeat is rather unusual (Figure 8), in particular the presence of two homeobox sequences within the same open reading frame. Homeobox genes, encoding homeodomain transcription factors, often play important roles in embryonic development.

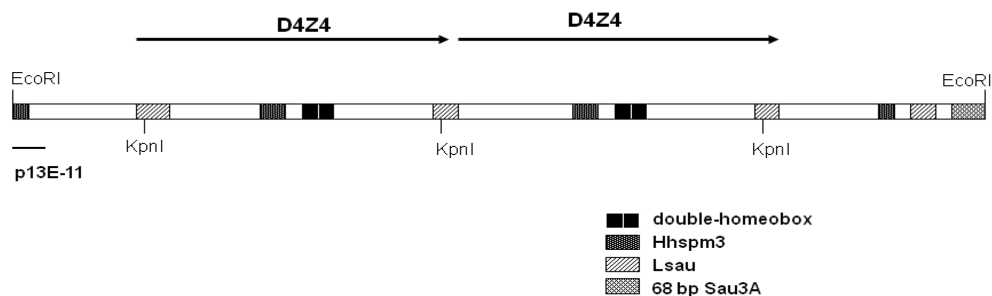

Figure 8: Schematic representation of the structure and the organization of the D4Z4 repeats. Each D4Z4 unit has LSau motifs, the hhspm3 motif and two homeobox sequences.

Because of the potential of D4Z4 to encode a protein, several groups have focused on searching for the transcripts from D4Z4. Although cDNAs and RT-PCR products containing closely related sequences have been identified, none originate from chromosome 4q35. For examples, several related cDNAs clones representing transcripts from acrocentric chromosome were isolated. However, these cDNAs contained in-frame stop codons within the predict homeodomains [Hewitt et al, 1994, Lyle et al, 1995]. Subsequently, clones that could encode such double homeodomain proteins have been identified; again these did not represent transcripts from D4Z4 but rather there are from other loci [Ding et al, 1998; Gabriels et al, 1999; Beckers et al, 2001].

Within each D4Z4 repeat, the homeoboxes are flanked by two classes of repetitive DNA; a GC-rich low copy repeat originally named hhspm3 [Zhang et al, 1987] and LSau [Agresti et al, 1987, 1989; Meneveri et al, 1993]. LSau repeats are dispersed, preferentially found in regions of heterochromatin and are often associated with 68bp (or β) satellite DNA [Meneveri et al, 1993]. On chromosome 4q35, 68 bp satellite DNA is not interspersed

between repeat units, but present as a block of about 8 kb immediately distal to D4Z4 [van Geel et al, 2002].

The analysis of the interaction between D4Z4 and nuclear proteins revealed the presence of a 27 bp binding site (DBE, D4Z4 Binding Element) able to recruit a multi-protein complex in vitro and in vivo comprising of YY1, HMGB2 and nucleolin [Gabellini et al, 2002]. The ubiquitous transcription factor Yin Yang 1 (YY1) is a recruiter of Polycomb group proteins (PcG), responsible for chromatin remodeling and epigenetic silencing in many fundamental biological processes. The activity of YY1 is modulated by histone deacetylases and histone acetyltransferases [Yao et al, 2001]. Studies suggest that the Trithorax group protein Ash1L, associated with transcriptionally active chromatin, is recruited by a non-coding RNA, called DBE-T (D4Z4 Binding Element Transcript) to the D4Z4 locus in FSHD patients in which D4Z4 deletion is associated with reduced Polycomb silencing thus promoting topological reorganization of the FSHD locus leading to de-repression of 4q35 genes [Cabianca et al, 2012].

1.5.2 Role of D4Z4 transcription in FSHD pathogenesis

Each D4Z4 unit contains a putative promoter and a single open redaing frame (ORF) encoding a putative double homeobox gene, named DUX4 [Gabriels et al, 1999; Hewitt et al, 1994; Lyle et al, 1995] (Figure 9).

Figure 9: Localization of the DUX4 gene within each D4Z4 unit. On the permissive chromosomes the last copy of the DUX4 gene spices to the third region immediately flanking and stabilizing the transcript owing to the presence of the poly(A) signal (PAS).

Homeodomain proteins are important for many early and late developmental processes. Thus, DUX4 is considered a strong candidate for FSHD pathogenesis. Initially, a putative promoter and the putative double homeodomain gene DUX4 were identified within each D4Z4 repeat unit. As D4Z4 was considered to be of heterochromatic nature, it was hypothesized that partial deletion of the D4Z4 repeat array resulted in destabilization of the D4Z4 heterochromatin and in the inappropriate upregulation of DUX4 [Hewitt et al, 1994; Gabriels et al, 1999]. DUX4 overexpression may induce cell death by apoptosis, induce caspase 3/7 activation and alter emerin distribution at the nuclear envelope [Kowaljow et al, 2007]. In addition, DUX4 overexpression may activate PITX1 (paired-like homeodomain

transcription factor 1), as was determined for both a reporter gene fused to the Pitx1 promoter and the endogenous Pitx1 gene. Interestingly, upregulation of the PITX1 protein was also observed in muscle biopsies of patients with FSHD [Kowaljow et al, 2007; Dixit et al, 2007]. Nevertheless, for a long time, the functionality of the DUX4 gene was questioned, because of lack of introns and polyadenylation signals and absence of evidence for in vivo transcription [Hewitt et al, 1994; Gabriels et al, 1999; Winokur et al, 2003; Osborne et al, 2007; Alexiadis et al, 2007]. However in the following years, D4Z4 homologues have been identified in several mammalian species and it was established that the DUX4 open reading frame (ORF) shows evolutionary conservation, disputing the non-functionality of DUX4 and suggesting a coding role, possibly during development. Interestingly, not only the ORF of DUX4, but also their organization in an array is evolutionary conserved [Clapp et al, 2007]. Importantly, this study provided evidence for bidirectional transcription of the mouse DUX4 array [Block et al, 2012]. Next, expression of two different DUX4 transcripts in cells transfected with D4Z4 elements and in FSHD myoblasts was reported. The first transcript lacks introns and is transcribed from internal D4Z4 repeat units, while the second transcript has two introns and is transcribed from the most distal D4Z4 repeat unit. Interestingly, the pLAM sequence distal to the second transcript may provide a polyadenylation signal [Kowaljow et al, 2007; Dixit et al, 2007].

Thus far, DUX4 expression seems to be restricted to FSHD myoblasts [Kowaljow et al, 2007; Dixit et al, 2007] and testicles [Young et al, 2013]. As most homeodomain proteins have a function as transcriptional regulators in developmental processes, DUX4 expression may normally be restricted to embryogenesis [Deschamps et al, 1992]. In fact, the DUX4 homeodomain shares high homology with the homeodomain of the proteins Pax3 and Pax7, which are involved in the development of skeletal muscle [Buckingham et al, 2007]. As FSHD is specifically linked to the 4qA161 haplotype [Lemmers et al, 2007], sequence variations residing within or close to the D4Z4 repeat array may play a role in the regulation of DUX4 transcription. More detailed analysis of DUX4 expression shows that the DUX4 pre-mRNA can be alternatively spliced and it has been suggested that the FSHD muscle expresses a different splice form of DUX4 mRNA compared to control muscle [Snider et al, 2011], however Jones and coworkers [2012] observed the DUX4 mRNA and protein expression in muscle biopsies and myogenic cells from genetically unaffected relatives of the FSHD. Recently, transgenic mouse models carrying human genomic constructs with the FSHD subtelomeric region permissive for somatic DUX4 expression were generated [Krom et al, 2013]. Data suggest that these mice maintain the transcriptional profile of the DUX4

retrogene as observed in FSHD patients and controls. However they do not show an obvious muscle pathological phenotype. Therefore the role of DUX4 in muscle disease still needs to be clarified.

1.5.3 Expression of proximal genes controlled by D4Z4

It was hypothesized that D4Z4 repeat array is involved in controlling the expression of genes located proximally on 4q35 [Hewitt et al, 1994]. Indeed we showed that a 27 bp binding element (DBE, D4Z4 Binding Element) within D4Z4 recruits a multi-protein complex comprising of YY1, HMGB2 and nucleolin and the reduced expression of these proteins leads to the anomalous derepression of genes proximal to the repeat array [Gabellini et al, 2002]. Recently the sequence of a non-coding RNA, called DBE-T (D4Z4 Binding Element Transcript), located in a proximal region of the repeat array, has been described. Studies showed that myoblasts from FSHD patients express this lncRNA, DBE-T, which promotes the recruitment of the Trithorax group protein Ash1L, associated with transcriptionally active chromatin, to the D4Z4 locus thus promoting topological reorganization of the FSHD locus leading to de-repression of 4q35 genes [Cabianca et al, 2012].

Four genes have been identified within the 4q35 region (Figure 10) and are considered, interesting candidates for FSHD pathogenesis.

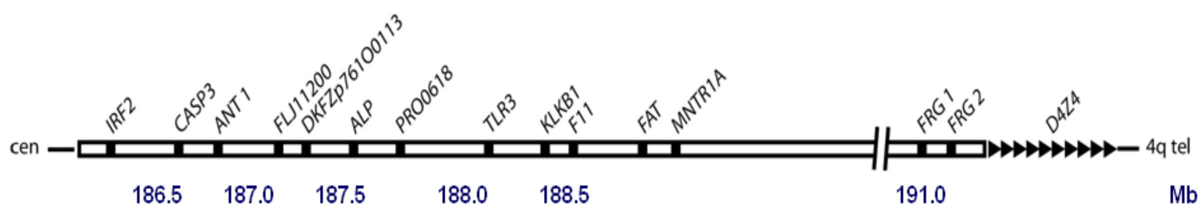

Figure 10: Schematic representation of 4q35 showing physical distances between the genes (Gabellini et al., 2002).

FSHD Region Gene 1 (FRG1) is located 100 kb proximal the D4Z4 sequences toward the centromere [van Deutekom et al, 1996b] (Figure 10). FRG1 has been shown to be crucial for muscle development in vertebrates and invertebrates [Hanel et al, 2009; Liu et al, 2010] and most importantly its over-expression in mice causes a dystrophic phenotype sharing several features with FSHD patients [Gabellini et al, 2006]. First, as in FSHD patients, FRG1 overexpression induces a progressive myopathy in which sarcolemma is not damaged and no elevated creatine kinase levels are detected. Second, FRG1 over-expressing mice display a reduced tolerance to exercise, taking a shorter time to reach exhaustion, similarly to the

muscle fatigue described by FSHD patients [Gabellini et al, 2006]. Third, only selective muscles in FRG1-overexpressing mice are dystrophic and display altered accumulation of myosin heavy chain (MyHC) [D'Antona et al, 2007, Sancisi et al, 2014]. Fourth, it has been shown that in muscles of both FSHD patients and FRG1 transgenic mice, specific pre-mRNAs undergo aberrant alternative splicing [Gabellini et al, 2006; Pistoni et al, 2013, Sancisi et al, 2014]. Finally, it has been recently shown that FRG1 overexpression impairs myogenic differentiation [Neguembor et al, 2013].

Additional genes located at 4q35 were found to be transcriptionally upregulated in FSHD muscle [Gabellini et al, 2006]. FRG2 (FSHD Region Gene 2) mapping 37 kb proximal to the D4Z4 repeat array (Figure 10) is a gene of unknown function, expressed only in myoblasts of FSHD patients and its transcript is undetectable in normal individuals. The interest in FRG2 has been dropped because its over-expression in muscles of transgenic mice did not cause myopathy [Gabellini et al, 2006], although recent studies show that FRG2 and DUX4 present the same expression pattern, suggesting that they reside in the same chromatin domain [Stadler et al, 2013].

ANT1 (Adenine Nucleotide Traslocator 1), positioned 4.8 kb to the proximal D4Z4 (Figure 10) encodes for a mitochondrial protein, which is the major component of the inner mitochondrial membrane and is involved in the response to oxidative stress. Mutations of ANT1 cause Progressive External Ophtalmoplegia and have been found in subjects with autosomal recessive mitochondrial disorder characterized by cardiomyopathy and myopathy [Echaniz-Laguna et al, 2012]. However, transgenic mice overexpressing ANT1 in muscle were normal [Gabellini et al, 2006].

FAT1 gene [Caruso et al, 2013] is located 3.6 Mb from the D4Z4 repeat array on 4q35 (Figure 10). FAT1, protocadherin gene in mouse is required in migrating muscle precursors and altered muscle shapes caused by Fat1 mutations are predictive of early onset defects in muscle integrity in adult mutants. The topography of muscle abnormalities caused by Fat1 loss-of-function resembles that of human patients with FSHD. Muscle-specific reduction of FAT1 expression and promoter silencing was observed in rare cases of biopsies from foetuses with a prenatal diagnosis of FSHD1 [Caruso et al, 2013].

1.5.4 Epigenetic hypothesis: chromatin modifications in FSHD

Each D4Z4 repeat unit harbors two classes of GC-rich sequences, hhspm3 and LSau, which are predominantly found in heterochromatic regions of the genome. This led to the hypothesis that deletion of D4Z4 repeats might modify the chromatin organization of the 4q subtelomeric region and alter gene expression [Winokur et al, 1994] (Figure 11).

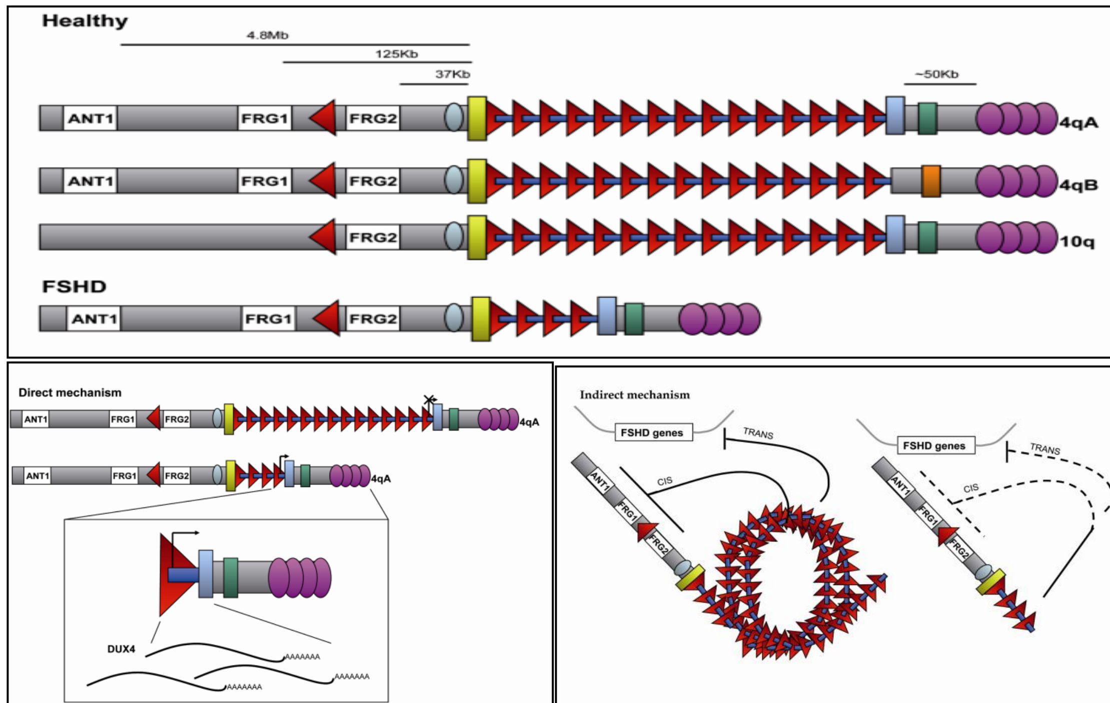

Figure 11: **Models for the molecular basis of FSHD.**
A) Healthy individuals carry 11–150 units of D4Z4, whereas FSHD patients have less than 11 repeats.
B) DIRECT MECHANISM: reduction of D4Z4 repeat array leads to the synthesis of DUX4 transcript, which is normally not transcribed, through changes in D4Z4 heterochromatin and/or stabilization of DUX4 mRNA.
C) INDIRECT MECHANISM: the reduction of D4Z4 repeats leads to modifications of the spatial and structural organization of chromatin generating changes of transcriptional control over the expression of candidate genes localized in cis or in trans.

Many studies indicate that the D4Z4 locus displays a chromatin structure more similar to euchromatin [Zeng et al, 2009]. Histone H4 acetylation levels at the p13E-11 region immediately proximal to D4Z4 were similar to those observed in the 5' regions of the FSHD candidate genes FRG1 (FSHD region gene 1) and ANT1 (adenine nucleotide translocator 1) and did not differ significantly between control and FSHD lymphoid cells. In conclusion, these results suggested that the nature of D4Z4 chromatin is that of unexpressed euchromatin rather than that of costitutive heterochromatin [Jiang et al, 2003]. These observations favored the hypothesis that this region might be more dynamic than expected. In FSHD1 and FSHD2 patients was observed loss of marks of unexpressed heterochromatin such as histone H3K9me3. Interestingly, the heterochromatin binding protein HP1, which mediates

transcriptional silencing, and the sister chromatid cohesion complex, cohesin, bind to D4Z4 in an H3K9me3-dependent manner and their recruitment is seriously compromised in FSHD [Zeng et al, 2009].

Interestingly in all the proposed model epigenetic changes such as methylation or histone modifications are used as an additional level of complexity that might help interpreting the complex correlation between genotype and phenotype in FSHD.

Considering that methylation status of CpG sites could play a critical role in chromatin configuration, D4Z4 methylation status was investigated. D4Z4 methylation was analyzed firstly at EagI sites in lymphoblastoid cell lines of patients with the ICF syndrome (Immunodeficiency Centromeric region instability Facial anomalies) who carry mutations in the DNA-methyltransferase 3B (DNMT3B) gene [Kondo et al, 2000] and healthy controls. That analysis showed that D4Z4 repeats are highly methylated in controls, while they resulted hypomethylated in ICF patients. This finding is of great interest because DNA Methyl-Transferase 3B (DNMT3B) contributes to *de novo* DNA methylation in early embryogenesis [Lana et al, 2012]. Consequently investigators tested whether D4Z4 are hypomethylated in FSHD.

The first study surveyed methylation at SmaI, MluI, SacII, and EagI methylation-sensitive restriction sites in FSHD and normal blood and skeletal muscle samples of FSHD and ICF (Figure 12). D4Z4 was found highly methylated in both normal and FSHD lymphoblasts, as well as in somatic tissues, including skeletal muscle. Nevertheless the study did not discriminate methylation status of the D4Z4 repeat array at chromosome 4 and chromosome 10 [Tsien et al, 2001].

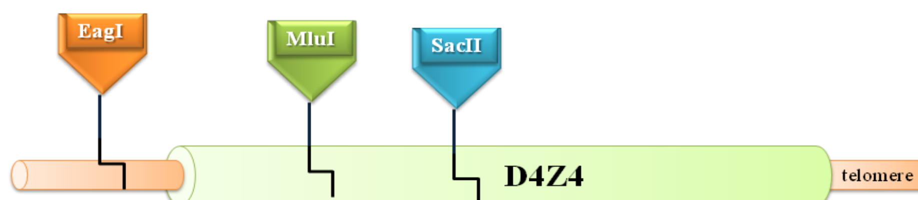

Figure 12: The first methylation study on D4Z4 with restriction enzymes MluI, SacII and EagI.

Subsequently DNA methylation was examined at two methylation-sensitive restriction sites, BsaAI and FseI (Figure 13), in the most proximal unit of D4Z4 array at 4q35, which was considered representative for the entire array [van Overveld et al, 2003].

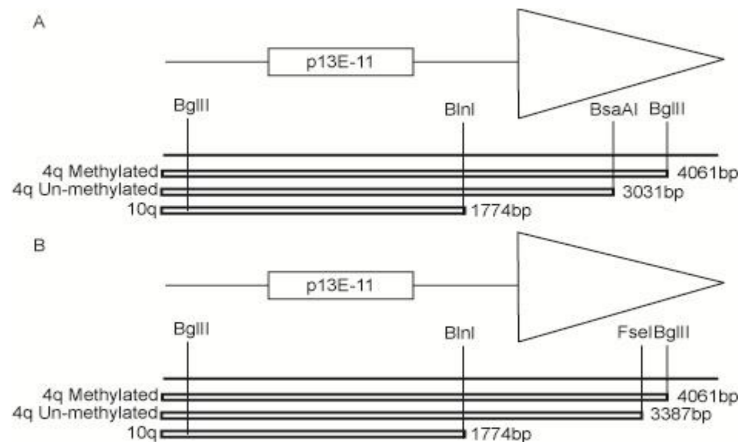

Figure 13: The two methylation-sensitive restriction sites in the proximal D4Z4 repeat on chromosome 4q; a) BsaAI and b) FseI.

Through this approach the D4Z4 methylation level can be assessed on both chromosomes 4, excluding chromosome 10 (Figure 13). The limitation of this test is due to possibility to analyze D4Z4 methylation status only in individuals carrying standard allele constitution of 4-type repeat units on chromosome 4 and 10-type on chromosome 10 (disomic), or on individuals carrying one array of 10-type repeat units at normal sized chromosome 4 (monosomic)(see Figure 14).

van Overveled and coworkers detected normal level of methylation in the control group, significant hypomethylation at both methylation sensitive sites in FSHD1 patients and similar level of hypomethylation in their non-penetrant relatives, carrying same D4Z4 reduced allele. Interestingly, in FSHD2 patients level of D4Z4 methylation on both chromosomes 4 was strongly decreased, while it was equivalent among unrelated individuals affected with muscular dystrophy different from FSHD and healthy controls [van Overveld, 2003]. Hypomethylation was pronounced in ICF patients who do not develop muscular dystrophy.

Further investigation of the methylation status in 21 monosomic FSHD1 patients and 19 monosomic healthy controls [van Overveld et al, 2005], performed to obtain exact level of methylation of single D4Z4 contracted allele and healthy allele, confirmed D4Z4 hypomethylation in FSHD1 subjects.

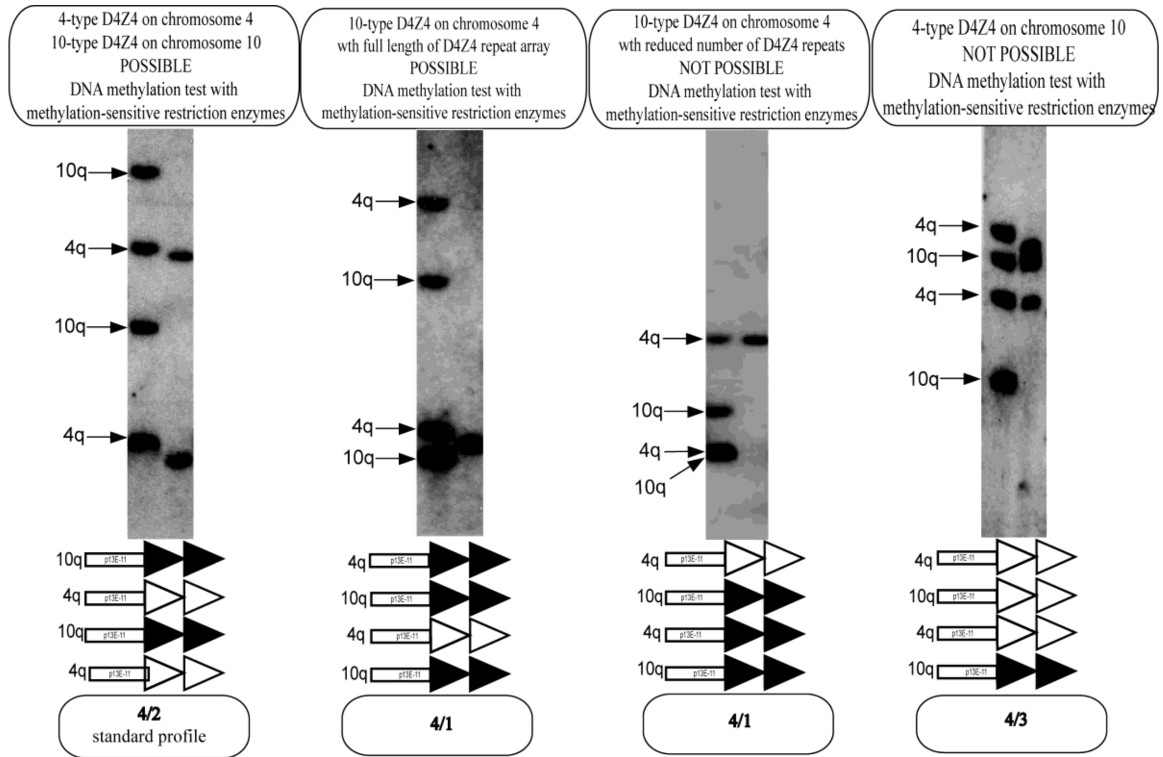

Figure 14: The presentation of 4q-10q allelic constitution.

De Greef and coworkers [2009] performed a detailed methylation assay at CpoI methylation-sensitive restriction site. This assay enables surveying the D4Z4 methylation status of most proximal and internal D4Z4 repeat units on chromosomes 4q and 10q separately. The authors confirmed previous methylation analysis at two sites in most proximal D4Z4 repeat unit as valid.

Considering that BsaAI and FseI analysis permit only to quantify methylation at two CpG dinucleotides of the first D4Z4 repeat of the tandem array recently a new methodology has been proposed for studying methylation status on D4Z4 repeats. By using bisulfite sequencing Hartweck and coworkers [2013] measured methylation level in three regions of the D4Z4 sequence (Figure 15). Interestingly bisulfite sequencing revealed that only one region, DR1, was hypomethylated in FSHD patients confronting to controls. This test provide a global survey of the average methylation level on both chromosomes 4q and both chromosome 10q, without permitting precise analysis of D4Z4 methylation on 4q allele of reduced size.

Figure 15: Schematic presentation of three domains analyzed by bisulfate sequencing by Hartweck et al [2013]

1.6 Recent findings from different genotype-phenotype studies and clinical reports on FSHD patients and families revealed several exceptions to the general outline

In pre-molecular era, the first observations performed on large families with clinical diagnosis of FSHD suggested an almost complete penetrance of the disease [Bailey et al, 1986; Becker 1953; Tyler and Stevens 1950; Chung and Morton 1959]. However, since the molecular test for FSHD has been put in place, subjects carrying a D4Z4 Reduced Allele (DRA) without signs of disease have been reported [Lunt et al, 1995a; Nakagawa et al, 1997; Zatz et al, 1995 and 1998; Ricci et al, 1999; Tonini et al, 2004; Sakellariou et al, 2012] challenging the notion that exclusively reduction of integral number of D4Z4 is responsible for full disease penetrance.

Two main studies regarding the FSHD penetrance were performed by Lunt and coworkers respectively in 1989 and 1991. The estimation of age dependent penetrance based on the presence of the characteristic clinical signs, was >95% by age 20, without difference between the families [Lunt et al 1989]. In particular, the penetrance of disease has been estimated <5% for ages 0 to 4 years, 21% for ages 5 to 9, 58% for ages 10 to 14, 86% for ages 15 to 19, and 95% penetrance for age 20 years and over at 95% for patients aged ≥ 20 years [Lunt et al, 1989]. The mode of inheritance was observed to be autosomal dominant.

Lunt et al. [1995a] from analysis of 30 FSHD families (carriers of 19 to 38 kb fragment size), observed small families with 4q35 cosegregating fragments of 30 kb and 34 kb in which subjects who inherit this fragment resulted clinically unaffected above age 20 years, supporting a lower penetrance in these families. The authors hypothesized that in families with a larger fragment size, >30 Kb, penetrance of the FSHD gene was lower than previous estimated value of 95% above age 20 years.

In the study of Zatz et al. [1995] on 34 Brazilian FSHD families, two unrelated multigenerational families with multiple affected patients were described, where the parents of affected patients (a female aged 46 years and a male aged 79 years) resulted asymptomatic carriers of FSHD allele; the author did not report the correlation with the fragment size.

In the subsequent study of Zatz et al. [1998], the estimated penetrance for FSHD allele on 52 families with fragment smaller than 35 kb was 85% for patients until age 30; furthermore, when the authors considered the sexes separately, the estimated penetrance of the FSHD allele resulted significantly greater for males (95%) than for females (69%). Interestingly, among 27 families with at least two clinically affected patients it was observed

that in 21 the pattern of inheritance was autosomal dominant (four of them with incomplete penetrance). Three pedigrees presented patterns of inheritance compatible with autosomal recessive trait, considering at least two affected sibs born from asymptomatic parents.

Ricci et al. [1999] reported 7 subjects, age 20 to 69 years, carrying a DRA ranging in size between 21 and 37 kb, without symptoms or signs of FSHD, classified as non-penetrant carriers. In this study, unaffected individuals belonging to families where D4Z4 allele smaller than 20 kb segregates, were not observed.

Tonini et al. [2004], analyzing 238 subjects carrying DRA (<35 kb) from 106 unrelated families, observed that about 20% of individuals, relatives of FSHD patients, carrying the same DRA, remained asymptomatic or was minimally affected with a significantly higher proportion of females than males; asymptomatic carriers were found in about 30% of the analyzed families.

Lastly, the most recent work of Sakellariou and coworkers [2012] reports the clinical and genetic analysis of 133 individuals carrying DRA (71 probands and 62 relatives) from 71 unrelated Greek families, revealing a high percentage (almost 50%) of asymptomatic relatives carrying a contracted 4q allele, older than 30 years. The percentage of unaffected carriers was lower among males. Notable, among 38 multiple-case families, 16 (42%) were found to have at least one symptom-free individual, with a greater proportion of asymptomatic minimally affected gene carriers concentrating in some pedigrees. Significant statistical association was observed concerning the gender and the clinical manifestation of the disease; the percentage of the symptomatic patients was found to be 66.7% among females whereas among males was 86.6%.

1.6.1 Cross-sectional genotype-phenotype study of FSHD families with D4Z4 reduced allele [Ricci et al, 2013]

Several studies conducted on FSHD families have described a high variability in clinical expression among and within FSHD families, as well as asymptomatic subjects carrying DRA, doubting the notion of an almost full penetrance of disease.

Thus we considered it is important to establish the penetrance of disease in FSHD families carrying DRA and to ascertain previously reported rough and inverse correlation between the size of DRA and the age at onset and severity of disease expression [Ricci et al, 2013] (Appendix 1).

The selection process was conducted on 418 FSHD index cases carrying DRA with 1-8 repeats (Figure 16).

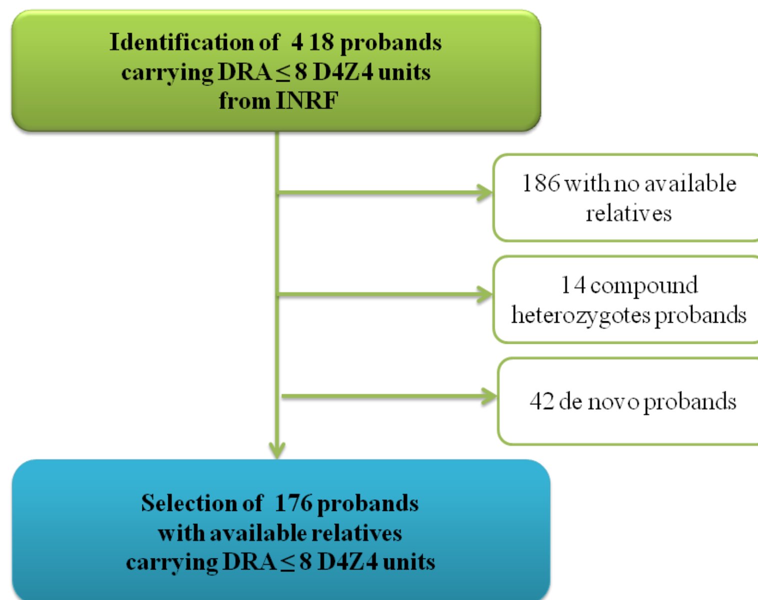

Figure 16: Preliminary selection of probands/families from the Italian National Registry for FSHD (INRF).

We divided subjects in three groups: subjects carrying DRA with 1-3 D4Z4 repeats; subjects carrying DRA with 4-6 D4Z4 repeats; subjects carrying DRA with 7-8 D4Z4 repeats (Figure 17).

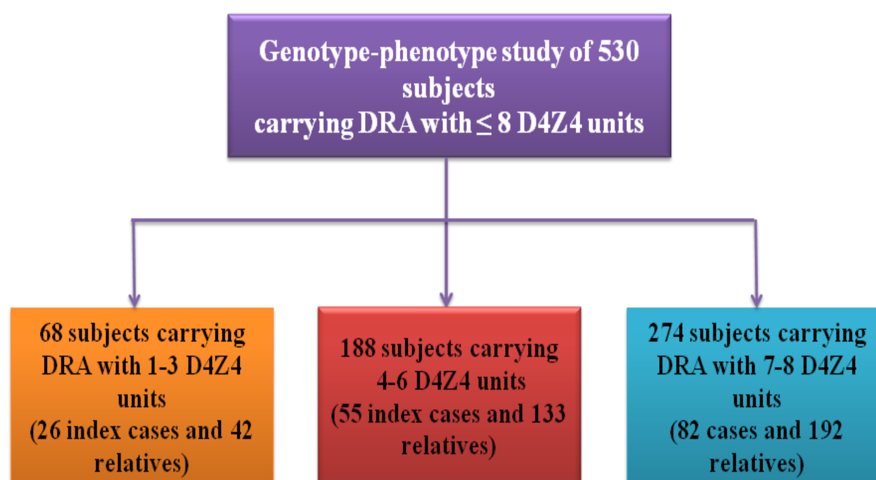

Figure 17: Selection of the cohort of probands and their relatives for genotype–phenotype correlation analysis.

The distribution of asymptomatic relatives was analyzed based on the size of DRA. Figure 18 shows that 9.5% (4 out of 42) of all carriers of DRA with 1-3 repeats did not

display motor impairment. This percentage increases among carriers of DRA with 4-6 and 7-8 repeats (28.6% and 39.6% respectively).

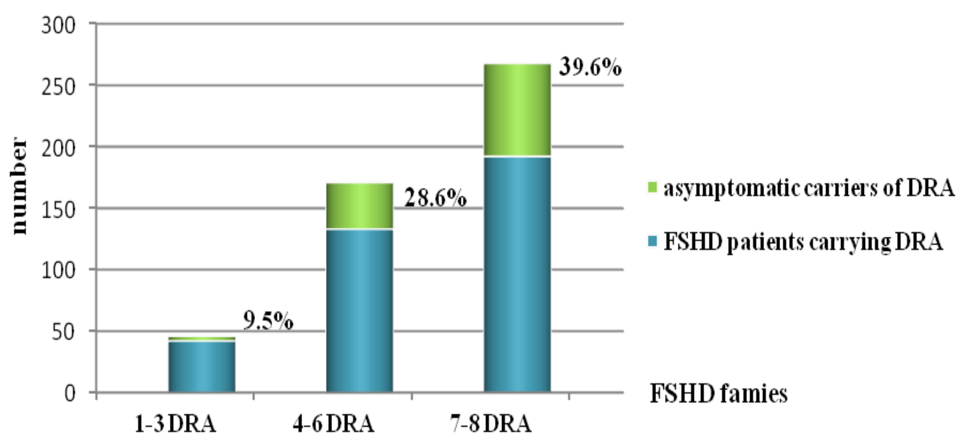

Figure 18: The distribution of asymptomatic DRA carriers in three groups.

In addition we calculated the distribution of asymptomatic carriers based on the age at examination: subjects aged between 18-30, 31-55, 56-70 years, and subjects over 70 years of age. As shown in Table 1 asymptomatic DRA carriers were found in all classes up to 70 years. In particular, almost one third of carriers of DRA with 4-6 and 7-8 repeats (27.6% and 35.9%, respectively) were asymptomatic between 56 and 70 years of age.

	Age (years)					
	18-30		31-55		56-70	
D4Z4 units	N of subjects	% score=0 (N)	N of subjects	% score=0 (N)	N of subjects	% score=0 (N)
1-3	8	12.5 (1)	23	8.7 (2)	9	11.1 (1)
4-6	31	25.8 (8)	65	33.8 (22)	29	27.6 (8)
7-8	42	54.8 (23)	85	40.0 (34)	39	35.9 (14)

Table 1: The percentage of asymptomatic DRA carriers in the group of 1-3, 4-6 and 6-8 DRA based at the age at onset.

Since the observed percentage of asymptomatic carriers varied among relatives carrying DRA of different sizes, the age-related risk of developing motor impairment was evaluated in correlation with D4Z4 size on the basis of data obtained and the age at onset of DRA carriers calculated with the Kaplan-Meier method (Kaplan and Meier, 1958).

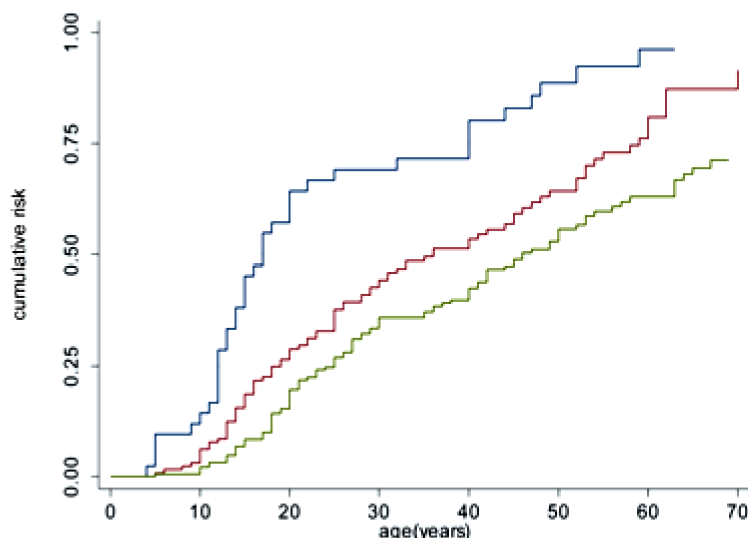

Figure 19: The age-related risk of developing motor impairment was evaluated in correlation with D4Z4 size on the basis of data obtained and the age at onset of DRA carriers. Blue line refers to carriers of 1-3 DRA, red line to 4-6 DRA carriers and green line to 7-8 DRA carriers. Long rank-test p value = 0.002.

Among subjects carrying DRA with 1-3 units the risk of developing motor impairment is 64.3% by age 20, 80.1% by age 40, 96.2% by age 60. Among subjects carrying DRA with 4-6 and 7-8 D4Z4 units these risks are 21.8% and 19.6% respectively by age 20, 44.8% and 42.5% by age 40, 71.5% and 62.9 % by age 60. Therefore, FSHD penetrance is almost complete by age 60 only for carriers of DRA with 1-3 units (Figure 19).

We tested whether the size of DRA correlates with age at onset and disease severity. Table 2 shows that the mean age at onset is statistically lower among subjects carrying DRA with 1-3 units (20.3 years) in comparison with those carrying DRA with 4-6 and 7-8 D4Z4 repeats (respectively 29.2 and 34.6 years) (p = 0.0002).

	Relatives			
D4Z4 units	N of subjects	Mean age at onset (yrs)	95% CI	p-value*
1-3	77	20.3	(15.5;25.2)	
4-6	41	29.2	(25.6:32.7)	
7-8	114	34.6	(30.1;39.1)	0.0002

Table 2: The age at onset in the group of 1-3, 4-6 and 7-8 DRA carriers.

Severity is also increased among carriers of DRA with 1-3 repeats. Indeed, as shown in Table 3, affected relatives carrying DRA with 1-3 repeats had a mean FSHD score of 7.2.

	Relatives			
D4Z4 units	N of subjects	Mean FSHD score	95% CI	p-value*
1-3	38	7.2	(5.8; 8.6)	
4-6	96	4.4	(3.8; 5.1)	
7-8	116	4.1	(3.5; 4.7)	0.0006

Table 3: The mean FSHD score in the group of 1-3, 4-6 and 7-8 DRA carriers.

By contrast, individuals carrying DRA with 4-6 and 7-8 D4Z4 units had mean FSHD score of 4.4 and 4.1 respectively. This association was statistically significant (p = 0.0006)

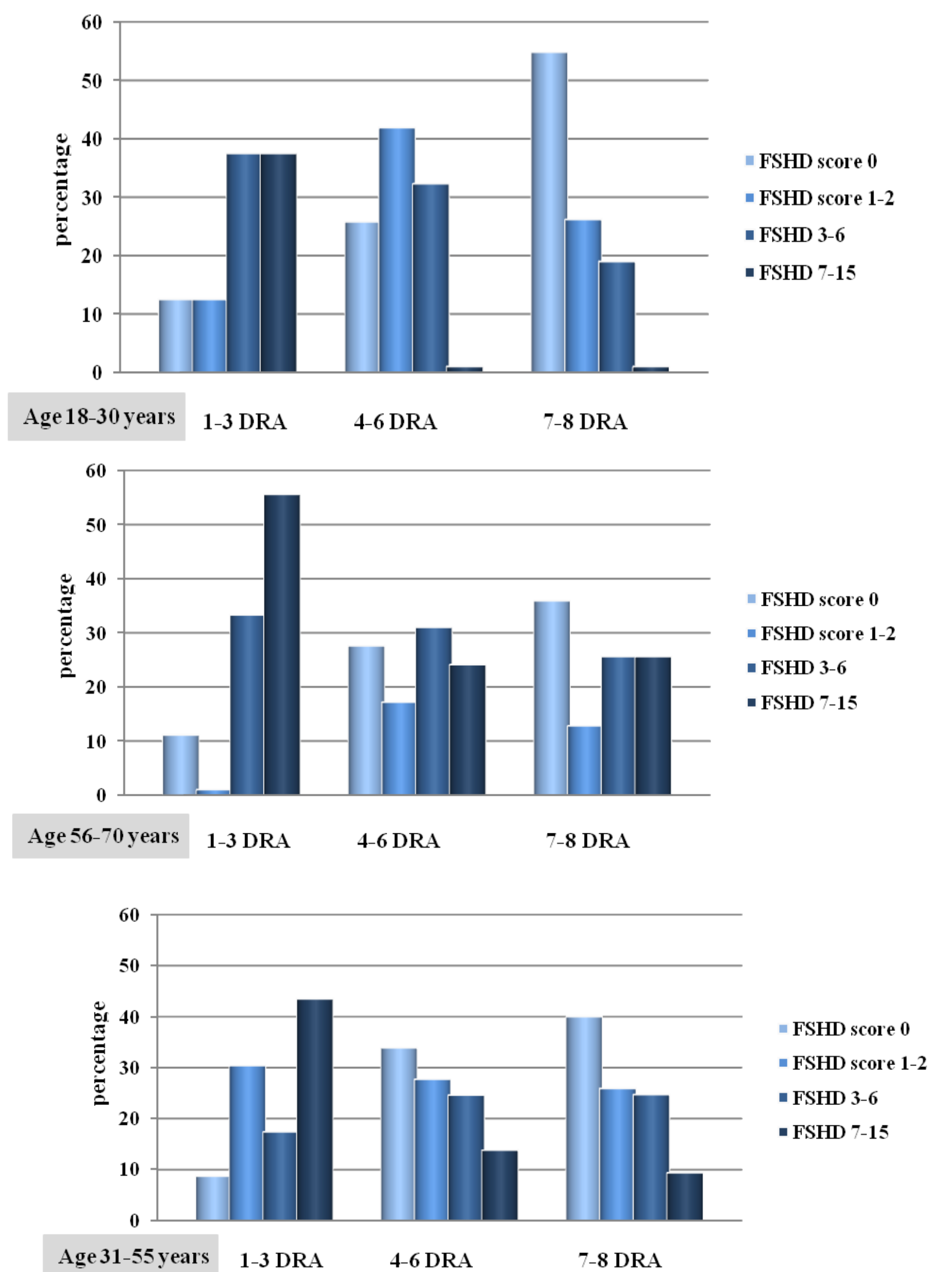

Figure 20: The degree of motor impairment among relatives was also evaluated in association with D4Z4 allele size and age at examination.

and was obtained by using linear regression model adjusted for age at examination.

The degree of motor impairment among relatives was also evaluated in association with D4Z4 allele size and age at examination. Figure 20 shows that approximately 40% of relatives carrying DRA with 1-3 units are severely affected (FSHD score ≥7) by age 30. In contrast, no relatives carrying DRA with 4-8 units had a FSHD score higher than 6 in this age window. Figure 17 shows that between age 31-55 and 56-70 a high percentage (ranging between 32% and 37%) of relatives carrying DRA with 4-8 units were asymptomatic (FSHD score equal to zero) or displayed minimal signs of functional motor impairment (FSHD score 1-2, ranging between 15% and 27%).

Our large scale genotype-phenotype study [Ricci et al, 2013] revealed that FSHD penetrance in DRA carriers is not complete by age 20, as previously proposed [Tawil et al, 2010], as asymptomatic carriers in all the classes of ages up to 70 years were found. It was shown that DRAs with 4–8 repeats have no definitive prognostic value, and that other prognostic parameters, beside DRAs, should be considered [Ricci et al, 2013]. Instead, the risk of developing the motor impairment by age 50 in FSHD family members is higher (83–93%) in subjects carrying DRA with 1–3 repeats. Instead, considering the cohort of relatives carrying DRA with 4–8 repeats, the risk of developing motor impairment is 48% for subjects with lower degree of kinship and raises to 55–63% for subjects with first degree of kinship with the proband. Interestingly, in our cohort, 19 of 148 FSHD families (13%) in which a DRA with 4–8 units segregates presented affected subjects only in one generation [Ricci et al, 2013]. In these cases the lack of autosomal dominant inheritance should prompt us to consider whether the disease develops because of the presence of additional genetic defect(s).

Collectively, the wide clinical variability among subjects carrying 4-8 D4Z4 repeats together with the high number of asymptomatic or minimally affected carriers suggests that additional factors, such as genetic, epigenetic and environmental factors, are involved in reaching the threshold of disease appearance and/or modifying the clinical outcome [Scionti et al, 2012a]. Remarkably, study shows a higher percentage of asymptomatic subjects between relatives with lower degree of relationship with proband, regardless age and D4Z4 size. The degree of kinship may influence the disease outcome, as result of genetic background "dispersion", suggesting a more complex mode of inheritance of FSHD.

2. Aims of the thesis

Facioscapulohumeral muscular dystrophy (FSHD [MIM 158900]) is the third most common muscular dystrophy with an estimated prevalence of 1:20,000 [Mostacciuolo et al, 2009]. FSHD is considered an autosomal dominant disorder [Flanigan et al, 2004] with a typical onset within the second decade of life [Lunt et al, 1989; Tawil et al, 2010]. The disease presents a remarkably wide variety of phenotypic expression, ranging from almost asymptomatic subjects to severe wheelchair-dependent patients [Lunt and Harper, 1991; Ricci et al, 2013]. The classical FSHD phenotype, firstly described as an independent nosological entity in 1884 by Landouzy and Dejerine [Landouzy and Dejerine, 1984], is characterized by progressive facial, shoulder girdle and pectoral muscle weakness and atrophy. Disease progression may lead to involvement of abdominal and pelvic muscles causing lumbar hyperlordosis and waddling gait. Weakness of anterior leg muscles results in steppage gait.

The FSHD molecular defect does not reside in any protein-coding gene. Instead, a large majority of FSHD patients carry rearrangements occurring in a 3.3 kilobase (kb) tandemly arrayed sequence (D4Z4) located at the 4q subtelomeric region, 4q35 [Lunt et al, 1995a; Upadhyaya et al, 1997; Funakoshi et al, 1998; Sakellariou et al, 2012; Scionti et al, 2012]. The various number of D4Z4 repetitive elements result in polymorphic EcoRI alleles detected by the p13E-11 probe such that the size of each allele is mainly due to the number of D4Z4 copies [van Deutekom et al, 1993]. It was established that normal subjects carry alleles greater than 45 kb (≥ 11 D4Z4 repeats), whereas alleles of 35 kb, corresponding to 8 D4Z4 units, or shorter, are present in the majority of either *de novo* or familial FSHD patients [Lunt et al, 1998; Scionti et al., 2012].

2.1 Detailed molecular and clinical characterization of probands and their relatives carrying the shortest D4Z4 reduced allele (1-3 DRA)

The alleles of extremely short sizes (1-3 D4Z4 repeats) were described to be associated with the most severe form of disease characterized by an early onset and rapid progression of muscle weakness [Jardine et al, 1994; Lunt et al, 1995; Ricci et al, 1999; Ricci et al, 2013]. Infantile FSHD has been subsequently described as a separate entity, defined by the onset of facial weakness by the age of 5 years and shoulder girdle weakness by the age of 10. A number of reports described cases carrying very short D4Z4 alleles with 1-2 repeats characterized by childhood onset, rapid progression of muscle weakness and extramuscular

clinical features [Okinaga et al, 1997; Nakagawa et al, 1997; Miura et al, 1998; Yamanaka et al, 2002; Wang et al, 2012]. However several studies reported differences in clinical expression between subjects carrying shorter alleles, varying from very severe forms of disease and complex phenotypes starting in infancy [Okinaga et al, 1997; Nakagawa et al, 1997; Miura et al, 1998; Yamanaka et al, 2002; Dorobek et al, 2004] to milder form or asymptomatic carriers [Sakellariou et al, 2012]. By revising the literature, we found that the severe cases not all had a childhood onset, or carried a D4Z4 allele of very reduced size, however due to the different design of these studies it is not possible to pool various observations to obtain a complete or more defined picture of clinical features of subjects carrying "very short" D4Z4 allele. Therefore, the goal of our research is to conduct a detailed clinical and molecular characterization of 66 index cases carrying 1-3 DRA alleles from Italian National Registry for FSHD.

2.2 Analysis of D4Z4 methylation status among FSHD patients and healthy controls

D4Z4 contains DNA elements characteristic of heterochromatic regions and various evidences indicate that the D4Z4 repeat array acts as locus control region that governs 4q35 chromatin conformation and expression of nearby genes [Gabellini et al, 2002 ; Zeng et al, 2009; Cabianca et al, 2012]. It is known that chromatin conformation might be influenced by numerous elements, including genetic and environmental factors. It is therefore possible that high variability of clinical expression of FSHD in carriers of DRAs as well as in subjects carrying D4Z4 alleles of normal size might be due to multiple factors influencing chromatin conformation at the D4Z4 locus, and thus effect the expression of proximal genes, or that participate to molecular pathways in which 4q35 genes are involved.

Consistent with this possibility, it has been recently shown that in a group of FSHD2 families, affected subjects carry mutations in SMCHD1 gene, encoding a chromatin modifier, in association with D4Z4 hypomethylation [Lemmers et al, 2012], hence it has been suggested that reduced penetrance in FSHD might be explained by the presence of additional factors either in association with D4Z4 reduction or independently.

On this basis, we consider that the study of epigenetic conformation at the D4Z4 locus in a large set of families might clarify some aspects influencing FSHD manifestation. We propose to investigate the relevance of D4Z4 methylation in the large collection of well clinically and molecularly-characterized FSHD families accrued through the Italian National Registry for FSHD and in healthy controls.

3. Materials and methods

3.1 Subject recruitment

The study has been conducted on FSHD families accrued through the Italian National Registry for FSHD (INRF), established in 2007 by the Italian Clinical Network for FSHD (ICNF) (www.fshd.it) [Lamperti et al, 2010]. The ICNF includes two diagnostic laboratories at the University di Modena and Reggio Emilia and at the Fondazione Santa Lucia in Rome and fourteen clinical centers, belonging to the Italian Association of Myology (www.miologia.org) with long standing expertise in diagnosis and management of neuromuscular disorders. The fourteen clinical centers are distributed on the whole Italian territory, from Northern to Southern regions, including islands. Index cases diagnosed as having FSHD according to clinical diagnostic criteria defined by the European Expert Group on FSHD [Padberg et al, 1991].

Clinical examination was extended to all available relatives of the probands. Each subject was evaluated in a prospective manner by trained neurologists of one of fourteen Italian Centers of clinical network during the time of the study (2011 - 2014). All subjects were clinically examined using the standardized FSHD clinical protocol with validated interrater reliability [Lamperti et al, 2010]. The FSHD clinical protocol has been developed by the ICNF in order to numerically define the clinical severity of the motor impairment, not to be used to diagnose FSHD. The FSHD scale quantifies muscle weakness through the functional evaluation of six muscle groups specifically affected in FSHD, belonging to I) face (score from zero to 2); II) shoulder girdle (score from zero to 3); III) upper limbs (score from zero to 2); IV) distal legs (score from zero to 2); V) pelvic girdle (score from zero to 5); VI) abdominal muscles (score from zero to 1). The FSHD score, which translates disability into a number, ranges from zero, when no objective evidence of muscle functional impairment is present, to 15, when all the muscle groups tested are severely impaired (www.fshd.it) [Lamperti et al, 2010]. DRA carriers who received an FSHD score equal to zero did not show an objective motor impairment and were considered clinically unaffected at time of examination. On the basis of the FSHD score, subjects were classified as affected by mild (FSHD score 1-2), moderate (FSHD score 3-6), or severe (FSHD score 7-15) motor impairment.

Age at onset was estimated on the basis of patient's records. Even though the perception of disease onset may be subjective and could depend on the specific motor skills required in daily activities, we considered that the timing of onset of motor impairment based on patient's

complaints provides a reliable estimate of the degree of functional disability related to disease.

In order to investigate the earliest signs of disease and to rule out pre- or perinatal events as possible causes of delayed achieving of motor milestones, an Anamnestic Infantile Form (AIF) was designed (Appendix 2). All data about: 1) pregnancy, 2) birth, 3) prenatal period and the first months after, and 4) psychomotor and language development were collected in retrospective manner. Items related to each section were scored as normal/altered. Anamnestic reports about neurological examinations in the first year of life, together with clinical and instrumental data in the following years when possible. To obtain a more objective evaluation of the age at onset of the facial weakness, we asked specific questions, such as “Have your relatives ever noticed that you were sleeping with half-closed eyes?”; “Have you ever been able drink with a straw?”; “Have you ever notice difficulty in blowing out candles?”. When subjects did not complained motor impairment, but a mild muscle weakness was observed, we arbitrarily considered the age at examination as the age at onset, according to previous reports [Lunt et al, 1995b].

Recruited healthy volunteers match the demographic origin of patients. Controls were selected as: randomly enrolled healthy volunteers for the genetic study.

The study was approved by the Local Ethics Committees of all participating Institutions. Informed consent, according to the Declaration of Helsinki, was obtained from each subject enrolled in the study.

3.2 Study design

3.2.1 Detailed molecular and clinical characterization of probands and their relatives carrying the shortest D4Z4 reduced allele (1-3 DRA)

We identified 114 index cases carrying DRA with 1-3 repeats and fulfilling the clinical diagnostic criteria defined for FSHD [Padberg et al, 1991]. Out of them, 66 index cases from unrelated families were recruited. Clinical and molecular analysis was extended to all available and willing to participate relatives.

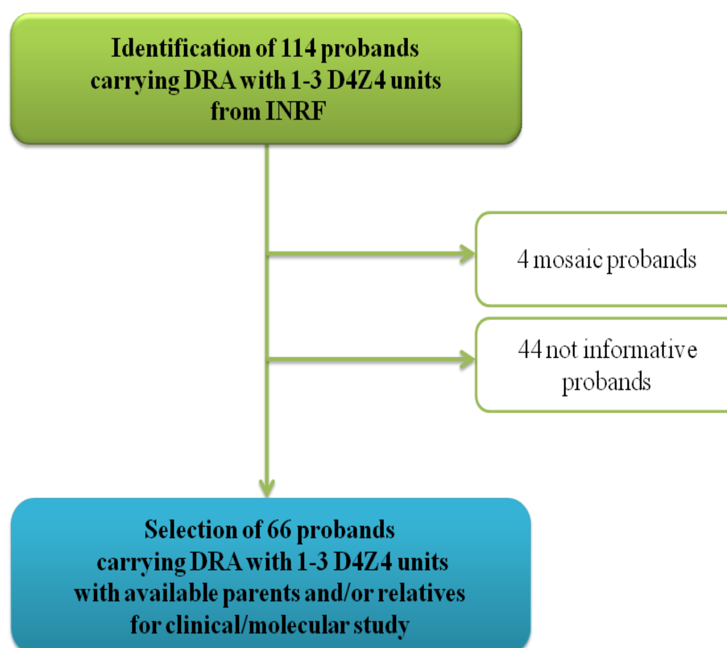

Figure 21: Preliminary selection of probands/families from the Italian National Registry for FSHD (INRF).

Screening for 1-3 DRA was performed on 236 individuals. *De novo* cases were defined as single subjects with neither parent carrying DRA; when the DRA was detected in one of the parents and/or other family members (brothers and sisters for example) the subject was considered as familial. 44 subjects were considered not informative since it was not possible to examine their parents and/or other informative family members.

3.2.2 Analysis of D4Z4 methylation status among FSHD patients and healthy controls

We selected 47 FSHD families from the INRF and 43 healthy controls for the methylation study; 97 subjects from 47 unrelated FSHD families and 43 controls (Figure 22) were analyzed as they displayed a standard allele constitution of 4-type D4Z4 repeat units on chromosomes 4 and 10-type D4Z4 repeat units on chromosome 10 (see Figure 14).

Clinically well-characterized FSHD families were subdivided in two groups: (1) 38 FSHD families where 1-8 DRAs segregate (FSHD1) and (2) 9 FSHD families where 9-11 DRAs segregates (borderline FSHD). Non-penetrant relatives (18 subjects) carrying same DRA as they FSHD family members, were selected in distinct group. Among healthy controls 40 carried normal sized alleles (>11 D4Z4 repeats), while two individuals carry one of the 4q alleles contracted (6 and 8 DRA) and 6 have borderline DRA (9-11 DRA).

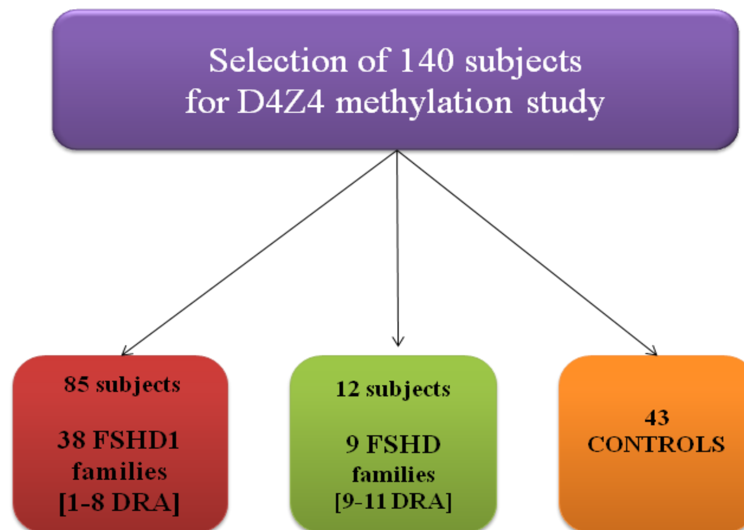

Figure 22: Preliminary selection of families from the Italian National Registry for FSHD (INRF) and healthy controls for methylation study.

For the in-depth methylation analysis were selected 6 FSHD patients, four carrying DRA with ≤ 8 D4Z4 repeats, one carrying borderline DRA (11 D4Z4) and one carrying D4Z4 allele of normal size (12 D4Z4).

3.3 Molecular analysis

3.3.1 Standard molecular test for FSHD

Lymphocytes isolation from whole blood

- 10 ml of fresh blood are collected into a tube containing EDTA as anticoagulant.
- Blood is diluted with an equal volume of 1X PBS.
- 20 ml of the diluted blood are carefully layered over 10 ml of Lympholyte-H in a 50 ml Falcon tube.
- Samples are centrifuged at 800g for 45 minutes at room temperature in a G8 swing-out rotor without brake.
- After centrifugation, it should appear a well-defined lymphocyte layer at the interface. The cells from the interface are removed using a sterile Pasteur pipette, without collecting any of the clear Ficoll layer and transferred into a new centrifuge tube.
- The transferred cells are diluted with 1X PBS up to 30 ml.
- Samples centrifugation at 500g for 5 minutes.
- Cell pellet is washed twice with 10 ml of 1X PBS.

Inclusion of cells in agarose plugs

- Pellet is resuspended gently in 10 ml of 1X PBS.
- Cells are counted with NucleoCounter (Chemometec, Denmark).
- After cell count, cells are spun down and cell pellet is resuspended in PBS to a homogeneous concentration of $2x10^6$ per 40µl. Avoiding formation of air bubbles is recommended.
- An equal volume of 1% Low Melt Agarose in PBS (kept at 50°C) is added to the cell suspension. Cells are mixed gently with a pipette to ensure even dispersion through the agarose and then dispensed immediately into the plug molds (BioRad). Plugs are placed at 4°C in order to allow the agarose plugs to solidify for 20-30 minutes. Each well, in BioRad's disposable plug mold, holds approximately 80 µl of samples and contains $2x10^6$ cells.

DNA isolation in agarose blocks

- Plugs are gently transferred into a 2 ml microfuge tube containing the extraction solution: 1X TE, 1% sarcosyl with EDTA 0,5M (pH 8.0) and Proteinase K (0,5 mg/ml).
- Samples are incubated at 50°C for 48 hours.
- After the Proteinase K treatment, plugs are put on ice for 30 minuts to allow blocks to set.
- After the 30 minutes on ice, the extraction solution is removed and plugs are washed for 5 minutes with distilled sterile water. Washing is repeated for 3 times.
- Plugs are then transferred in a new 2 ml microfuge tube and, after addition of PMSF solution (40µg/ml with TE pH 8.0 – 1ml/plug), are incubated at 50°C for 30 minutes.
- After PMSF treatment, plugs should solidify on ice for 30 minutes, allowing blocks to set.
- PMSF solution is removed and plugs are washed for 5 minutes with distilled sterile water, and washing is repeated for 3 times.
- Plugs are stored in a 0,5M EDTA, pH 8.0 (at least 1 ml per plug), at 4°C.

Restriction endonuclease digestion of DNA in agarose blocks

- EDTA solution is discarded and plugs are washed for 3 times, for 5 minutes with distilled sterile water.
- ½ of plug (corresponding to $1,0x10^6$ cells) is used for each DNA digestion.

- Each plug slice is transferred in a 2 ml microfuge tube and equilibration mix is prepared (appropriate restriction buffer 1X and water). Samples are incubated for at least 1 hour at 37°C.
- After equilibration, the restriction buffer is discarded and digestion mix is prepared (appropriate restriction buffer 1X, restriction enzyme and water):
- EcoRI: 6U/µl (final volume: 200 µl), overnight at 37°C. 1 plug is digested.
- BlnI: 0,25U/µl (final volume: 200 µl), 4 hours at 37°C. 1/2 plug is digested.
- After digestion, plugs are placed on ice for 30 minutes.

Pulsed Field Gel Electrophoresis (PFGE)

- DNA is separated by PFGE on a 1% agarose gel (Megabase agarose, Biorad) . The electrophoresis is performed in 0,5X TBE at 14°C.
- To detect EcoRI, interval is 10-200 kb, switch times increases from 0,74 sec to 17,33 sec at the end of each cycle and it runs for 20 hours and 18 minutes.
- After electrophoresis, the gel is stained with ethidium bromide and blotted to a Nylon$^+$ membrane (ZetaProbe, Biorad).
- Sizes of each chromosome are estimated according to a HMW marker (8-48 kb, Roche).

Isolation of High Molecular weight DNA from frozen blood samples

- Frozen blood in tubes containing EDTA as anticoagulant is thawed.
- To disrupt red blood cells 5 volumes of sterile cold TE 20:5 are added to each blood sample and let it set on ice for 1 hour.
- Samples are centrifuged at 3000 rpm for 10 minutes at 4°C. Supernatant is decanted. Cell pellet is washed with 15 ml of sterile cold TE 20:5 and spun again as previously described, until the buffy coat becomes white and all red cells have been removed.
- Cell pellet is resuspended in lysis solution: 1/2 blood volume of TE 1X, Sarcosyl 1% and Proteinase K (final concentration 100 µg/ml).
- Samples are incubated over night at 37°C or 2 hours at 55°C.
- After the proteinase K treatment, the solution is transferred into a 15 ml tube. An equal volume of phenol equilibrated with 0,1M Tris-HCl (pH 8.0) is added and mixed by gentle agitation for 15 minutes, at room temperature. To separate the aqueous phase, the sample is centrifuged for 15 minutes at 3000 rpm.

- The 90% of the upper viscous aqueous phase is transferred to a clean centrifuge tube, carefully avoiding protein localized at the aqueous:phenol interface.
- An equal volume of phenol:chloroform:isoamilic alcohol (24:25:1) is added to the aqueous phase. Samples are then mixed for 15 minutes, centrifuged for 15 minutes and aqueous phase is then transferred to a clean tube, as described above.
- An equal volume of chloroform:isoamilic alcohol 25:1 is added. Samples are then mixed for 10 minutes, centrifuged for 10 minutes and the aqueous phase is transferred to a clean tube.
- DNA is precipitated with 5M NaCl at final concentration of 200 mM and 2 volumes of cold ethanol absolute.
- Dried DNA is dissolved in TE 10:1

Digestion of DNA and linear gel electrophoresis (LGE)

- 14 µg of DNA are digested with EcoRI (50U) overnight. Digestion is checked on a 1% agarose gel.
- DNA is precipitated with 5M NaCl at a 200 mM final concentration and 2-3 volume of cold absolute ethanol.
- DNA is washed with ethanol 70% and dried DNA is dissolved in BlnI buffer.
- Half of the DNA is transferred in a clean 1,5 ml microfuge tube and 20U of BlnI enzyme are added to the samples.
- DNA samples are incubated at 37°C for 2 hours.
- Seven micrograms of genomic DNA, digested with EcoRI, double digested with EcoRI/BlnI is separated on a 0,4% agarose gel. The electrophoresis is performed in 1X TAE buffer at 4°C for 48 hours.
- After electrophoresis, the gel is stained with ethidium bromide, photographed under UV light and blotted to a Nylon membrane (ZetaProbe, Biorad).
- Sizes of each chromosome are estimated according to a HMW marker (8-48 kb, Roche) and 5 kb marker (5-50 kb, Roche).

Southern Blotting

- After staining of the gel with ethidium bromide solution, DNA is nicked by agitating the gel in a tray containing a fair amount of 0,25 M HCl for 10 minutes.
- After 10 minutes, HCl solution is removed through abundance of distilled water.

- Gel is then soaked in 0,5M NaOH solution for 20 minutes. DNA is transferred onto ZetaProbe GT Nylon membrane (BioRad) using 1 liter of 0,5M NaOH, through capillary transfer.
- Capillary transfer is set up as a follow, from bottom to top:
  - Pyrex glass dish, containing 1 liter of NaOH solution
  - Glass plate
  - One sheet of blotting paper (3mm) as a wick
  - Agarose gel (top side down)
  - ZetaProbe GT membrane cut to the size of the gel and prewetted with NaOH solution
  - Four sheets of blotting paper
  - A stack of paper towels 10 cm high
  - Glass plate
  - 1 kg of weight
- DNA is transferred for 18 hours.
- Paper towel and blotting papers are carefully removed.
- Membrane is rinsed in 2X SSC solution.
- Membrane is then dried by blotting onto 3mm paper, at 80°C, for 40 minutes and it can be used for hybridization.

Probe labeling and hybridization

- Membrane is incubated in pre-warmed "Myracle hybridization solution" (Stratagene) for 1 hour at 68°C in a roller bottle.
- Probe p13E-11 (D4F104S1) is labeled with ^{32}P-dATP using High Prime DNA labeling Kit (Roche).
- Radio-active labeled probe is denatured by heating in a boiling water bath for 5 minutes and chilled quickly in ice.
- Probe (1-2X10^6 cpm/ml) is added to pre-warmed "Myracle hybridization solution".
- Membrane with probe is incubated overnight at 68°C.
- Hybridization solution is removed and membrane is washed quickly with solution 1 (2X SSC with 0,1% SDS) and then incubated for 20 minutes, at room temperature, with the same solution.

- Solution 1 is removed and the rinse solution is replaced with 50 ml of pre-warmed solution 2 (0,5X SSC with 0,1% SDS). Membrane is incubated for 30 minutes at 65°C.
- Solution 2 is removed and the rinse solution is replaced with 100 ml of pre-warmed solution 3 (0,1% SDS). Membrane is incubated for 30 minutes at 68°C.
- Membrane is then removed from the bottle and covered with plastic seal wrap.
- Membrane is placed into a cassette and exposed to X-ray film (Kodak X-Omat LS) at -80°C with an intensifying screen to obtain an autoradiographic image.

3.3.2 BsaAI and FseI methylation assay

For the methylation analysis of the proximal D4Z4 repeats units on chromosome 4q, which has been shown to be representative of the entire D4Z4 array [van Overveld et al, 2003] we applied method illustrated in Figure 13.

- 10 µg of DNA extracted from isolated lymphocytes are digested for 5 hours with 100U of BglII (NEB). Digestion is checked on 1% agarose gel.
- DNA is precipitated with 3M NaOAc in a 300 mM final concentration and 2.5 volume of cold absolute ethanol.
- DNA is washed with ethanol 70% and dried DNA is dissolved with CutSmart buffer (NEB) at 37ºC for 1 hour.
- Dissolved DNA is subsequently digested with 40U of AvrII (NEB) at 37ºC for 2 hours.
- Mix is divided in two (5 µg each) separate 1,5 ml microfuge tube.
- 5 µg are digested with 20U of methylation sensitive restriction enzymes BsaAI (NEB), and other 5 µg are digested with 14U (NEB) of methylation sensitive restriction enzymes FseI (NEB).
- 5 µg DNA digested with BglII/AvrII/BsaAI and 5 µg DNA digested with BglII/AvrII/FseI are separated on 0,8% agarose gel. The electrophoresis is performed in 1X TAE buffer for 24 hours.
- After electrophoresis, the gel is stained with ethidium bromide and blotted to a Nylon membrane (ZetaProbe, Biorad).
- DNA is hybridized with the radioactive-labeled probe p13E-11.
- Digested fragments are detected by autoradiography with Typhoon Trio system (GE Healthcare) and the signal intensities are quantified with.
- The intensity of the bands was acquired with Phosphor imager and percentage of methylation was estimated for the methylated band in comparison to unmethylated band

and considered as a level of methylation on the examined BsaAI/FseI site (Figure 23).

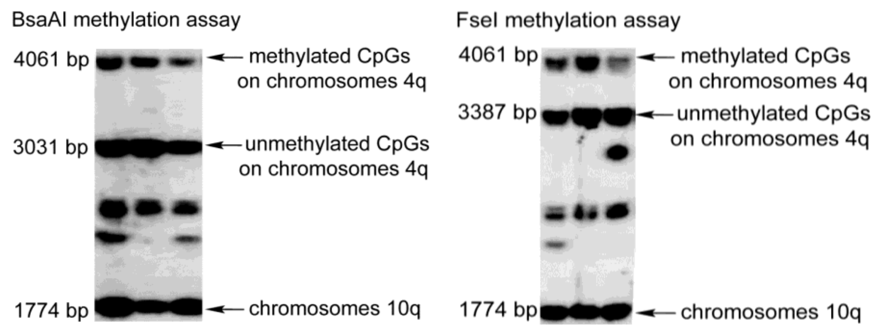

Figure 23: Autoradiography of blotted DNA, digested with methylation-sensitive restriction enzymes, BsaAI and FseI and hybridized by with probe p13E-11.

3.3.3 In-depth methylation assay: Bisulfite D4Z4 conversion and sequencing

- Linear gel electrophoresis of EcoRI-digested DNA is used to isolate D4Z4-reduced allele (thanks to molecular analysis previously done and HMW used, each D4Z4 reduced allele originated from chromosome 4 is identified).
- DNA is extracted from agarose blocks and purified with Gel Extraction Kit (Qiagen).
- EpiTect Bisulfite Kit (Qiagen) is used for complete bisulfite conversion of previously isolated DNA. Conversion and cleanup is performed following procedure described in EpiTect Bisulfite Handbook.
- 2 µl of bisulfite converted DNA is used in each of four separate PCR reactions amplifying D4Z4-region 1 (DR1), D4Z4-region 2 (DR2), D4Z4-region 3 (DR3) and DBE-region (DBE-R).
- DR1, DR2 and DR3 fragments are PCR amplified with specific oligonucleotides previously reported by Hartweck [Hartweck et al, 2013]:
    1. DR1-F:5'-GAAGGTAGGGAGGAAAAG-3'
       DR1-R: 5'-ACTCAACCTAAAAATATACAATCT-3';
    2. DR2-F: 5'-AAATATGTAGGGAAGGGTGTAAGTT-3'
       DR2-R: 5'-CTTAAATATACCAAACCCTCTCTCC-3';
    3. DR3-F: 5'-GTAGAGGGGATTTTTTAATTTGTTT-3'
       DR3R: 5'-CAAACACCCCTTAACCCTAC-3'.
- DBE-R is amplified with nested PCR, using newly designed oligonucleotides:
    1. DBE-R-F: 5'-GGGTTGTTTTAGGGGGGTTTAT-3',
    2. DBE-R-R1: 5'-TAACCTCTCATTCTAAAACC-3',
    3. DBE-R-R2: 5'-CAAATCTAAACCCTAAACTC-3'.

For each PCR 10 µl of reaction mix is prepared with: 10 µM primers (forward and reverse) in a final concentration of 0,125 µM; 2,5 mM dNTPs in a final concentration of 0,2 mM; 5X buffer (GoTaq Flexi) in a final concentration 1X; 25mM $MgCl_2$ in a final concentration of 1,5 mM and 0,5U of GoTaq DNA polymerase. PCR program was as follows:

- 94 ºC for 3 minutes
- [94 ºC for 20 seconds, 53 ºC (DR1, DBR)/ 60 ºC (DR2, DR3) for 30 seconds,
  72 ºC for 40 seconds] for 35 times
- 72 ºC for 10 minutes.

For amplification of DBE-R, 2 µl of obtained PCR product is further processed in the same conditions, using the same forward primer (DBE-R-F) and more internal reverse primer (DBE-R-R1) and amplified in the second PCR round with identical program.

- All PCR products are run on 2% agarose gel and specific molecular size-bands are cut from gel in agarose blocks.
- Specific bisulfite converted DNA fragments are extracted from agarose blocks, purified with Gel Extraction Kit (Qiagen) and dissolved in 20 µl of H_20.
- 2 µl of purified DR1, DR2, DR3 and DBR-DNA bisulfite converted fragments are subcloned into 4-TOPO vector (Invitorgen).
- 4 µl of cloning mix is used for transfection of high-competent E. Colli on selective LB broth medium with ampiciline, overnight at 37ºC.
- At least 10 single colonies are randomly picked and incolutaed in 3 ml LB with ampicillin, overnight (14-16 hours) at 37ºC and 220 rpm.
- Plasmids are isolated and purified with mini-prep kit (Quck lyse mini-prep, Qiagen).
- 500 ng of each isolated plasmide is digested with 10U EcoRI (EuroClone) for 1 hour at 37ºC.
- Digested plasmids are run on 1.5% agarose gel. Described EcoRI digestion enables detection of plasmids with correct molecular-size insert. Plasmids with insert of correct molecular size are selected for sequencing with T3 or T7 universal primers and 3.1 BigDye kit in the standard conditions.
- All sequences for each specific fragment are aligned and number of cytosines are calculated.
- Percentage of methylation is calculated as a portion of number of cytosines in bisulfite converted sequences, of total number of cytosines from CpGs in original non-converted DNA fragment.

3.3.4 Statistical analysis

For the cohorts of probands and family members, Kaplan-Meier survival analysis (Kaplan and Meier, 1958) as used to estimate the age-specific cumulative risk of reported age at onset, muscle impairment incidence and loss of independent walking between, with the corresponding 95% confidence interval (CI). For each individual time from birth to the earliest age at onset of motor impairment was estimated.

The Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (r_S) measures the strength of association between two ranked variables.

Student's t test, statistical test involving confidence limits for the random variable *t* of a t distribution , is used especially in testing hypotheses about means of normal distributions when the standard deviations are unknown

The Kruskal-Wallis test is most commonly used when there is one attribute variable and one measurement variable, and the measurement variable does not meet the assumptions of an anova (normality and homoscedasticity). It is the non-parametric analogue of a single-classification anova. A single-classification anova may yield inaccurate estimates of the p-value when the data are very far from normally distributed, or when the variances of the different groups are very different from each other. The Kruskal-Wallis test does not make assumptions about normality and homoscedasticity. Like most non-parametric tests, it is performed on ranked data.

4. Results

4.1 Detailed molecular and clinical characterization of probands carrying the shortest D4Z4 reduced allele (1-3 DRA) and their relatives

Several studies suggest that carriers of 1-3 DRA develop a severe form of disease. In particular, in FSHD the age at onset has been considered the clinical feature discriminating different phenotypic entities [Brouwer et al, 1994] and thus has been used as the prognostic parameter for phenotype severity [Lunt et al, 1995b]. Consistently, previously we found that FSHD onset occurs earlier in families in which 1-3 DRA segregate [Ricci et al, 2013]. Therefore, we extended our analysis to all probands carrying 1-3 DRA, subdivided in two groups, *de novo* and familial. *De novo* cases were defined as single subjects with neither parent carrying DRA; when the DRA was detected in one of the parents and/or other family members (brothers and sisters for example) the subject was considered as familial.

In 66 unrelated index cases carrying 1-3 DRA we extended molecular characterization to parents and/or other relatives. To this purpose we analyzed 236 subjects clinically and molecularly. We found that 26 probands were familial (39.4%) index cases and 40 probands were *de novo* (60.6%)(Figure 24).

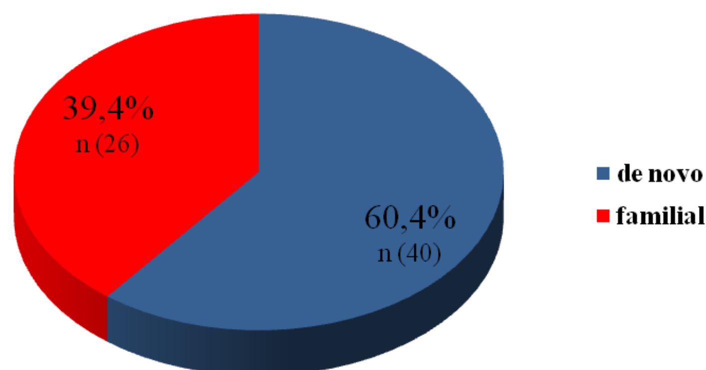

Figure 24: Distribution of *de novo* and familial index cases, carriers of 1-3 DRA.

Interestingly, the mean age at onset observed among *de novo* probands was 8.1 yrs; whereas it was 13.1 yrs among familial index cases. Thus muscle weakness appears significantly earlier in *de novo* cases than in familial (Long-rank p test value=0.020). The difference in the onset of disease between two groups was calculated with non-parametric Kruskal-Wallis test.

1-3 DRA index cases	N subjects	Age at onset 0-10 yrs N subjects	%	> 10 yrs N subjects	%
de novo	40	26	65	14	35
familial	26	10	38.5	16	61.5
tot	66	36	54.5	30	45.5

Table 4: Age at onset among *de novo* and familial index cases, carriers of 1-3 DRA.

Subsequently, we subdivided these two groups of patients on the basis of the age at onset, from 0 to 10 years and over 10 years (Table 4). For the two cohorts of probands, Kaplan-Meier survival analysis (Kaplan and Meier, 1958) was used to estimate the age-specific cumulative motor impairment incidence. For each individual time from birth to the earliest age at onset of motor impairment was estimated.

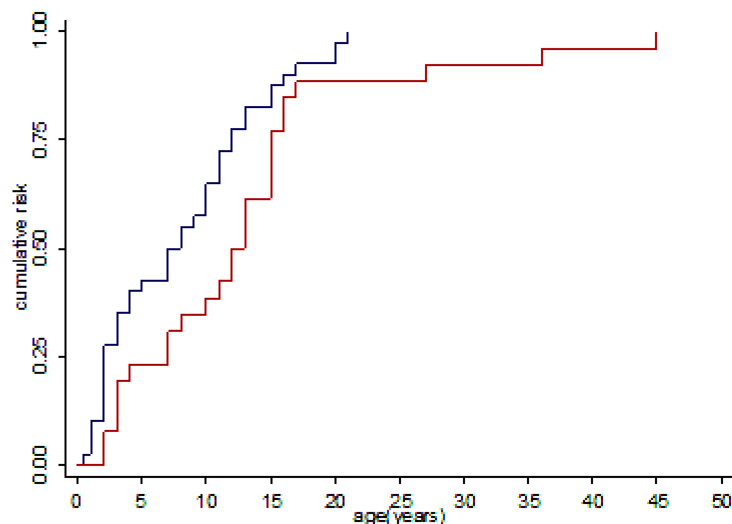

Figure 25: Age-specific cumulative risk of reported age at onset between *de novo* and familial index cases, carriers of 1-3 DRA. Blue line refers to *de novo* index cases carriers of 1-3 DRA, red line refers to familial index cases carriers of 1-3 DRA. Long-rank test p value=0.026.

Figure 25 compares FSHD penetrance between carriers of *de novo* or familial 1-3 DRA calculated with the Kaplan-Meier method. Among subjects carrying a *de novo* 1-3 DRA the risk of developing motor impairment is 65% by age 10, 88% by age 15 and 98% by age 20. Among subjects carrying a familial 1-3 DRA the risk is 38% by age 10, 77% by age 15 and 88% by age 20. Therefore the risk of developing FSHD in childhood, before age 10, is significantly higher in subjects carrying a *de novo* 1-3 DRA.

It has been reported that alleles of extremely short dimension (1-3 D4Z4 repeats) are associated with most severe form of disease [Jardine et al, 1994; Lunt et al, 1995; Ricci et al, 1999; Ricci et al, 2013].

Considering the observed differences in age at onset between *de novo* and familial 1-3 DRA carriers, we tested whether the disease expression is more severe in *de novo* index cases. To this aim, we measured the motor disability by the FSHD score [Lamperti et al, 2010] in *de novo* and familial index cases. Statistical evaluation failed to detect any significant difference in mean FSHD score between the two groups (*de novo* vs familial index cases, 9.7 vs 11.1; Long-rank p test value=0.145).

In eight families we were able to extend molecular analysis to three generations, and we noticed that in six families the molecular defect appeared *de novo* and was transmitted to the infant. Three carriers of *de novo* mutation were recruited as probands. Other three carriers of *de novo* mutation were included in study as relatives of probands. Interestingly, two of them were asymptomatic at the moment of examination (45 and 41 years) and one has a mild form of FSHD (FSHD score 3 at 38 years of age), showing the presence of a wide degree of clinical variability also among 1-3 DRA carriers.

Statistical evaluation failed to detect any significant difference in mean FSHD score adjusted by sex and age, between the two groups (*de novo* vs familial index cases, 9.7 vs 11.1; Long-rank p test value=0.145).

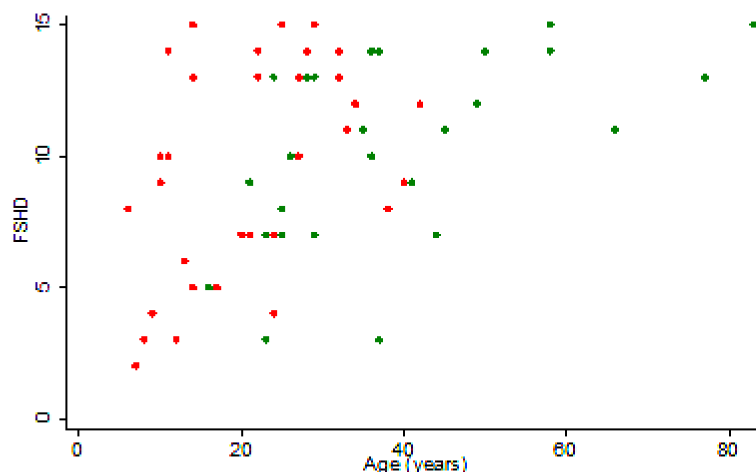

Figure 26: **FSHD score in correlation to the age at examination**. Red spots refer to *de novo* index cases carriers of 1-3 DRA, green spots refer to familial index cases carriers of 1-3 DRA.

By using the Kaplan-Meier method (Figure 26) we further evaluated the relative risk of loss of independent walking, considered as an important feature of motor disability, in carriers of *de novo* rearrangement and familial index cases. Our analysis showed that the cohort of *de*

novo carriers has a higher risk of loss of independent walking versus the familial index cases, even though this difference did not reach a statistically significant value.

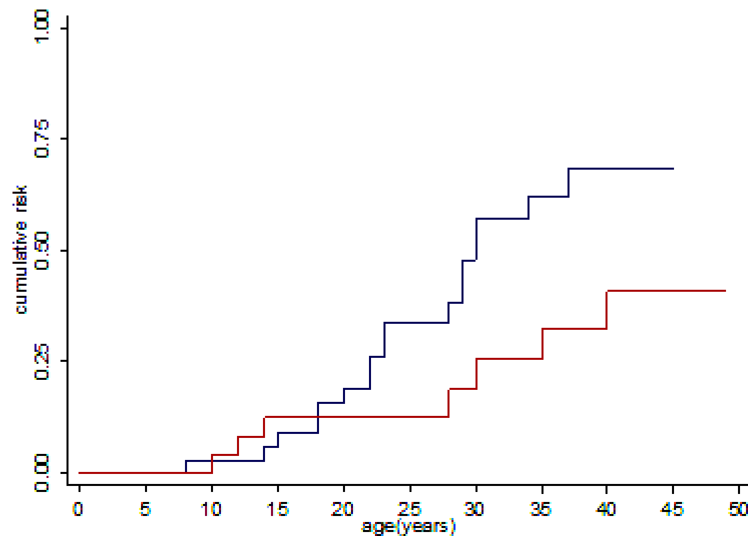

Figure 27: **Age-specific cumulative risk of reported loss of independent walking between *novo* and familial index cases, carriers of 1-3 DRA.** Blue line refers to *de novo* index cases carriers of 1-3 DRA, red line refers to familial index cases carriers of 1-3 DRA. Long-rank test p value=0.062.

In familial cases analysis was also extended to 42 relatives carrying a DRA. We compared the age of the disease onset detected in the group of probands with that recorded in the group of relatives. This comparison displayed that affected relatives present a later onset of FSHD than the probands (relatives versus probands, 17.6 yrs vs 13.1 yrs. Long-rank p test value=0.019). We also compared the degree of motor impairment, recorded as FSHD score, detected in the two groups. The mean FSHD score received by relatives was significantly lower than that recorded in probands (6.1 vs 10.5, Long rank test p value<0.0001). Four relatives (6.1%), respectively aged 33, 42, 47, 50, did not present any muscle weakness.

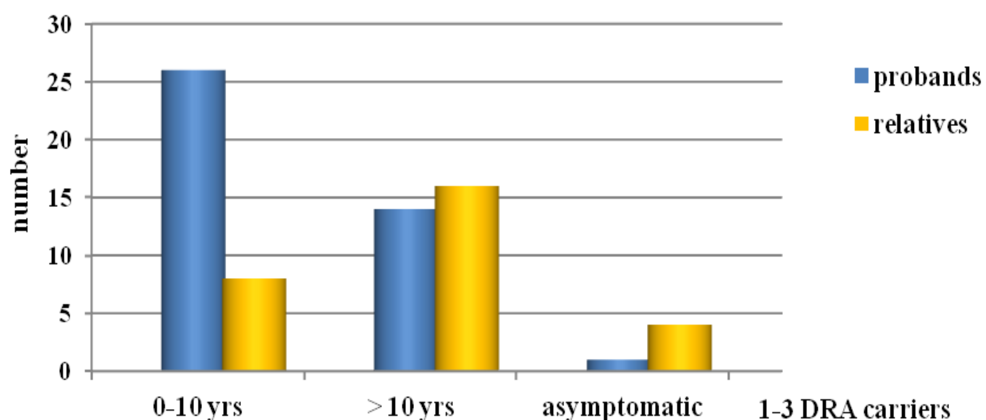

Figure 28: The distribution of probands and their relatives by the reported age at onset.

To systematically obtain information about the perinatal period and the appearance of the first signs and/or symptoms of disease we designed the Infantile Anamnestic Questionnaire (Appendix 2) and gathered anamnestic data about pregnancy, delivery and birth.

Retrospective data were collected from 76 carriers of 1-3 DRA (31 *de novo* cases, 13 familial index cases and 20 relatives, 10 sporadic index cases and 2 mosaic cases). Table 5 shows that no significant alteration in pregnancy, delivery and birth were reported. We had no accounts of any floppy infant at birth. Out of 76 examined subjects patients, psychomotor development milestones were reached regularly in 72 (94%). Seven cases (4 *de novo*, 3 familial) (9.2%) reported reduced sucking, which is defined as a difficulty creating sufficient vacuum without swallowing abnormalities. Notably, all these subjects presented a severe form of disease (FSHD score ranging from 9 to 15 before age 30); six of them suffered loss of independent walking in third decade of life (22-30 yrs). Interestingly, only three of them referred an early age at onset (≤ 10 yrs). Three subjects reported difficulty in pronouncing some phonemes at age of 3-4 years, possibly suggestive of facial muscle weakness.

Infantile anamnestic records of 70 carriers of 1-3 DRA			N of subjects		N of subjects		N of subjects
prenatal period	active fetal movements	normal	**69**	reduced	**2**	NA	**5**
delivery	partum	eutocic	**60**	dystocic	**14**	NA	**2**
	fetal position	cephalic	**58**	podalic	**6**	NA	**12**
birth	weight	normal	**69**	low	**0**	NA	**7**
	revived	no	**70**	yes	**2**	NA	**4**
	clubfoot	"	**70**	"	**1**	NA	**5**
perinatal period	reduced suction	no	**68**	yes	**7**	NA	**1**
	facial nerve palsy's diagnosis	"	**72**	"	**2**	NA	**2**
	Moebius syndrome's diagnosis	"	**72**	"	**1**	NA	**3**
	hip dysplasia	"	**70**	"	**2**	NA	**4**
	facial hypomimia	"	**72**	"	**3**	NA	**1**
	floppy	"	**72**	"	**0**	NA	**4**
psychomotor development	social smile	normal	**68**	altered	**5**	NA	**3**
	walk independently			<15 months	**65**		
				15-18 months	**7**		
				>18 months	**3**		
				NA	**1**		

Table 5: Collected anamnestic records of 70 subjects, carriers of 1-3 DRA, referring to prenatal period, delivery, birth, perinatal period and psychomotor development.

Overall this in-depth genotype-phenotype correlation study on 1-3 DRA carriers reveals that very short D4Z4 alleles do not always associate with a severe phenotype, being relatives less severely affected that probands, including healthy DRA carriers. Moreover our study shows that infantile inset was found in half of the probands carrying 1-3 DRA. Therefore additional factors must influence the disease outcome.

4.2 Analysis of D4Z4 methylation status among FSHD patients and healthy controls

In the recent years it was proposed that an epigenetic mechanism plays a role in the pathogenesis of FSHD. C-5 methylation of cytosine, the most common epigenetic modification of mammalian DNA, is known to be involved in development, X-chromosome inactivation, imprinting, and gene silencing [Robertson and Wolffe, 2000]. CpG methylation can affect occupancy of specific genomic regions since several transcription factors and chromatin-binding proteins, such as CTCF and YY1, are sensitive to it [Hark et al, 2000; Kim et al, 2003]. Each D4Z4 repeat unit harbors two classes of GC-rich sequences, namely the low copy-repeats hhspm3 and LSau. On these bases DNA methylation has been considered a marker of epigenetic changes at the D4Z4 locus. Indeed various methylation assay on D4Z4 were used by different groups to estimate the level of methylation on two different sites in D4Z4 repeat on chromosome 4q35 and revealed D4Z4 hypomethylation in both FSHD1 and FSHD2 patients.

Based on these studies we tested this model in our cohort of Italian patients. We selected 47 FSHD families from the INRF and 43 healthy controls; 97 subjects from 47 unrelated FSHD families and 43 controls were analyzed as they displayed a standard allele constitution of 4-type D4Z4 repeat units on chromosomes 4 and 10-type D4Z4 repeat units on chromosome 10 (see Figure 14).

Clinically well-characterized FSHD families were subdivided in two groups: (1) 38 FSHD families in which a 1-8 DRA segregates (FSHD1) and (2) 9 FSHD families in which 9-11 DRAs segregates (‘’borderline’’ FSHD). The asymptomatic family members of FSHD patients carrying the same DRA were considered as non-penetrant and grouped separately (18 subjects). Among healthy controls 40 carried normal sized alleles (>11 D4Z4 repeats), while two individuals carry one of the 4q alleles contracted (6 and 8 DRA) and 6 have borderline DRA (9-11 DRA).

4.2.1 The correlation of the level of methylation and the total number of D4Z4 repeats on 4q in FSHD patients and healthy controls

Previous studies demonstrated that both FSHD1 and FSHD2 patients presented a hypomethylation on chromosomes 4q, thus we tested if the methylation level at chromosomes 4q correlates with the presence of the disease in our cohort of patients. We firstly tested whether the level of methylation measured on two analyzed CpGs correlates to a total number of D4Z4 on both chromosomes 4q in the FSHD patients and healthy controls. Secondly we tested weather there is a differences in methylation status between the FSHD patients and their non-penetrant relatives, carrying the same DRA.

Figure 29 is representing the mean percentage of methylation on two D4Z4 CpG dinucleotides in correlation to a total number of D4Z4 repeats on both chromosomes 4q in 3 groups: (1) FSHD patients carrying 1-8 DRA; (2) FSHD patients carrying borderline DRA (9-11 DRA) and (3) healthy controls.

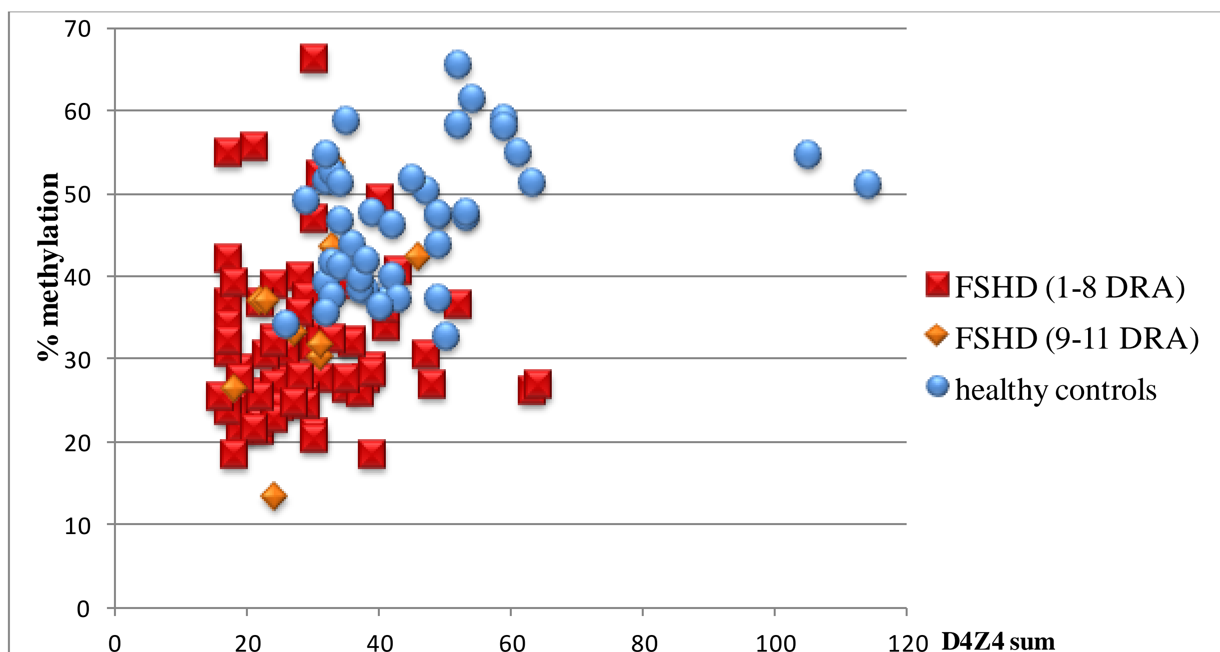

Figure 29: The correlation of measured D4Z4 level of methylation with the total number of D4Z4 repeats on both chromosomes 4q.

A Spearman's correlation was run to determine the relationship between the percentage of methylation and the total number of D4Z4 repeats on both chromosomes 4q in: 1) FSHD patients (1-11 DRA) and 2) healthy controls. Statistical analysis revealed a strong, positive monotonic correlation between the percentage of methylation and the total number of D4Z4 repeats in healthy controls ($r_s = 0.014$, Spearman's test p-value < 0.001), while we failed to

observe any correlation between the number of D4Z4 units and the D4Z4 methylation degree in FSHD patients (r_s =0.073, Spearman's test p-value > 0.001).

4.2.2 The methylation status in FSHD patients and their non-penetrant relatives, carriers of the same DRA

Although the penetrance of FSHD is considered complete by age 20 [Lunt, 1998; Tawil et al, 2010], healthy relatives carrying a D4Z4 reduced allele (DRA) as well as non-concordant monozygotic twins, have been described [Griggs et al, 1995; Tupler et al, 1998; Ricci et al, 1999; Tonini et al, 2004; Ricci et al, 2014].

It was proposed that in order to have FSHD DRA at 4q35 must be combined with a permissive haplotype, namely 4A(159/161/168)DUX4-polyadenylation signal (PAS), which provides the possibility of expressing the most distal copy of the DUX4 gene [Lemmers et al, 2010]. However, our recent observation that 1,3% of healthy individuals from the general population carry reduced D4Z4 allele associated with 4A161PAS haplotype [Scionti et al, 2012b] points at the possibility that in the heterozygous state a D4Z4 reduction might produce a subclinical sensitized condition that requires other epigenetic mechanisms or a contributing factor to cause the development of disease.

This possibility is supported by our most recent observations [Ricci et al, 2013]: 1) among relatives carrying DRA degree of motor impairment is lower than in FSHD probands; 2) in FSHD families the percentage of non-manifesting carriers is higher among second- through fifth-degree relatives in comparison with first-degree family members; 3) in 13% of families in which D4Z4 alleles with 4-8 repeats segregate, the diagnosis of FSHD was reported only in one generation.

Thus we focus our attention on the FSHD families where DRA segregate and were the reduced penetrance in family is present, and we tested possibility that differential clinical expression might lay in a differet degree of D4Z4 methylation between FSHD patients and their non-penetrant relatives carrying the same DRA (1-11 D4Z4 units). We analyzed the methylation level of the proximal D4Z4 repeat at chromosomes 4q in 67 FSHD patients and 18 familial non-penetrant DRA carriers, who received an FSHD score equal to 0, at the time of examination (age range 25-80 years).

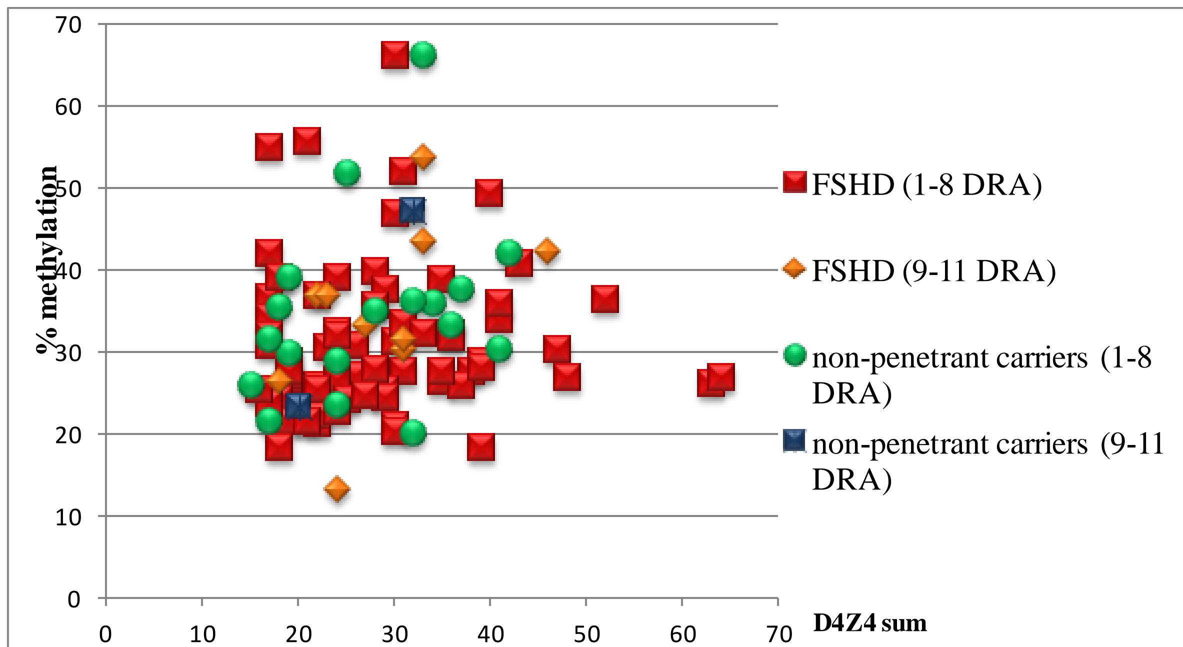

Figure 29: The correlation of measured D4Z4 level of methylation with the total number of D4Z4 repeats on both chromosomes 4q between FSHD patients and their non-penetrant relatives carrying the same DRA.

Comparing the D4Z4 methylation status in non-penetrant DRA carriers with the level of D4Z4 methylation assessed in the FSHD affected relatives, we failed to observe a significant difference of methylation status between the two groups (Student's T test p-value = 0.193).

We also tested the correlation between the methylation status and a total number of D4Z4 repeats at 4q in 43 healthy controls and 18 non-penetrant DRA carriers, relatives of FSHD patients. To this purpose we merged the methylation data obtained from both groups. Interestingly, Spearman test revealed a strong positive monotonic correlation between the percentage of methylation and the total number of D4Z4 repeats on the 4q (r_s = 0.000, Spearman's test p-value < 0.001). This observation seems to indicate that the methylation status detected by using methylation-sensitive restriction enzymes depend on the total number of D4Z4 units carried by each individual. By contrast data obtained from the analysis of FSHD patients show a lack of correlation that might indicate the complexity of the molecular mechanisms underlying FSHD pathogenesis.

4.2.3 The inter-familial phenotypic variability does not correlates with the methylation status in a FSHD family members

The observed phenotypic variability in the FSHD families and the presence of asymptomatic subjects carrying the same DRA as their FSHD relatives, prone us to study the intra-familial methylation status. Here we present FSHD family 197 (Figure 31) from the

INRF, where the DRA with 4 D4Z4 repeats was found in 11 subjects. Six of them received an FSHD score equal to 0, at the time of examination (age range 25-65 years). The 5 family members have the FSHD phenotype.

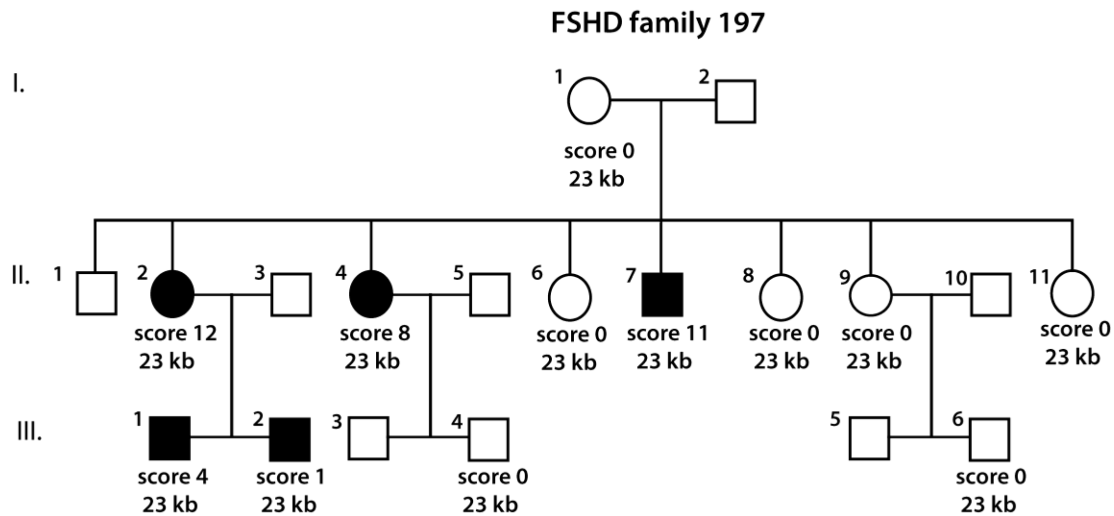

Figure 31: The pedigree of the FSHD family 197, showing with the reduced penetrance of the disease.

We analyzed the methylation status on two distinct methylation-sensitive restriction sites (BsaAI and FseI) and revealed that assessed methylation status does not correlate with the disease (Figure 32 and 33). The level of methylation measured in each family member is presented in the Table 6.

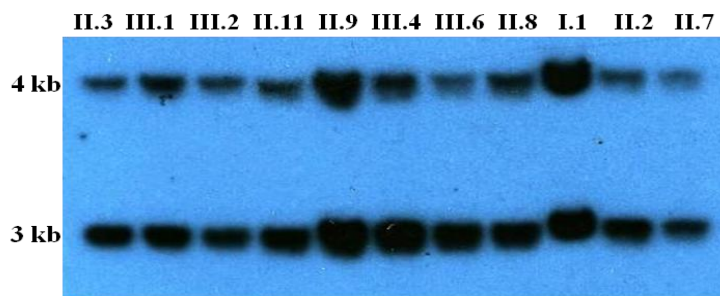

Figure 32: The Southern blot of DNA digested with methylation-sensitive restriction enzyme BsaAI, hybridizied with p13E-11 probe. The upper line corresponds to the methylated CpG sites in D4Z repeat in 4q array and the line bellow corresponds to the unmethylated CpG sites in D4Z4 repeats in 4q array.

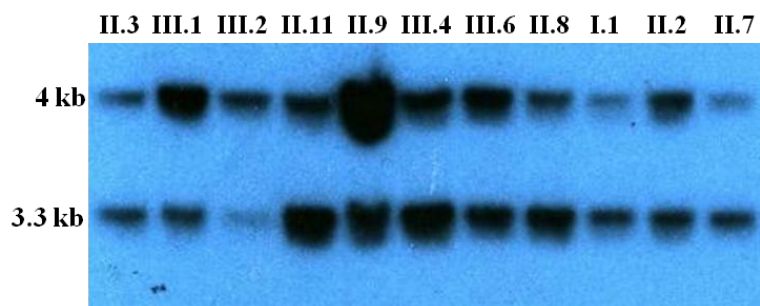

Figure 33: The Southern blot of DNA digested with methylation-sensitive restriction enzyme FseI, hybridizied with p13E-11 probe. The upper line corresponds to the methylated CpG site in D4Z repeats in 4q array and the line bellow corresponds to the unmethylated CpG sites in D4Z4 repeats in 4q array.

family member	FSHD score	BsaAI % of met	FseI % of met	mean % methylation
II.3	8	27	46	36
III.1	4	35	73	54
III.2	1	39	71	55
II.11	0	33	30	31
II.9	0	39	81	60
III.4	0	26	40	33
III.6	0	24	54	39
II.8	0	33	37	35
I.1	0	51	35	43
II.2	12	29	54	42
II.7	11	30	35	32

Table 6: The detailed presentation of methylation status in each member of FSHD family 197 and their FSHD score received at the time of examination.

As shown in Table 6, we found a higher percentage of methylation in subjects III.1, III.2, II.2 affected by FSHD in comparison with their asymptomatic relatives. The degree of methylation assessed in other family members was~ approximately 30% regardless the presence of disease.

4.2.4 The methylation status in the autosomal recessive family

Through the analysis of families collected in the INRF we identified a family in which two sisters show clinical features of severe FSHD, while neither the parents nor the other three siblings display myopathic signs (Figure 34). The family history is negative for the presence of neuromuscular diseases and no other distant relative show signs of muscle weakness. Instead three sibs, II.5, II.8 and II.11 died before age 5 because of beta-thalassemia. The single daughter of one of the FSHD probands is healthy. The family is from a small village in inner Sardinia, Italy. Molecular analysis of the FSHD locus at 4q35 failed to detect alleles with a reduced number of D4Z4 repeats.

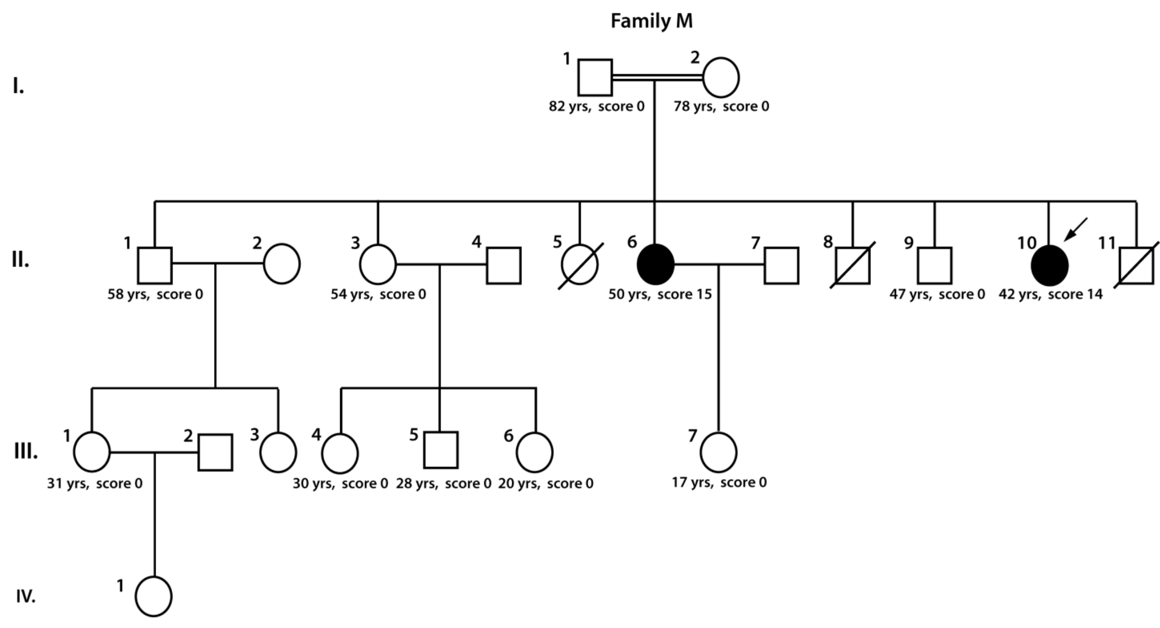

Figure 34: The pedigree of the family M indicating autosomal recessive transmission of disease.

Under the hypothesis that in this family FSHD occurs as an autosomal recessive trait, we decided to identify if the D4Z4 methylation, proposed as the epigenetic factor which contributes to the pathophysiology of FSHD also in the subjects carrying D4Z4 repeats of the normal size [de Greef et al, 2010], we decided to analyze the methylation status of the 4q alleles in the family members.

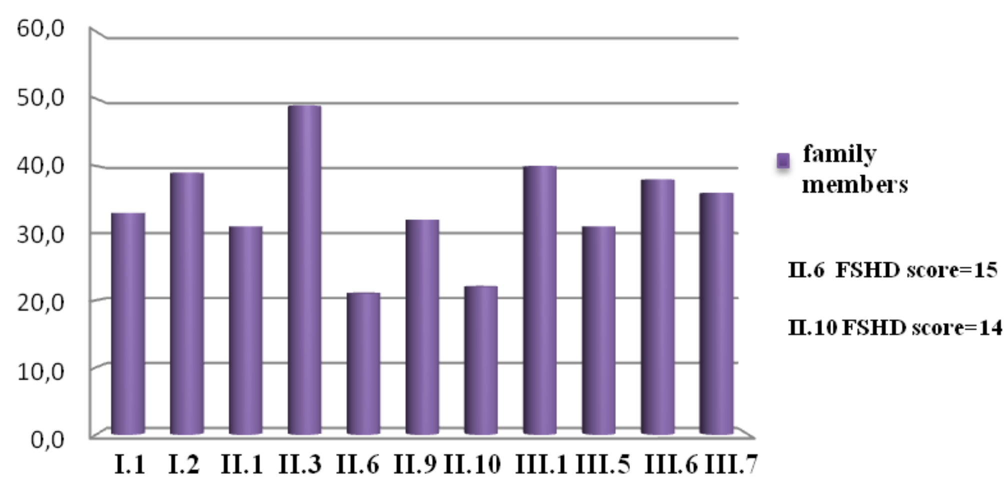

Figure 35: The methylation status on BsaAI methylation-sensitive restriction site in members of autosomal recessive FSHD family M .

The performed D4Z4 methylation assay in this family showed the decreased level of methylation in the affected sisters II.6 and II.10, but interestingly significant demethylation was present also in other family members I.1, II.1 II.9 and III.5 which do not have any sign and/or symptom of the disease (Figure 35).

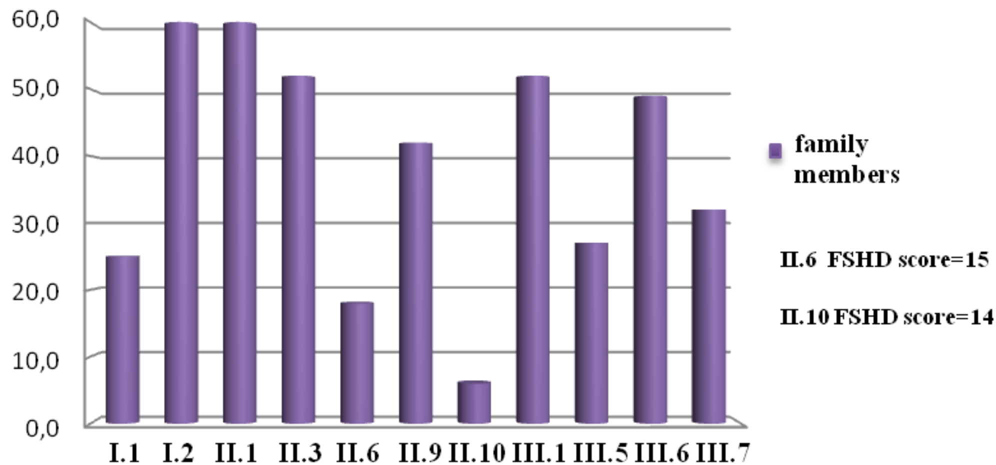

Figure 36: The methylation status on FseI methylation-sensitive restriction site in members of autosomal recessive FSHD family M.

The same analysis we performed on the other methylation-sensitive restriction site (restricted site of FseI) (Figure 36), positioned more distally in proximal D4Z4 repeat and we observed the different result. Affected sisters, II.6 and II.10 had significant demethylation on this site, but the demethylation was detected in the healthy father I.1.

The obtained results from both methylation-sensitive restriction site assays does not attribute a pathogenic significance to D4Z4 hypomethylation since it was found in both patients but also in healthy relatives. In addition, Next-Generation Sequencing (NGS) did not reveal any mutation in SMCHD1, a candidate gene for FSHD2. All these observations highlight the genetic heterogeneity of FSHD and its genetic complexity, either in association with D4Z4 reduction or independently. On these bases we conclude that family M represent a novel nosologic FSHD entity, distinct from FSHD1 and FSHD2.

4.2.5 In-depth methylation analysis

As we did not observe a strict correlation between the level of D4Z4 rmethylation and the presence of the FSHD, we decided to study in more detail methylation pattern of CpG rich D4Z4 repeat array to define the significance of D4Z4 methylation status in FSHD patients.

As described (see Figure 14) the methylation status of the D4Z4 repeat array cannot be assessed in all FSHD patients by using methylation-sensitive restriction enzymes. In addition this assay quantifies methylation only on two CpG dinucleotides. To obtain more detailed information about methylation of D4Z4 sequence Hartweck and colleagues [2013] performed bisulfite sequencing of three domains within the D4Z4 repeat. Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected.

The bisulfite conversion depends on the presence of individual methyl-cytosine, yielding single-nucleotide resolution information about the methylation status of DNA segment of interest.

By bisulfite sequencing Hartweck and coworkers [2013] assessed the methylation status in three distinct regions of D4Z4 in FSHD patients and controls. This approach allowed them to screen globally level of methylation on both chromosome 4q and both chromosome 10q, without providing a specific information on the methylation status of the D4Z4 reduced allele at 4q.

To overcome this limitation, we developed a novel technique, which allows the specific analysis of the methylation status of a single 4q D4Z4 alleles. In order to distinguish between chromosome 4q and chromosome 10q sequences, we performed a comparison between virtually bisulfite-converted D4Z4 sequences belonging to 4q and 10q. This comparison allowed us to identify single nucleotide polymorphism (SNP), located downstream of the D4Z4 binding element (DBE), T at chromosome 4q and G at chromosome 10q (Figure 37).

```
4       ggggtttatCGttatttatgaaggggtggagtttgtttgtttgtgggttttataagggc
10      ggggtttatCGttatttatgaaggggtggagtttgtttgttCGtgggttttataagggc
        ***************************************** *****************

4       ggttggttggttggttggttgttCGggtaggtttttggttgtatttgtCGtagtgtata
10      ggttggttggttggttggttgttCGggtaggtttttggttgtatttgtCGtagtgtata
        ***********************************************************

4       gttCGgttgaggtgtaCGggagttCGtCGgtttttttttgttCGCGttCGttCGtgaaat
10      gttCGgttgaggtgtaCGggagttCGtCGgtttttttttgttCGCGttCGttCGtgaaat
        ************************************************************

4       ttCGgtCGgggtttatCGCGatggttttttCGatattttCGgatagtattttttttCGCGg
10      tgCGgtCGgggtttatCGCGatggttttttCGatattttCGgatagtattttttttCGCGg
        * ***********************************************************
```

Figure 37: Alignment of bisulfite-converted sequences of a D4Z4 fragment on chromosome 4q and chromosome 10q. DBE position is highlighted in red. The polymorphism used to distinguish chromosome 4 from chromosome 10 is circled.

To perform in-depth methylation screening of high number of CpG dinucleotides inside each D4Z4 repeat on chromosome 4q we analyzed four different regions encompassing total~ 1.2 kb DNA sequence on D4Z4 and 113 CpG dinucleotides (see Appendix 3). Bisulfite converted DNA extracted from a band corresponding to a single 4q allele (described in Material and Methods, chapter 3.3.3) was amplified in 4 different DNA fragments DR1, DBE-R, DR2 and DR3. Amplified fragments were cloned and single plasmids were sequenced. Percentage of methylation was estimated as the number of cytosines from a total number of CpGs in non-bisulfite converted DNA sequence. DBE-R region containing a SNP

(T at chromosome 4 and G at chromosome 10, see Figure 37) permitted us to verify that analyzed sequences originate from 4q.

Our analysis revealed a reduced level of CpG methylation in regions DR1 and DBE-R in all 6 analyzed 4q arrays (mean level of methylation, 11.5% and 8.4% respectively); D4Z4 region 2 (DR2) displayed a variable level of CpG methylation in the samples analyzed (13.5%-52%, mean 31.5%), while DR3 was methylated (57.8%) in the majority of D4Z4 arrays 4q-alleles as described in Figures 38-39 and Table 7.

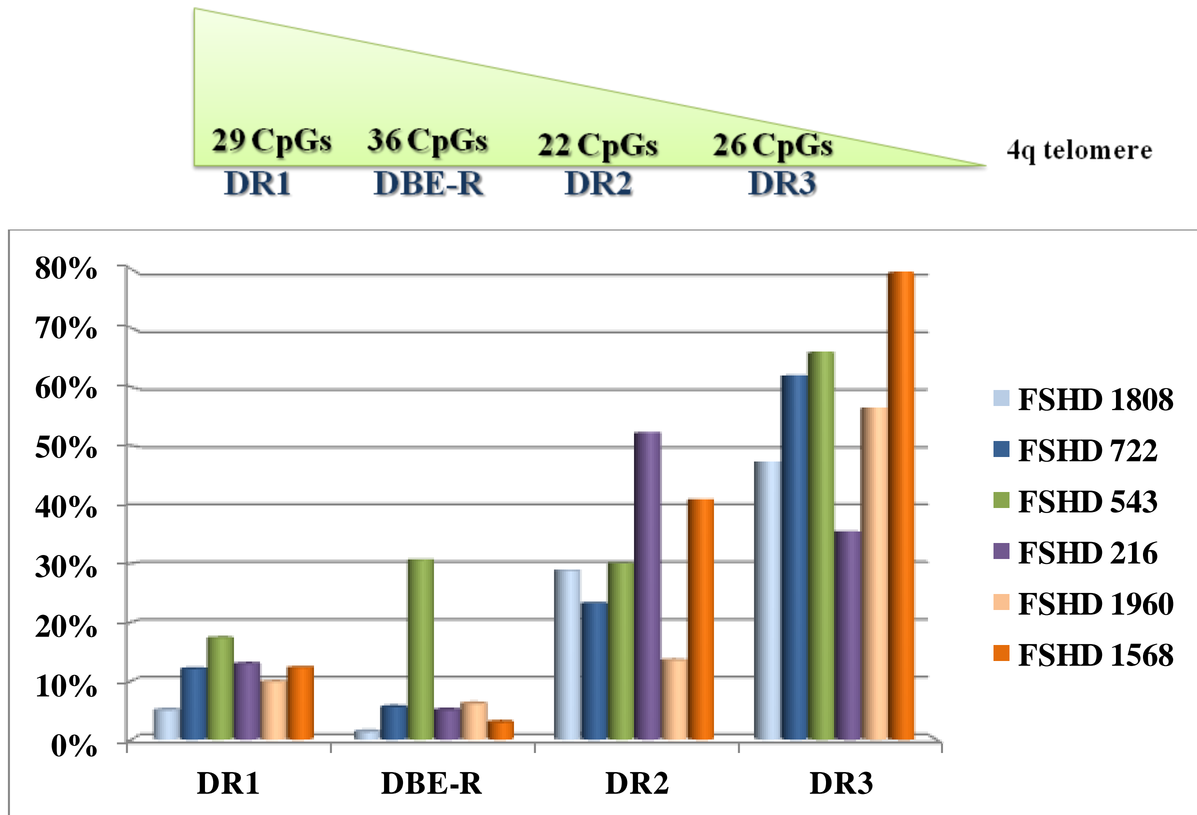

Figure 38: The mean percentage of methylation measured in 4 distinct domains of D4Z4 repeat on chromosome 4q.

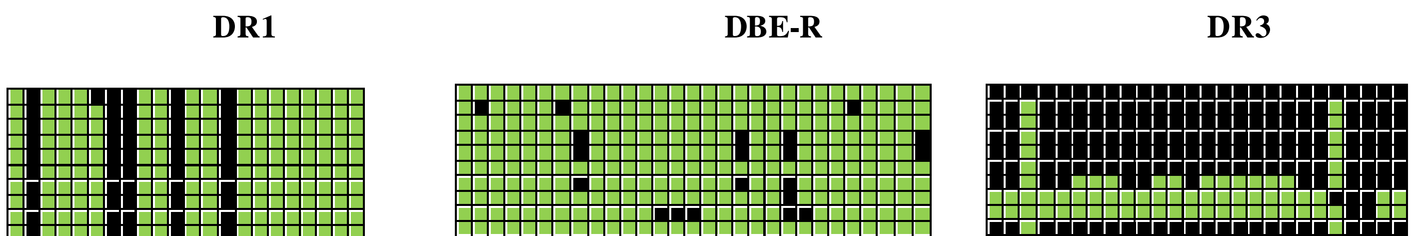

Figure 39: The example of methylation status of each CpG dinucleotide for DR1 (12%), DBE-R (6%) and DR3 (62.3%) in patient FSHD 722.

ID	FSHD score	D4Z4 repeats	DR1 % met	DBE-R % met	DR2 %met	DR3 % met
FSHD 1808	9	3	5	1.3	28.8	47.3
FSHD 722	10	4	12	6	23.2	62.3
FSHD 543	12	5	17.3	30.6	30	66
FSHD 216	7	8	12.9	5	52.3	35.4
FSHD 1960	10	11	9.8	6.1	13.5	56.5
FSHD 1568	7	12	12.2	2.9	40.9	79.6

Table 7: Presentation of FSHD score, size of D4Z4 allele and the percentages of measured methylation level in DR1, DR2, DBE-R and DR3, for each analyzed patient.

The D4Z4 region 1 (DR1) and D4Z4 binding element-region (DBE-R) were selected for the in-depth CpG methylation screening because of the presence of the binding sites of the transcription modifiers CCCTC-binding factor (CCTF) and Ying-Yang 1(YY1).

DR1 is juxtaposed to the binding site of CTCF [Ottaviani et al, 2009], a known insulator protein [Bell et al, 1999]. CTCF functions as the enhancer blocking boundary [Hark et al, 2000]. The binding of CTCF is methylation sensitive [Bell and Felsenfeld 2000; Hark et al, 2000] and its enharcer blocker function is hampered by the methylation of its binding site.
DBE is involved in the recruitment of the D4Z4 repressor complex (DRC) as the putative binding site of the YY1, a PcG protein, was found to be positioned inside of the DBE region [Gabellini et al, 2002]. YY1 binding site encompasses the CpG (CGCCATnTT) and was reported to be inaccessible to YY1 when this CpG dinucleotide is methylated [Do Kim et al, 2006; Sekimata et al, 2011]. Our analysis revealed that in all the DBE-R sequences from the analyzed 4q alleles this CpG was unmethylated in all analyzed sequences.

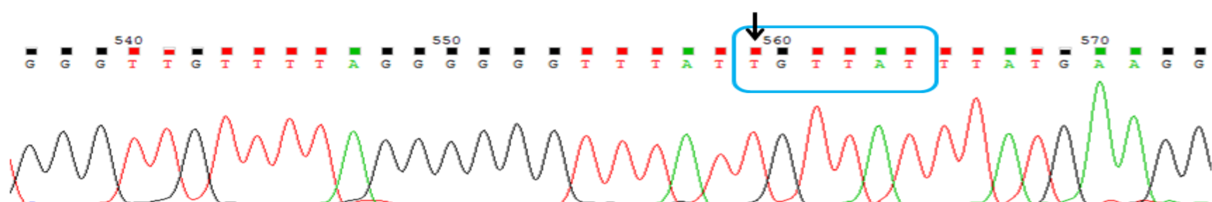

Figure 40: Methylated CpG of the putative binding site in the DBE region.

D4Z4 region 2 (DR2) was selected for the CpG screening because of the presence of the BsaAI methylation-sensitive restriction site.

ID	DR2 %met	BsaAI % met
FSHD 1808	28.8	18.1
FSHD 722	23.2	23.3
FSHD 543	30	25.3
FSHD 216	52.3	51.9
FSHD 1960	13.5	42.2
FSHD 1568	40.9	47.8

Table 8: The level of methylation obtained by bisulfate sequencing of DBE-R and by use ofmethylation-sensitive restriction enzyme BsaAI. Methylation-sensitive restriction site inside the DBE-R domain allows to compare the degree of methylation obtained by two different methylation tests.

As shown in table 8, comparison of the methylation data obtained by two different D4Z4 methylation test allowed us to estimate the reliability of the methylation-sensitive site assay. We can conclude that assessed degree of methylation with both assays is not always comparable, considering the lack of correspondence of the methylation data attained from the patients FSHD 1808 and FSHD 1960.

D4Z4 region 3 (DR3) is domain positioned inside the ORF of DUX4, thus it was from a specific importance to screen the methylation status of 36 CpGs present in this region. We found this region methylated in the majority of 4q arrays. This observation might have implications on understanding expression of DUX4.

5. Discussion

FSHD has been classified as an autosomal dominant myopathy associated to rearrangements occurring in a 3.3 kb tandemly repeated sequence (D4Z4) located at the 4q subtelomeric region [Wijmenga et al, 1992]. It was established that normal subjects carry p13E-11 EcoRI alleles greater than 45 kb (≥ 11 D4Z4 repeats) originating from chromosome 4q, whereas alleles of 35 kb, corresponding to 8 D4Z4 units, or shorter, are present in the majority of *de novo* and familial FSHD patients [Upadhayaya et al, 1996]. Age-dependent penetrance based on the presence of the characteristic clinical signs was estimated >95% by age 20, without inter-familial differences [Lunt et al, 1989]. Therefore, in the last 20 years, for the diagnosis of FSHD has been used molecular test considered highly sensitive and specific for the disease [Tawil et al, 2010].

Genotype-phenotype correlation studies showed a rough and inverse correlation between the number of D4Z4 repeats and the severity of FSHD. Alleles with 1-3 D4Z4 repeats (EcoRI fragment <20 kb) are generally associated with a severe form of disease that presents in childhood, 4-7 D4Z4 repeats (EcoRI fragment 21-34kb) with the classical form of FSHD, and 8-10 D4Z4 repeats (EcoRI fragment 35-45 kb) with a milder disease [Lunt et al, 1995a; Tawil et al, 1996; Ricci et al, 1999]. In addition, EcoRI alleles between 38-45 kb in size (9-11 D4Z4 repeats) have been detected in healthy subjects and individuals affected by FSHD or other myopathies [Butz et al, 2003; Vitelli et al, 1999]. Nevertheless, since the advent of molecular diagnosis for FSHD, several studies have reported FSHD families with subjects carrying D4Z4 reduce alleles and no signs of the disease, defined as non-penetrant carriers [Zatz et al, 1998; Ricci et al, 1999; Tonini et al, 2004; Scionti et al, 2012a, Sakellariou et al, 2012]. Moreover, a growing number of evidences have emerged to complicate the evaluation of patients, reporting a wide and unexpected variability of FSHD clinical outcomes, also among subjects carrying the same D4Z4 allele, even within the same family [Fitzsimons, 1999; Galluzzi et al., 1999; Felice et al., 2000]. For some non-penetrant cases with the D4Z4 reduced allele, it has been proposed that the reduction of D4Z4 repeats on chromosome 4q35 could be pathogenic only in certain chromosomal backgrounds, as "permissive" specific haplotypes [Lemmers et al, 2002; Lemmers et al, 2004b; Lemmers et al, 2010), namely (1) reduction of D4Z4 elements, (2) presence of the 4A(159/161/168) haplotype, and of (3) a single nucleotide polymorphism which provides a polyadenilation signal (PAS) for the DUX4 transcript [Lemmers et al, 2010]. However, our most recent studies establish that, although the majority of FSHD patients (70%) carry D4Z4 alleles with 4-8 units, this size range is also present in 3% of healthy subjects from the general population.

Additionally, there is little predictive value of the 4qA161PAS haplotype in the absence of family history since our recent study found "permissive" haplotype in 1.3% of healthy subjects. Thus the FSHD molecular signature has the frequency of a common polymorphism rather than a rare mutation. All these observations strengthened the notion that is important to re-evaluate clinical phenotype and systematically analyze a large cohort of FSHD families, in order to establish the predictive value of DNA test and potentially identify other modifying factors as prognostic tools in clinical practice and genetic counseling. Our studies fall in this contest.

5.1 Are the 4q alleles with 1-3 D4Z4 repeats always predictive for a severe form of FSHD?

Our most recent genotype-phenotype study [Ricci et al, 2013] performed on FSHD families showed that FSHD display almost full penetrance in families in which 1-3 DRA segregates, with a more severe phenotype and reduced clinical variability, in comparison with carriers of DRA with 4-8 units. However, all *de novo* index cases were excluded from the study because of its design. Therefore we considered that analysis of all index cases carrying of DRA with 1-3 repeats might offer more precise information about this group of patients. In particular, we aimed at verifying whether this subjects present FSHD with infantile onset and whether detection of a 1-3 DRA is always predictive of a severe phenotype.

First, our analysis showed that the majority of carriers of 1-3 DRA (60.6%) are *de novo*. Notably, this percentage is higher than that previously described in the whole FSHD population [Padberg 1982; Padberg et al, 1995; Zatz et al, 1995 ; Tawil et al, 1996; Zatz et al 1998; Lunt, 1998; van der Maarel et al, 2000]. We verified the percentage of *de novo* cases among 4-8 DRA probands from the INRF and surprisingly identified only 14 (5.7%) out 246 index cases, carrying 4-8 DRA. Thus we concluded that among probands from the INRF, carriers of *de novo* rearrangements are significantly more frequent in the cohort of subjects carrying 1-3 DRA, than in the cohort of 4-8 DRA carriers. This data, associated with the observation that 1-3 DRAs previously have never been detected in general population [Scionti et al, 2012b], support the idea that the D4Z4 repeat array is highly recombinogenic and therefore prone to high mutation rate. Second, we found that the majority of cases presenting disease onset before age 10, are isolated and carry a *de novo* rearranged DRA (Table 4); while the majority of familial cases develops FSHD around the second decade of life. Interestingly, among the 66 carriers of 1-3 D4Z4 alleles 6.1% were asymptomatic (Figure 28). However, even though there is a trend towards a more severe progression among *de novo* cases in

comparison with familial probands (Figure 25 and 27) and infantile disease onset is more common among *de novo* carriers of 1-3 DRA (Table 4), we failed to observe a significant difference in disease outcome between the two groups (mean FSHD score in *de novo* versus familial index cases, 9.7 vs 11.1). In addition it is important to emphasize that in this cohort of subjects carrying the shortest D4Z4 allele, not all presented an infantile onset. Indeed 45.5% of index cases reported age at onset after 10 years of age, moreover our Anamnestic Infantile Form revealed no presence of the congenital form of the FSHD. Therefore our study establishes that the group of 1-3 DRA carriers does not present a distinct "Infantile FSHD" as it was reported in the past.

Collectively, our genotype-phenotype study on the large cohort of subjects carrying the shortest allele, revealed that: 1) first signs and/or symptoms were not detected in pre- and peri-natal period; 2) half of the subjects carrying 1-3 DRA develop FSHD in infantile period (1-10 years); 3) not all subjects carrying 1-3 DRA display more severe form of disease, regardless infantile onset; 4) asymptomatic individuals were found to carry 1-3 DRA. These observations suggest that additional factors might contribute to complex FSHD pathogenesis.

5.2 How to explain the phenotypic variability in FSHD?

In the recent years it was proposed that an epigenetic mechanism plays a role in the pathogenesis of FSHD. C-5 methylation of cytosine, the most common epigenetic modification of mammalian DNA [Robertson and Wolffe, 2000]. CpG methylation can affect occupancy of specific genomic regions since several transcription factors and chromatin-binding proteins, such as CTCF and YY1, are sensitive to it [Hark et al, 2000; Kim et al, 2003]. Each D4Z4 repeat unit harbors two classes of GC-rich sequences, namely the low copy-repeats hhspm3 and LSau. On these bases DNA methylation has been considered a marker of epigenetic changes at the D4Z4 locus. Indeed various methylation assays on D4Z4 used by different groups to estimate the level of methylation on two different sites in D4Z4 repeat on chromosome 4q35 found D4Z4 hypomethylation in both FSHD1 and FSHD2 patients. The idea that additional epigenetic factors could be responsible for the FSHD was supported by the observation that in FSHD2 patients, who carry D4Z4 alleles with more than 11 repeats, DNA is hypomethylated at the D4Z4 locus. Consistent with this possibility, it has been recently shown that in a group of FSHD2 families, affected subjects carry mutations in SMCHD1 gene, encoding a chromatin modifier, in association with D4Z4 hypomethylation. In addition a hypomethylation of D4Z4 repeat units has been observed in FSHD patients [de

Greef et al, 2010]. Therefore it has been suggested that reduced penetrance observed in FSHD, might be explained by the presence of additional factors either in association with D4Z4 reduction or independently, contributing to the inter- and intra-familial phenotypic variability observed in FSHD population [Sacconi et al, 2013].

Thus we tested this possibility in FSHD families where segregate 1-8 DRAs and FSHD families where "borderline" 9-11 D4Z4 4q allele segregate. It has to be considered that in 10% of the population, subtelomeric exchanges between the D4Z4 repeats on 4qter and 10qter have been observed [van Deutekom et al, 1997]. These subtelomeric exchanges result in the formation of hybrid alleles containing a mixture of 4-type and 10-type D4Z4 repeat units [Lemmers et al, 1998], which interfere with the Southern blot-based methylation analysis, therefore D4Z4 repeat composition was selective for the study. We investigated the methylation level at two sites on proximal D4Z4 repeat units on chromosome 4q in 140 clinically well characterized subjects, carrying the standard allele 4q-10q constitution (see Figure 14). Using this test we found a strong correlation between the percentage of methylation and total number of D4Z4 in healthy controls ($r_s = 0.075$, Spearman's test p-value < 0.001), while we failed to observe any correlation in FSHD patients ($r_s = 0.014$, Spearman's test p-value > 0.001) carrying a 4q allele with 1-11 D4Z4 repeats. This observation seems to indicate that the methylation status detected by using methylation-sensitive restriction enzymes depend on the total number of D4Z4 units carried by each individual. Comparing the D4Z4 methylation status in non-penetrant DRA carriers with the level of D4Z4 methylation assessed in their FSHD affected relatives, we failed to observe a significant difference of methylation status between the two groups (Student's T test p-value = 0.193). Thus, data obtained from FSHD patients show a lack of correlation between the methylation status of D4Z4 4q arrays and the presence of the disease. By contrast, we detected that degree of D4Z4 methylation correlates with the total number of D4Z4 repeats in healthy individuals. Therefore our study demonstrates that degree of methylation assessed with methylation-sensitive site assay, has no specific role in the complexity of the molecular mechanisms underlying FSHD pathogenesis.

Since the analysis of methylation-sensitive restriction sites does not allow the precise definition of the methylation status of the D4Z4 reduced allele associated with the disease, we decided to analyze the precise methylation of CpGs within the D4Z4 sequence of the DRA. To overcome the technical challenge represented by 98% of homology between D4Z4 repeats on 4q and 10q allele, we develop a method to analyze a single 4q allele. This approach enables us to assess the specific information of D4Z4 methylation status on 4q, in contrast to

the previous study which analyzed global level of methylation on both chromosomes 4q and 10q.

To perform in-depth methylation screening of high number of CpG dinucleotides inside each D4Z4 repeat on chromosome 4q we choose to analyze four different regions encompassing total~ 1.2 kb DNA sequence on D4Z4 and 113 CpGs (see Figure 38). For this study we selected 4q alleles with 3-12 D4Z4 units from 6 unrelated FSHD patients. This approach revealed unexpected data. We found the D4Z4 region 1 (DR1) and D4Z4 binding element region (DBE-R) hypomethylated in all 6 alleles. Interestingly a binding site for CTCF, a zing finger protein, is positioned just proximally to DR1 [Ottaviani et al, 2009]. This methylation-sensitive transcription factor [Amy et al, 2000] was implicated to function as enhancer-blocker for controlling allele specific expression of H19 and Igf2 [Schoenherr et al, 2003, Fedoriw et al, 2004]. The fact that DR1, the region juxtaposed to the CTCF binding site, was hypomethylated in all 4q alleles indicate that binding of CTCF to D4Z4 is possible, enabling transcriptional repression of a nearby genes (Figure 41). Moreover, we also found that the CpG dinucleotide inside the putative binding site of YY1 (Ying-Yang 1) positioned at DBE region (Figure 40) was demethylated in all analyzed sequences.

Studies show that the YY1 binding to the putative motif (CGCCATnTT) is epigenetically modulated hence the YY1 binding capability is abolished when site is methylated [Do Kim et al, 2006; Sekimata et al, 2011]. Our methylation studies are consistent with previous findings showing that YY1 binds DBE and consequently recruit HMGB2 and nucleolin [Gabellini et al, 2002], thus promoting topological repression of D4Z4 region. Overall these results indicate that in D4Z4 reduced allele from FSHD patients the D4Z4 element retains the capability of binding proteins that involved in maintenance epigenetic silencing.

Our study also shows that the DR3, region including the ORF of DUX4, is methylated in the majority of analyzed D4Z4 arrays. This observation might be of particular significance for the possibility of expressing DUX4.

The discovered differences in the methylation status between the proximal, more centromeric, (DR1 and DBE-R) and telomeric DR3 domain in 4q alleles of FSHD patients open new questions. Hypomethylated regions might allow accessibility of epigenetic imprinting-involved modifiers CTCF and YY1, which could maintain the expression of DUX4 silenced. These notions are in concordance with observation that ORF of DUX is methylated in D4Z4 4q alleles of analyzed FSHD patients. By contrast the rmethylation of

D4Z4 region positioned within the DUX4 ORF might explain the very low level of its expression.

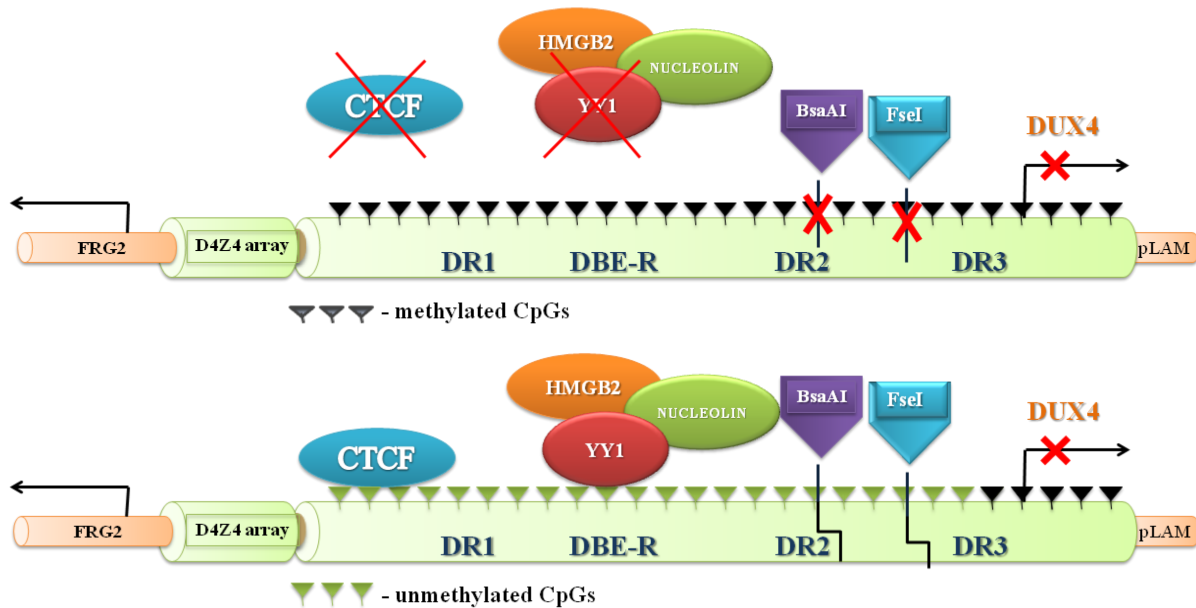

Figure 41: Schematic presentation of a possible epigenetic events in a differentially methylated sequences of D4Z4 repeat.

Altogether our studies confirm the epigenetic control of the D4Z4 region. However the significance of the high methylation level of DR3 should be unraveled in view of the present model for FSHD pathogenesis, which envisions the DUX4 expression as a key event for FSHD pathogenesis.

6. Conclusions

The practice of medical genetics requires a clear, definite evaluation of the significance of mutations/variations of DNA sequences for diagnosis to provide prognostic information and genetic counseling. This is particularly important for a progressive disease with unpredictable onset and a high variability of clinical expression such as FSHD.

The extensive use over the past 20 years of DNA analysis for studying Mendelian disorders has revealed many complex mechanisms in addition to single mutant genes that cause disease. Identical phenotypes may be produced by mutations in different genes [Casanovas et al, 2008], the same mutation can cause different phenotypes [Takashi et al, 1998], and distinct mutations in the same gene may result in different disorders that segregate with diverse Mendelian or even multifactorial patterns [Kanagawa et al, 2006]. In addition the incomplete penetrance of certain mutations argues for the importance of modifying loci or epigenetic mechanisms influencing the clinical expression in many Mendelian disorders [Chahwan et al, 2011]. Thus, establishing the value of mutational events underlying genetic diseases may be complex even when there are simple patterns of inheritance in diseases with a well-characterized pathologic course.

FSHD seems to fall in this complex pattern even though it is currently considered a fully penetrant disease with a wide variability in clinical spectrum, ranging from subjects with very mild muscle weakness to wheelchair bound patients [Tawil et al, 2010; Lunt et al, 1995a] The molecular test initially used for FSHD diagnosis was based on the observation that 95% of FSHD patients carry a reduction of integral numbers of D4Z4 repeats at 4q35 with full penetrance [van Deutekom et al, 1996a]. However the wide use of this test revealed several exceptions to the original model. Through the years the threshold size of D4Z4 alleles has been increased from the original 28 kb (6 repeats) [Wijmenga et al, 1992] to 35 kb (8 repeats) [van Deutekom, 1996a], with FSHD cases carrying D4Z4 alleles of 38-41 kb (9-11 repeats) considered borderline alleles [Butz et al, 2003; Vitelli et al, 1999] . Additional genotype-phenotype studies led to the identification of subjects carrying D4Z4 reduced alleles with no sign of muscle weakness in FSHD families [Ricci et al, 1999; Tonini et al, 2004] as well as in normal controls [van Overveld et al, 2000; Weiffenbach et al, 1992]. Finally FSHD affected subjects who do not carry D4Z4 reduced allele, have been described. In those cases D4Z4 hypomethylation associated with SMCHD1 mutation have been considered at the basis of disease.

Our systematic clinical and molecular analysis of FSHD patients from the Italian National Registry for FSHD as well as a large number of healthy controls display that no simple model can be used to describe FSHD pathophysiology.

Remarkably, our data establish that no general rule can be applied to diagnose FSHD. Although the majority of FSHD patients (70%) carry D4Z4 alleles with 4-8 units, this size range is carried by 3% of healthy subjects from the general population. By extending clinical analysis to a large number of subjects carrying 1-3 DRA we observed a wide range of clinical expression also in this category, which does not fit with the knowledge of a rough inverse correlation with the size of D4Z4 alleles. In addition the methylation status of the D4Z4 region does not strictly correlate with the presence of a clinical phenotype. Besides we observed no correlation between the D4Z4 methylation degree and the length of the D4Z4 array among FSHD patients. Finally we identified a family without DRA and mutation in SMCHD1 in which FSHD segregates as autosomal recessive trait.

In summary, our study indicates that a profound re-thinking of the genetic disease mechanism and modes of inheritance of FSHD are now required and entirely new models and approaches are needed. Our results indicate that other candidate factors are involved. Indeed, our data point at the possibility that a complex mechanism beyond current understanding is at the basis of this complex disease. At our present understanding it is probably necessary to envision different pathogenic models. It is possible that in the heterozygous state a D4Z4 reduction might produce a subclinical sensitized condition that requires other epigenetic mechanisms or a contributing factor to cause overt myopathy. In some rare cases, that could be by becoming homozygous [Scionti et al, 2012a] and doubling the dose of a dominant factor such as DUX4. In others, it might be by the simultaneous heterozygosity for a different and recessive myopathy, as suggested by many reports in which the FSHD contractions are found in association with a second molecular defect [Lecky et al, 1991; Tonini et al, 2002; Reilich et al, 2010]. In other cases FSHD might be the result of homozygous mutations in yet-to-discover genes causing an autosomal recessive form of disease. Finally it is also plausible that drugs or toxic agents might contribute to the disease onset and clinical variability.

It is hoped that broadening the scope of investigations including next generation deep sequencing in particular in families with unusual segregation of FSHD may finally lead to an understanding of the molecular pathogenesis of this complex disease. These findings have important clinical implications for genetic counseling of patients and families with FSHD with particular regard to the interpretation of data in prenatal diagnosis.

Appendix 1

doi:10.1093/brain/awt226 Brain 2013: 136; 3408–3417 | 3408

BRAIN
A JOURNAL OF NEUROLOGY

Large scale genotype–phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy

Giulia Ricci,[1,2,*] Isabella Scionti,[2,3,*] Francesco Sera,[4,*] Monica Govi,[2] Roberto D'Amico,[5] Ilaria Frambolli,[2] Fabiano Mele,[2] Massimiliano Filosto,[6] Liliana Vercelli,[7] Lucia Ruggiero,[8] Angela Berardinelli,[9] Corrado Angelini,[10] Giovanni Antonini,[11] Elisabetta Bucci,[11] Michelangelo Cao,[10] Jessica Daolio,[2] Antonio Di Muzio,[12] Rita Di Leo,[13,14] Giuliana Galluzzi,[15] Elisabetta Iannaccone,[16] Lorenzo Maggi,[17] Valerio Maruotti,[12] Maurizio Moggio,[18] Tiziana Mongini,[7] Lucia Morandi,[17] Ana Nikolic,[2] Ebe Pastorello,[10] Enzo Ricci,[16] Carmelo Rodolico,[13] Lucio Santoro,[8] Maura Servida,[18] Gabriele Siciliano,[1] Giuliano Tomelleri[19] and Rossella Tupler[2,3]

1 Department of Clinical and Experimental Medicine, Neurological Clinic, University of Pisa, via Roma 67, 56126 Pisa, Italy
2 Department of Life Sciences, University of Modena and Reggio Emilia, via G. Campi 287, 41125 Modena, Italy
3 Program in Gene Function and Expression, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA
4 MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, UK
5 Unit of Biostatistics, Department of Clinical and Diagnostic Medicine and Public Health, University of Modena and Reggio Emilia, Via del Pozzo 71, 41124 Modena, Italy
6 Clinical Neurology, University Hospital "Spedali Civili", Piazzale Spedali Civili 1, 25123 Brescia, Italy
7 Department of Neurosciences "Rita Levi Montalcini", University of Turin, via Cherasco 15, 10126 Turin, Italy
8 Department of Neurosciences and Reproductive and Odontostomatologic Sciences, University Federico II, via Pansini 5, 80131 Naples, Italy
9 Unit of Child Neurology and Psychiatry, IRCCS 'C.Mondino' Foundation, via Mondino 2, 27100 Pavia, Italy
10 Department of Neurosciences, University of Padua, via Orus 2, 35129 Padua, Italy
11 Department of Neurology, S. Andrea Hospital, University of Rome 'Sapienza', via Grottarossa 1035, 00189 Rome, Italy
12 Centre for Neuromuscular Disease, CeSI, University 'G. d'Annunzio', via Colle dell'Ara, 66100 Chieti, Italy
13 Department of Neurosciences, Psychiatry and Anaesthesiology, University of Messina, via C. Valeria, 98125 Messina, Italy
14 IRCCS San Camillo Venezia Via Alberoni 70, 30126 Venezia
15 Molecular Genetics Laboratory of UILDM, Lazio Section, IRCCS Santa Lucia Foundation, via Ardeatina 306, 00179 Rome, Italy
16 Department of Neurosciences, Università Cattolica Policlinico A. Gemelli, Largo A. Gemelli 8, 00168 Rome, Italy
17 IRCCS Foundation, C. Besta Neurological Institute, via Celoria 11, 20133 Milan, Italy
18 Neuromuscular Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Dino Ferrari Centre, University of Milan, via F. Sforza 35, 20122 Milan, Italy
19 Department of Neurological and Movement Sciences, University of Verona, Piazzale L.A. Scuro 10, 37134 Verona, Italy

*These authors contributed equally to this work.

Correspondence to: Rossella Tupler,
Program in Gene Function and Expression,
University of Massachusetts Medical School,
364 Plantation Street, Worcester,
MA 01605, USA
E-mail: rossella.tupler@umassmed.edu

Received November 19, 2012. Revised June 23, 2013. Accepted June 27, 2013. Advance Access publication September 11, 2013
© The Author (2013). Published by Oxford University Press on behalf of the Guarantors of Brain.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Facioscapulohumeral muscular dystrophy has been genetically linked to reduced numbers (≤8) of D4Z4 repeats at 4q35 combined with 4A(159/161/168) *DUX4* polyadenylation signal haplotype. However, we have recently reported that 1.3% of healthy individuals carry this molecular signature and 19% of subjects affected by facioscapulohumeral muscular dystrophy do not carry alleles with eight or fewer D4Z4 repeats. Therefore, prognosis for subjects carrying or at risk of carrying D4Z4 reduced alleles has become more complicated. To test for additional prognostic factors, we measured the degree of motor impairment in a large group of patients affected by facioscapulohumeral muscular dystrophy and their relatives who are carrying D4Z4 reduced alleles. The clinical expression of motor impairment was assessed in 530 subjects, 163 probands and 367 relatives, from 176 unrelated families according to a standardized clinical score. The associations between clinical severity and size of D4Z4 allele, degree of kinship, gender, age and 4q haplotype were evaluated. Overall, 32.2% of relatives did not display any muscle functional impairment. This phenotype was influenced by the degree of relation with proband, because 47.1% of second-through fifth-degree relatives were unaffected, whereas only 27.5% of first-degree family members did not show motor impairment. The estimated risk of developing motor impairment by age 50 for relatives carrying a D4Z4 reduced allele with 1–3 repeats or 4–8 repeats was 88.7% and 55%, respectively. Male relatives had a mean score significantly higher than females (5.4 versus 4.0, $P = 0.003$). No 4q haplotype was exclusively associated with the presence of disease. In 13% of families in which D4Z4 alleles with 4–8 repeats segregate, the diagnosis of facioscapulohumeral muscular dystrophy was reported only in one generation. In conclusion, this large-scale analysis provides further information that should be taken into account when counselling families in which a reduced allele with 4–8 D4Z4 repeats segregates. In addition, the reduced expression of disease observed in distant relatives suggests that a family's genetic background plays a role in the occurrence of facioscapulohumeral muscular dystrophy. These results indicate that the identification of new susceptibility factors for this disease will require an accurate classification of families.

Keywords: facioscapulohumeral muscular dystrophy; D4Z4 reduced allele; genotype–phenotype correlations; penetrance; disease expression

Abbreviations: DRA = D4Z4 alleles of reduced size; FSHD = facioscapulohumeral muscular dystrophy

Introduction

Facioscapulohumeral muscular dystrophy (FSHD, OMIM #158900) is the third most common hereditary myopathy with prevalence of 1 in 20 000 (Mostacciuolo *et al.*, 2009). FSHD is characterized by progressive, asymmetric atrophy and weakness of a highly selective set of muscle groups (Padberg *et al.*, 1991; Lamperti *et al.*, 2010; Tawil *et al.*, 2010) and wide inter- and intra-familial variability of clinical expression (Padberg, 1982; Tawil and van der Maarel, 2006). Age-dependent penetrance based on the presence of the characteristic clinical signs was estimated >95% by age 20 (Lunt *et al.*, 1989; Tawil *et al.*, 2010). The mode of inheritance is considered autosomal dominant (Flanigan, 2004).

A large majority of patients with FSHD carry rearrangements occurring in a 3.3 kb tandemly repeated sequence (D4Z4) located at the 4q subtelomeric region (Wijmenga *et al.*, 1992; Lunt *et al.*, 1995a; Upadhyaya *et al.*, 1997). These rearrangements result in polymorphic EcoRI alleles detected by the p13E-11 probe (Wijmenga *et al.*, 1992; Upadhyaya *et al.*, 1997). Early studies of small numbers of individuals observed that both *de novo* and familial patients with FSHD carry p13E-11 EcoRI alleles of 35 kb, corresponding to eight D4Z4 units, or shorter (Griggs *et al.*, 1993; Lunt *et al.*, 1995a; van Deutekom *et al.*, 1996). For the last 20 years the clinical diagnosis of FSHD has been supported by this type of D4Z4 DNA testing, which has been considered highly sensitive and specific for disease (Lunt *et al.*, 1995*a*, *b*; Lunt, 1998; Tawil *et al.*, 2010). However, several studies on FSHD families describe subjects carrying D4Z4 alleles of reduced size (DRA) and no signs of the disease, defined as non-penetrant carriers (Tawil *et al.*, 1996; Zatz *et al.*, 1998; Ricci *et al.*, 1999; van Overveld *et al.*, 2000; Goto *et al.*, 2004; Tonini *et al.*, 2004; Sakellariou *et al.*, 2012; Scionti *et al.*, 2012*a*). As a possible explanation of some non-penetrant cases, it has been proposed that reduction of D4Z4 repeats on chromosome 4q35 is pathogenic only in certain chromosomal backgrounds, defined by 'permissive' specific haplotypes, namely (i) reduction of D4Z4 elements; (ii) presence of the 4A(159/161/168) haplotype; and (iii) a single nucleotide polymorphism that provides a polyadenylation signal (PAS) for the *DUX4* transcript (Lemmers *et al.*, 2002, 2007, 2010).

Nonetheless, our most recent studies (Scionti *et al.*, 2012*b*) showed that although the majority of FSHD index cases (70%) carry DRA with 4–8 units, this size range is also carried by 3% of healthy subjects from the general population. Additionally, our work raised the possibility that there is little predictive value of the 4A161PAS haplotype in the absence of family history because 1.3% of healthy subjects carry this haplotype, which is a frequency similar to other common polymorphisms. Finally, we found that 19% of FSHD probands do not carry D4Z4 alleles with 1–8 repeats and only 50% of the probands carry the 4A161PAS permissive haplotype associated with DRA (Scionti *et al.*, 2012*b*). These observations suggest that the genetic factors leading to FSHD might be incompletely described. Consistent with this idea, Lemmers *et al.* (2012) recently described mutations in *SMCHD1* gene in patients with FSHD and hypothesized that these mutations influence the disease penetrance (Lemmers *et al.*, 2012).

Here, we evaluate FSHD occurrence among relatives carrying DRA in relation to D4Z4 reduced allele size, gender, age, degree of kinship and 4q haplotype.

Materials and methods

Study design and subjects selection

The study has been performed on FSHD families accrued through the Italian National Registry for FSHD (INRF), established in 2007 by the Italian Clinical Network for FSHD (ICNF) (www.fshd.it) (Lamperti *et al.*, 2010). The ICNF includes two diagnostic laboratories at the University di Modena and Reggio Emilia and at the Fondazione Santa Lucia in Rome and 14 clinical centres, networked within the Italian Association of Myology (www.miologia.org) and distributed across all of Italy, from northern to southern regions, including the islands.

The study was conducted from 2008 to 2012. As outlined in Fig. 1, initially the selection process regarded 418 index cases carrying DRA with 1–8 repeats and fulfilling the clinical features of FSHD (Padberg *et al.*, 1991). One hundred and eighty-six cases were considered not eligible because they had no available relatives. Fourteen compound heterozygotes for DRA alleles were excluded from this study and analysed separately (Scionti *et al.*, 2012a). Forty-two *de novo* cases, defined as single subjects with neither parent carrying DRA, were excluded because they would not be informative for this study. For each proband the clinical and molecular examinations were extended to the available relatives at various degrees of kinship. Among the 645 relatives identified, 367 were found to be carriers of DRA. All clinical and molecular data were collected in the INRF database at Miogen Laboratory of University of Modena for data analysis.

Clinical examination

Each subject recruited during the time of the study was examined by a trained neurologist of the ICNF using the standardized FSHD clinical protocol with validated inter-rater reliability (Lamperti *et al.*, 2010). The FSHD clinical protocol was developed by the ICNF in order to numerically define the clinical severity of the motor impairment, and is not to be used to diagnose FSHD. The FSHD scale quantifies muscle weakness through the functional evaluation of six muscle groups specifically affected in FSHD, belonging to (i) face (score 0–2); (ii) shoulder girdle (score 0–3); (iii) upper limbs (score 0–2); (iv) distal legs (score 0–2); (v) pelvic girdle (score 0–5); and (vi) abdominal muscles (score 0–1). The FSHD score, which translates disability into a number, ranges from zero, when no objective evidence of muscle functional impairment is present, to 15, when all the muscle groups tested are severely impaired (www.fshd.it) (Lamperti *et al.*, 2010). DRA carriers who did not show an objective motor impairment received an FSHD score equal to zero and were considered clinically unaffected at the time of examination. On the basis of the FSHD score, subjects were classified as affected by mild (FSHD score 1–2), moderate (FSHD score 3–6), or severe (FSHD score 7–15) motor impairment.

Probands from 13 families were not re-evaluated as they were not alive at the time of this study.

Age at onset was estimated on the basis of patient records. When subjects did not complain of motor impairment, but a mild muscle weakness was observed, the age at examination was set as the age at onset, according to previous reports (Lunt *et al.*, 1995*b*). In six subjects it was not possible to obtain information about the age at onset of motor impairment due to their poor compliance.

The study was approved by the Local Ethics Committees of all participating Institutions. Informed consent, according to the Declaration of Helsinki, was obtained from each subject enrolled in the study.

Molecular characterization

Allele sizes were estimated by Southern hybridization using probe p13E-11. Genomic DNA extracted from peripheral blood lymphocytes was digested with EcoRI, EcoRI/BlnI or XapI, electrophoresed in a 0.4% agarose gel for 45–48 h at 35 V alongside an 8–48 kb marker (Bio-Rad) as previously described (Scionti *et al.*, 2012*b*). To assess the chromosomal origin of D4Z4-reduced alleles, DNA from each subject was analysed by NotI digestion and hybridization with the B31 probe (Scionti *et al.*, 2012*b*). Restriction fragments were detected by autoradiography or using a Typhoon Trio system (GE Healthcare). 4qA/4qB allelic variants were defined in all 530 subjects included in the study, using HindIII-digested DNA, pulsed field gel electrophoresis electrophoresis and Southern blot hybridization with radiolabeled 4qB and 4qA probes according to standard procedures (Scionti *et al.*, 2012*b*).

The Simple Sequence Length Polymorphism (SSLP) and the pLAM Single Nucleotide Polymorphism (SNP) [AT(T/C)AAA] sequences flanking the D4Z4 repeat units were defined in 294 relatives according to published procedures (Scionti *et al.*, 2012*b*).

Statistical analysis

The association between the risk of being asymptomatic (FSHD score equal to zero) and D4Z4 allele size and age was evaluated by using the multivariate logistic regression model. The association between age at onset or FSHD score and D4Z4 allele size and sex among symptomatic relatives was assessed by using a general linear model. Association estimates and their relative 95% confidence intervals (CI) were also reported.

The prevalence of FSHD score, classified into two categories (0 versus 1–15), among relatives was estimated and its association with D4Z4 allele sizes was also evaluated. Univariate and multivariate logistic regression models were fitted with D4Z4 allele size, sex and family degree as predictors.

An interaction test was also carried out to assess whether the difference in terms of FSHD scores between probands and relatives varied in relation to D4Z4 allele sizes.

In order to minimize any ascertainment bias, all the genotype–phenotype correlation analyses were performed on relatives and probands separately.

In all general and generalized linear models estimated, the sandwich estimator of covariance matrix of parameters was used to take into account any clustering effect within families (Williams, 2000). Wald tests were used to evaluate the effect of predictors and to evaluate the effects of predictors on outcomes (McCullagh and Nelder, 1989).

For the cohorts of probands and family members, Kaplan-Meier survival analysis (Kaplan and Meier, 1958) was used to estimate the age-specific cumulative motor impairment incidence, with the corresponding 95% CI. For each individual, time from birth to the earliest age at onset of motor impairment was estimated. The analysis was stratified by D4Z4 allele size only for relatives, and also by gender for relatives and probands.

The risk prediction algorithm was developed and validated using established methods (Hippisley-Cox *et al.*, 2007). The original cohort of 367 relatives was randomly split in the derivation and validation

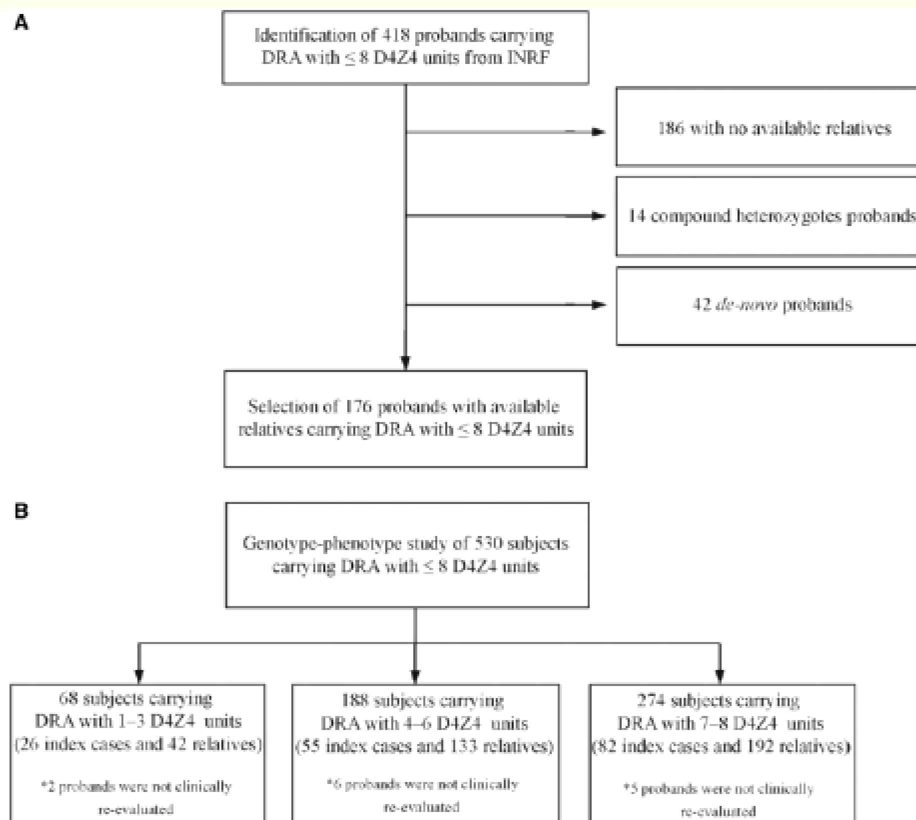

Figure 1 (A) Preliminary selection of probands/families from the Italian National Registry for FSHD (INRF). (B) Selection of the cohort of probands and their relatives for genotype–phenotype correlation analysis.

samples. The coefficients for D4Z4 allele size, sex and family degree were estimated by using the Cox proportional hazard model. The coefficients were used as weights, which were combined with the baseline survivor function to derive risk equations at age 50 years. The risk equation was applied to the validation sample and measures of discrimination were calculated (R^2, D statistics and area under the receiver operating characteristic curve).

Results

We examined 530 carriers of DRA (367 relatives and 163 index cases) from 176 unrelated families, in which at least one subject was affected by FSHD (Fig. 1B). Considering the cohort of relatives carrying DRA as a whole (367 subjects, 152 males and 215 females, mean age 46.4 ± 17.2, Supplementary Fig. 1), we observed that 118 (32.2%) did not show any functional motor impairment (FSHD score equal to zero) and 249 (67.8%) displayed muscle impairment to various degrees (FSHD score ≥1) (Table 1).

The distribution of asymptomatic relatives was also analysed based on the size of DRA. We divided subjects in three groups: subjects carrying DRA with 1–3 D4Z4 repeats; subjects carrying DRA with 4–6 D4Z4 repeats; subjects carrying DRA with 7–8 D4Z4 repeats (Supplementary Table 1 and Supplementary Fig. 1). Table 1 also shows that 9.5% (4 out of 42) of all carriers of DRA with 1–3 repeats displayed no motor impairment. This percentage increases among carriers of DRA with 4–6 and 7–8 repeats (28.6% and 39.6%, respectively).

We then calculated the distribution of asymptomatic carriers based on age at examination, separately analysing four classes: those aged 18–30 years, 31–55 years, 56–70 years, and those over 70 years of age (Supplementary Table 2). Four classes were formed. As shown in Table 1, asymptomatic carriers were found in all classes up to 70 years. In particular, almost one-third of carriers of DRA with 4–6 and 7–8 repeats (27.6% and 35.9%, respectively) were asymptomatic between 56 and 70 years of age. Since the percentage of asymptomatic carriers varies among relatives carrying DRA of different sizes, the age-related risk of developing

Table 1 Distribution of unaffected relatives according to D4Z4 allele size and age at examination

	Age (years)						Total		
	18–30		31–55		56–70				
D4Z4 units	*n* subjects	% Score = 0 (*n*)	*n* subjects	% Score = 0 (*n*)	*n* subjects	% Score = 0 (*n*)	*n* subjects	% Score = 0 (*n*)	P-value*
1–3	8	12.5 (1)	23	8.7 (2)	9	11.1 (1)	42	9.5 (4)	0.707
4–6	31	25.8 (8)	65	33.8 (22)	29	27.6 (8)	133	28.6 (38)	0.461
7–8	42	54.8 (23)	85	40.0 (34)	39	35.9 (14)	192	39.6 (76)	0.013
Total							367	32.2 (118)	

*Wald test of the age's coefficient as ordinal predictor in the logistic model adjusted by sex.

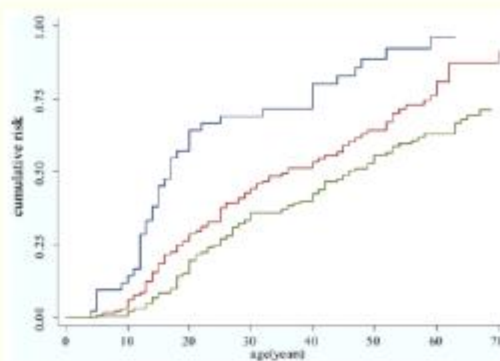

Figure 2 Age-specific cumulative risk of reported muscle impairment according to D4Z4 allele size. Estimates obtained on 361 relatives using the Kaplan-Meier analysis. Blue line refers to carriers of 1–3 D4Z4 units; red line refers to carriers of 4–6 D4Z4 units; green line refers to carriers of 7–8 D4Z4 units. Carriers of 7–8 versus 4–6 units Log rank test *P* value = 0.002.

Table 2 Estimates of the age-specific cumulative risk obtained using the Kaplan-Meyer analysis

Age (years)	Carriers of 1–3 D4Z4 units		Carriers of 4–6 D4Z4 units		Carriers of 7–8 D4Z4 units	
	Risk	95% CI	Risk	95% CI	Risk	95% CI
20	64.3	(50.1; 78.3)	21.8	(21.8; 37.5)	19.6	(14.6; 26.0)
30	69.1	(55.0; 82.2)	36.1	(36.1; 53.4)	35.9	(29.4; 43.5)
40	80.1	(66.5; 90.8)	44.8	(44.8; 62.5)	42.5	(35.5; 50.3)
50	88.7	(76.3; 96.3)	55.0	(55.0; 73.3)	55.7	(47.9; 63.8)
60	96.2	(84.6; 99.7)	71.5	(71.5; 88.8)	62.9	(54.7; 71.2)
70			80.3	(80.3; 97.6)	71.3	(62.3; 79.7)
80					82.2	(72.1; 90.4)

Table 3 Distribution of mean age at onset among affected relatives according to D4Z4 allele size

	Relatives			
D4Z4 units	Number of subjects	Mean age at onset (years)	95% CI	P-value†
1–3	38	20.3	15.5–25.2	
4–6	91	29.2	25.6–32.7	
7–8	114	34.6	30.1–39.1	0.0002
Total	243			

†Wald test of equality to zero of D4Z4 allele size's coefficients parametrized as categorical variable in a general linear model with age at onset as dependent variable and sex and D4Z4 allele size as predictors.

motor impairment was evaluated in correlation with D4Z4 size on the basis of data obtained from 361 relatives. Figure 2 and Table 2 show the penetrance estimates for DRA carriers calculated with the Kaplan-Meier method. Among subjects carrying DRA with 1–3 units the risk of developing motor impairment is 64.3% by age 20, 80.1% by age 40 and 96.2% by age 60. Among subjects carrying DRA with 4–6 and 7–8 D4Z4 units these risks are 21.8% and 19.6%, respectively, by age 20, 44.8% and 42.5% by age 40, and 71.5% and 62.9% by age 60. Therefore, FSHD penetrance is almost complete by age 60 only for carriers of DRA with 1–3 units.

We tested whether the size of DRA correlates with age at onset and disease severity. Table 3 shows that the mean age at onset is statistically lower among subjects carrying DRA with 1–3 units (20.3 years) in comparison with those carrying DRA with 4–6 and 7–8 D4Z4 repeats (29.2 and 34.6 years, respectively) ($P = 0.0002$). Overall, we found that 60.6% of affected relatives experienced scapular girdle onset, 19.0% facial muscle onset, 16.7% lower limbs onset, 0.9% upper limbs onset and 2.8% abdominal muscle onset (Supplementary Table 3).

Severity is also increased among carriers of DRA with 1–3 repeats. Indeed, as shown in Table 4, affected relatives carrying DRA with 1–3 repeats had a mean FSHD score of 7.2. By contrast, individuals carrying DRA with 4–6 and 7–8 D4Z4 units had mean FSHD score of 4.4 and 4.1, respectively. This association was statistically significant ($P = 0.0006$).

The degree of motor impairment among relatives was also evaluated in relationship to D4Z4 allele size and age at examination. Figure 3A shows that ~40% of relatives carrying DRA with 1–3 units are severely affected (FSHD score ≥7) by age 30. In contrast, no relatives carrying DRA with 4–8 units had a FSHD score

higher than 6 in this age window. Figure 3B and C shows that between age 31–55 and 56–70 a high percentage of relatives carrying DRA with 4–8 units were asymptomatic (FSHD score 0) or displayed minimal signs of functional motor impairment (FSHD score 1–2).

Table 4 Distribution of FSHD score calculated on affected relatives according to D4Z4 allele size and age at examination

	Relatives			
D4Z4 units	Number of subjects	FSHD score mean	95% CI	P-value†
1–3	38	7.2	5.8–8.6	
4–6	95	4.4	3.8–5.1	
7–8	116	4.1	3.5–4.7	0.0006

†Wald test of equality to zero of D4Z4 allele size and age at examination coefficients parametrized as categorical variable in a general linear model with FSHD score as dependent variable and sex, D4Z4 allele size and age at examination as predictors.

We then evaluated whether there is a correlation between the clinical status of the proband and his/her relatives. As shown in Table 5, intra-familial analysis on 163 families with 217 affected relatives revealed a positive correlation [0.72 (95% CI 0.40–1.04)] in families in which D4Z4 alleles with 1–3 D4Z4 repeats segregate. In contrast, in families with 4–6 and 7–8 D4Z4 alleles, a lower degree of correlation between the clinical status of the proband and his/her relatives was observed [0.01 (95% CI −0.23–0.26) and −0.14 (95% CI −0.35–0.07), respectively].

Remarkably, in 19 of 148 families (13%) in which 4–8 D4Z4 alleles segregate, we found affected individuals only within a single generation, and with older unaffected relatives carrying the DRA (Supplementary Fig. 2). In each of these 19 families, molecular testing excluded the presence of somatic mosaicism in the unaffected parent carrying the DRA. The finding of affected subjects in only one generation or the presence of only one affected subject in the entire family suggests that a complex mode of inheritance might be at the basis of FSHD development in these families.

To test this hypothesis we assessed whether the prevalence of disease varies among relatives according to degree of kinship with

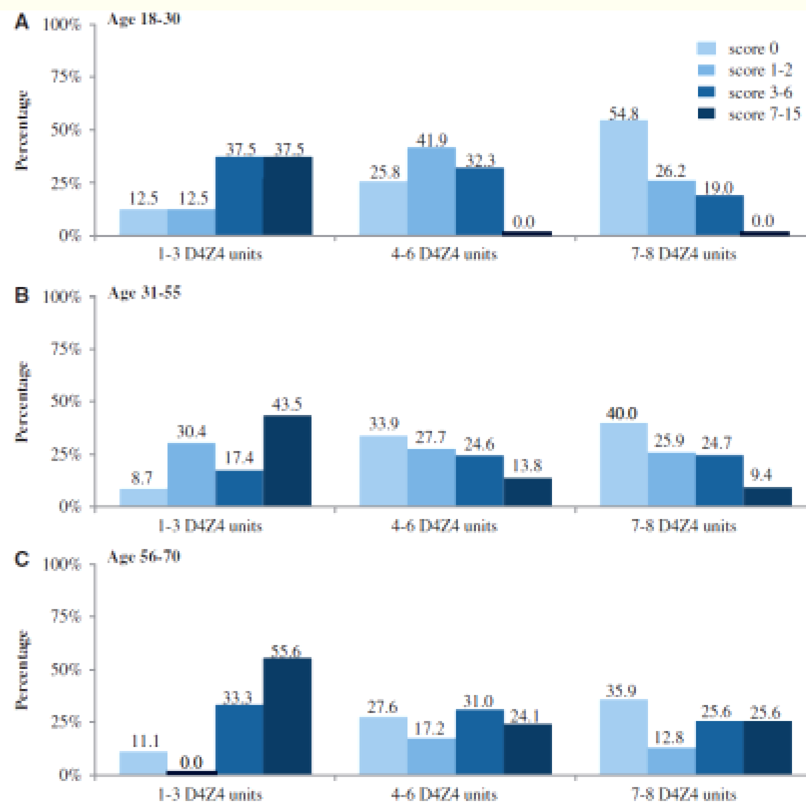

Figure 3 Distribution of clinical severity among relatives carrying D4Z4 reduced allele according to D4Z4 allele size and age. Subjects were subdivided by age: (A) 18–30 years, (B) 31–55 years, (C) 56–70 years and by D4Z4 allele size: 1–3, 4–6 and 7–8 units. In each subgroup, percentages of subjects who received FSHD score equal to 0, 1–2, 3–6 and ⩾ 7 are reported.

Table 5 Standardized regression coefficient between FSHD score of probands and relatives

Number of subjects	D4Z4-allele size 1–3		D4Z4-allele size 4–6		D4Z4-allele size 7–8		P-value
	Correlation coefficient	95% CI	Correlation coefficient	95% CI	Correlation coefficient	95% CI	
217	0.72	0.40–1.04	0.01	−0.23–0.26	−0.14	−0.35–0.07	<0.0001

General linear models with FSHD score of the relative as outcome and probands' FSHD score, age at examination, D4Z4 allele size and sex as predictors. Interaction test between proband's FSHD score and D4Z4 allele sizes.

the proband (distribution is reported in Supplementary Table 4). Table 6 shows that 72.5% of first-degree relatives are affected. This percentage significantly decreases to 52.9% among relatives with lower degree of kinship (from second- to fifth-degree), irrespective of D4Z4 size allele, sex and age at examination ($P = 0.018$), supporting the hypothesis that beside DRA, additional genetic factors may be necessary to develop FSHD. Conversely 47.1% of second- through fifth-degree relatives was unaffected, while only 27.5% of first-degree family members did not show any motor impairment.

It has also been observed that FSHD affects males more severely and more frequently than females (Zatz *et al*., 1998; Tonini *et al*., 2004; Sakellariou *et al*., 2012). We thus evaluated whether gender influences expression and severity of motor impairment. We observed that the percentage of asymptomatic carriers does not significantly differ between genders (data not shown). Instead, as shown in Table 7, male relatives had a significantly higher mean FSHD score (5.4 versus 4.0, $P = 0.003$) and they developed motor impairment on average 7.3 years before than females ($P = 0.003$). Thus male relatives who develop motor impairment had a more severe disease than affected female relatives. We then calculated the risk of developing motor impairment between 20–50 years in females and males separately using the Kaplan-Meier method. As shown in Fig. 4A, the risk is higher in male relatives than females, although the difference is not statistically significant (log rank test *P*-value 0.113). Among probands the risk of developing motor impairment after age 20 is higher in males than in females (log rank test *P*-value = 0.028) (Fig. 4B). Remarkably, the risk becomes similar between genders after age 50.

All index cases and their relatives recruited in the present study carried the 4qA allele. As it has been recently proposed that FSHD occurs only when DRA at 4q35 are in combination with the 4A(159/161/168)PAS haplotype (Lemmers *et al*., 2010), we further characterized DNA polymorphisms flanking the D4Z4 reduced array in 294 subjects (203 affected and 91 unaffected) belonging to 133 families from the cohort selected for this study. Table 8 reports the various haplotypes detected. All were associated with the polyadenylation signal (ATTAAA) that stabilized transcripts from *DUX4* gene. Notably, the 4A161PAS haplotype previously considered 'permissive' and the 4A166PAS haplotype previously considered 'non-permissive' for FSHD disease were detected in both DRA carriers with motor impairment (FSHD score ≥ 1) and without motor impairment (FSHD score 0). On this basis we conclude that no specific 4q haplotype can be considered as predictive of disease.

Collectively, the statistical analysis conducted on the entire cohort of relatives carrying DRA with 1–3 or 4–8 repeats indicates

Table 6 Prevalence of FSHD scores according to degree of kinship

Degree of kinship	FSHD score 0		FSHD score 1–15		P-value[‡]
	Number of subjects	%	Number of subjects	%	
First	77	27.5	203	72.5	
Second/Fifth	41	47.1	46	52.9	0.018

[‡]Wald test of coefficients associated to second or third degree of kinship in logistic models adjusted by D4Z4 allele size, sex and age at examination.

that individuals carrying DRA with 1–3 repeats have a high risk of developing motor impairment by age 50 (83–93%), regardless of sex or degree of kinship. In contrast, in the group with 4–8 repeats the reduced risk of becoming symptomatic (55-63% by age 50) is also modulated by sex (males show a higher risk than females) and degree of kinship (first degree relatives show a higher risk than second-fifth degree relatives).

Discussion

Before the discovery of rearranged D4Z4 alleles, the diagnosis and counselling of FSHD families was entirely based on clinical evidence (Lunt *et al*., 1989). Over the years, DNA testing of the D4Z4 locus and flanking polymorphisms has been considered highly sensitive and specific and extensively used to diagnose FSHD (Tawil *et al*., 2010). However, two recent discoveries have challenged the current understanding of the prognostic value of D4Z4 reduced alleles (DRA) in FSHD families: (i) alleles with reduced numbers (≤ 8) of D4Z4 repeats at 4q35 combined with 4A(159/161/168)PAS haplotype, have a frequency of 1.3% among healthy subjects from the general population; and (ii) only 50% of FSHD probands carry the 4A161PAS permissive haplotype associated with DRA (Scionti *et al*., 2012*b*). Therefore, our understanding of the factors that cause FSHD is incomplete and we conclude that it is crucial for clinical practice to define further elements that can influence motor impairment and can support the interpretation of molecular testing in FSHD families.

The present results rely on a population-based study involving index cases recruited from all regions of Italy. In order to minimize any ascertainment bias the analyses were performed excluding index cases and the evaluation of motor impairment was based

Table 7 **Distribution of FSHD score and age at onset calculated on affected relatives according to sex**

Sex	FSHD Score				Age at onset			
	Number of subjects	FSHD score mean	95% CI	P-value†	Number of subjects	Mean age at onset (years)	95% CI	P-value§
Male	102	5.4	4.7–6.1		99	26.8	(23.2; 30.5)	
Female	147	4.0	3.5–4.5	0.003	144	34.1	(30.5; 37.7)	0.003
Total	249				243			

†Wald test of equality to zero of female sex coefficients in a general linear model with FSHD score as dependent variable sex, D4Z4 allele size and age at examination as predictors.
§Wald test of equality to zero of female sex coefficients in a general linear model with age at onset as dependent variable and sex and D4Z4 allele size as predictors.

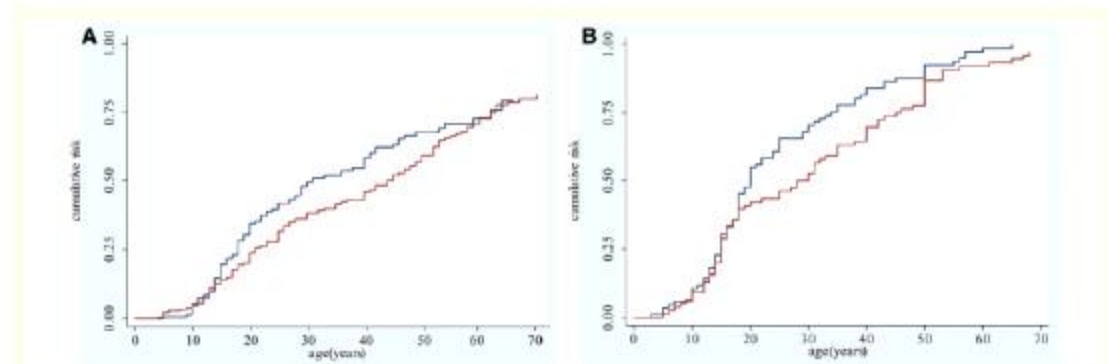

Figure 4 Age-specific cumulative risk of reported muscle impairment according to sex. (**A**) Estimates obtained on 361 relatives using the Kaplan-Meier analysis. Log-rank test *P*-value = 0.113. (**B**) Estimates obtained on 160 probands using the Kaplan-Meier analysis. Blue line refers to male; red line refers to female. Log-rank test *P*-value = 0.028.

Table 8 **Distribution of haplotypes on 294 relatives**

	Relatives					
	4A161 (n = 204)	4A162 (n = 14)	4A163 (n = 5)	4A164 (n = 1)	4A166 (n = 69)	4A167 (n = 1)
FSHD score = 0						
Number of subjects (%)	72 (79.1)	5 (5.5)	2 (2.2)	0 (0.0)	12 (13.2)	0 (0.0)
FSHD score ≥ 1						
Number of subjects (%)	132 (65.0)	9 (4.4)	3 (1.5)	1 (0.5)	57 (28.1)	1 (0.5)

on a standardized protocol shared within the ICNF. However, beside these strengths, the study has some limitations. First, the genetic background and the socio-demographic characteristics of the Italian population might restrict the external validity of the study results. Second, even though the study has a good coverage of index cases, the involvement of relatives might be due to the presence of any symptoms with the consequence that the healthy ones might be under-represented in the study. In that case the true estimated prevalence of disease among relatives would be lower. Third, given that FSHD is a rare disease and no routinely collected diagnosis records are available (Lunt *et al*., 1989) the age at onset was collected retrospectively. Therefore we cannot rule out the presence of recall bias. Indeed, the perception of disease onset may be subjective and could depend on the specific motor skills required in daily activities. It is thus possible that in a number of subjects the motor impairment of limbs may be perceived as early symptom because more disabling. According to this possibility and consistent with previous works (Tawil and van der Maarel, 2006; Pastorello *et al*., 2012), in our cohort the most frequently complained symptom at onset was also the impairment of upper girdle (Supplementary Table 3). Nevertheless, we considered that patient's complaints provide a reliable estimate of the time of functional disability onset related to disease. When subjects did not refer any motor impairment, but a mild

muscle weakness was observed at the clinical evaluation, the age at examination was arbitrarily set as the age at onset (Lunt *et al.*, 1995*b*).

Given these premises, our study shows that FSHD penetrance in DRA carriers is not complete by age 20, as previously proposed (Tawil *et al.*, 2010), as asymptomatic carriers in all the classes of ages up to 70 years were found.

The present analysis highlights different prognostic values of DRA with 1–3 units when compared with DRA with 4–8 units. First, among carriers of DRA with 1–3 units FSHD penetrance is almost complete; in contrast, ~30% of carriers of DRA with 4–8 units older than 55 years display no muscle weakness (Table 1). Second, the estimated risk of developing motor impairment by age 50 differs between the two classes of alleles. Carriers of DRA with 1–3 units have a risk of 88.5% of developing motor impairment by age 50; instead the risk among carriers of DRA with 4–8 units by the same age is 55% (Fig. 2). Third, 44% of carriers of DRA with 1–3 units develop severe FSHD (FSHD score ≥ 7) by age 55; whereas only 24% of carriers of DRA with 4–8 units develop disease with high degree of severity by the same age (Fig. 3B). Fourth, the clinical phenotype is more homogeneous in families with DRA with 1–3 units, as shown by the intra-familial analysis (Table 5). In contrast, the clinical status of probands does not seem to be predictive of disease severity in relatives carrying DRA with 4–8. Importantly, in these families with DRA with 4–8 units, the penetrance of FSHD is lower as the degree of relationship to the affected individual becomes more distant, indicating that the genetic background can affect the disease outcome.

Our study also shows that gender influences disease expression, because males are characterized by a lower mean age at onset of motor impairment (26.8 years in males versus 34.1 years in females, Table 7) and by a more severe disability in terms of FSHD score (5.4 in males versus 4.0 in females). Interestingly, the risk of developing motor impairment is higher in male relatives during adult age (range 18–55 years), whereas it is similar between males and females in childhood/teens and elderly age (Fig. 4). Overall, these data indicate that variables related to gender, including genetic, hormonal, and/or lifestyle factors, may be considered and should be further investigated. Finally, our study suggests that the predictive value of 4q haplotypes must be carefully considered because no specific 4q haplotype was exclusively associated with the presence of disease.

In summary, the genotype-phenotype correlation study presented here confirms that DRAs with 4–8 repeats have no definitive prognostic value, and that other prognostic parameters, beside DRAs, such as sex and degree of kinship with the proband should be considered. We estimated that the risk of developing the motor impairment by age 50 in FSHD family members is higher (83–93%) in subjects carrying DRA with 1–3 repeats. Instead, considering the cohort of relatives carrying DRA with 4–8 repeats, the risk of developing motor impairment is 48% for females and/or subjects with lower degree of kinship and raises to 55–63% for males and/or subjects with first degree of kinship with the proband. None of the various 4q haplotypes detected in FSHD families studied here were exclusively associated with the presence of disease, as reported in Table 8.

Interestingly, in our cohort, 19 of 148 FSHD families (13%) in which a DRA with 4–8 units segregates presented affected subjects only in one generation (Supplementary Fig. 2). In these cases the lack of autosomal dominant inheritance should prompt us to consider whether the disease develops because of the presence of additional genetic defect(s). This possibility is supported by recent observation that mutations in the *SMCHD1* gene segregate independently from the FSHD permissive D4Z4 allele on chromosome 4 in FSHD subjects that do not carry a DRA, also defined as patients with FSHD2 (Lemmers *et al.*, 2012). Therefore searches for secondary FSHD loci should be considered in all cases in which the ratio between affected and unaffected individuals expected for an autosomal dominant disease is not observed and random association between the DRA and FSHD cannot be excluded.

For all of these reasons, to define the predictive value of DRA, it is necessary to carry out clinical evaluation and collection of DNA samples of all of the proband's family members, not only in a research setting but also in clinical practice. We believe that broadening the analysis of FSHD families may facilitate genetic counselling of patients and families with FSHD in particular when interpreting the data for prenatal diagnosis.

Acknowledgements

We are indebted to all patients with FSHD and their families for participating in this study. The Associazione Amici del Centro Dino Ferrari-University of Milan is gratefully acknowledged. We thank Dr. Hulya Gundesli, Professor Woodring E. Wright and Dr. Paul D. Kaufman for their in-depth critique of the manuscript.

Funding

This work was supported by Telethon Italy GUP08004, by Telethon Italy GUP11009, by Association Française contre les Myopathies (AFM) grant number 14339 and by National Institutes of Health (NIH)-National Institute of Neurological Disorders and Stroke (NINDS) grant number RO1 NS047584.

Supplementary material

Supplementary material is available at *Brain* online.

References

Flanigan KM. Facioscapulohumeral muscular dystrophy and scapulohumeral syndrome. In: Engel AG, Franzini-Armstrong C, editors. Myology. New York: McGraw-Hill; 2004. p. 123–34.

Goto K, Nishino I, Hayashi YK. Very low penetrance in 85 Japanese families with facioscapulohumeral muscular dystrophy 1A. J Med Genet 2004; 41: e12.

Griggs RC, Tawil R, Storvick D, Mendell JR, Altherr MR. Genetics of facioscapulohumeral muscular dystrophy: new mutations in sporadic cases. Neurology 1993; 43: 2369–72.

Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular

disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007; 335: 136.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53: 457–81.

Lamperti C, Fabbri G, Vercelli L, D'Amico R, Frusciante R, Bonifazi E, et al. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: the FSHD clinical score. Muscle Nerve 2010; 42: 213–17.

Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, et al. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet 2002; 32: 235–6.

Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet 2012; 44: 1370–4.

Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, et al. A uniying genetic model for facioscapulohumeral muscular dystrophy. Science 2010; 329: 1650–3.

Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, de Knijff P, et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet 2007; 81: 884–94.

Lunt PW, Compston DA, Harper PS. Estimation of age dependent penetrance in facioscapulohumeral muscular dystrophy by minimising ascertainment bias. J Med Genet 1989; 26: 755–60.

Lunt PW, Jardine PE, Koch MC, Maynard J, Osborn M, Williams M, et al. Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). Hum Mol Genet 1995a; 4: 951–8.

Lunt PW, Jardine PE, Koch MC, Maynard J, Osborn M, Williams M, et al. Phenotypic-genotypic correlation will assist genetic counseling in 4q35-facioscapulohumeral muscular dystrophy. Muscle Nerve 1995b; 2: S103–9.

Lunt PW. 44th ENMC international workshop: facioscapulohumeral muscular dystrophy: molecular studies 19–21 July 1996, Naarden, The Netherlands. Neuromuscul Disord 1998; 8: 126–30.

McCullagh P, Nelder JA. Generalized linear models. 2nd edn. New York: Chapman and Hall, Inc; 1989.

Mostacciuolo ML, Pastorello E, Vazza G, Miorin M, Angelini C, Tomelleri G, et al. Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. Clin Genet 2009; 75: 550–55.

Padberg G. Facioscapulohumeral disease. MD thesis. University of Leiden; 1982.

Padberg GW, Lunt PW, Koch M, Fardeau M. Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul Disord 1991; 1: 231–4.

Pastorello E, Cao M, Trevisan CP. Atypical onset in a series of 122 cases with facioscapulohumeral muscular dystrophy. Clin Neurol Neurosurg 2012; 114: 230–4.

Ricci E, Galluzzi G, Deidda G, Cacurri S, Colantoni L, Merico B, et al. 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 1999; 45: 751–7.

Sakellariou P, Kekou K, Fryssira H, Sofocleous C, Manta P, Panousopoulou A, et al. Mutation spectrum and phenotypic manifestation in FSHD Greek patients. Neuromuscul Disord 2012; 22: 339–49.

Scionti I, Fabbri G, Fiorillo C, Ricci G, Greco F, D'Amico R, et al. Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. J Med Genet 2012a; 49: 171–8.

Scionti I, Greco F, Ricci G, Govi M, Arashiro P, Vercelli L, et al. Large scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy (FSHD). Am J Hum Genet 2012b; 90: 628–65.

Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, et al. Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. Ann Neurol 1996; 39: 744–48.

Tawil R, van der Maarel SM. Facioscapulohumeral muscular dystrophy. Muscle Nerve 2006; 34: 1–15.

Tawil R, van der Maarel S, Padberg GW, van Engelen BG. 171st ENMC international workshop: standards of care and management of facioscapulohumeral muscular dystrophy. Neuromuscul Disord 2010; 20: 471–5.

Tonini MM, Passos-Bueno MR, Cerqueira A, Matioli SR, Pavanello R, Zatz M. Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromuscul Disord 2004; 14: 33–8.

Upadhyaya M, Maynard J, Rogers MT, Lunt PW, Jardine P, Ravine D, et al. Improved molecular diagnosis of facioscapulohumeral muscular dystrophy (FSHD): validation of the differential double digestion for FSHD. J. Med. Genet 1997; 34: 476–9.

van Deutekom JC, Bakker E, Lemmers RJ, van der Wielen MJ, Bik E, Hofker MH, et al. Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. Hum Mol Genet 1996; 5: 1997–2003.

van Overveld PG, Lemmers RJ, Deidda G, Sandkuijl L, Padberg GW, Frants RR, van der Maarel SM. Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum Mol Genet 2000; 9: 2879–84.

Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet 1992; 2: 26–30.

Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics 2000; 56: 645–6.

Zatz M, Marie SK, Cerqueira A, Vainzof M, Pavanello RC, Passos-Bueno MR. The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. Am J Med Genet 1998; 77: 155–61.

Appendix 2

Referring Hospital:
Referring physician:
Date:
Patient's code: *Initials:*
Date of birth: *Sex:* □ M □ F
Family geographical origin:

Pregnancy

Active fetal movements	□ Yes	□ Reduced	□ Normal
	□ No		
Fetal cardiac beats	□ Normal	□ Altered	
Placental dysfunction	□ Yes	□ No	weight:
Diseases/infections	□ Yes	□ No	
Ultrasound	□ Yes	□ No	
	Routine ultrasound	□ Normal	□ Altered
	Morphological ultrasound	□ Normal	□ Altered
Therapy during pregnancy	□ Yes	Drugs:	
	□ No		

Birth

Partum	□ Eutocic	□ Dystocic	week:
Fetal position	□ Normal	□ Podalic	
Weight	Lenght	Apgar	
Revived	□ Yes	□ No	
Cianosis	□ Yes	□ No	
Mechanical ventilation	□ Yes	□ No	
Clubfoot	□ Yes	□ No	
Neonatal jaundice	□ Yes	□ No	

Perinatal period/ months after

Breastfeeding	□ Yes	□ No	
Breast Suction	□ Vigorous	□ Normal	□ Reduced
Facial nerve palsy's diagnosis	□ Yes	□ No	
Moebius syndrome's diagnosis	□ Yes	□ No	
Hip dysplasia	□ Yes	□ No	
Upper limbs	□ Normal	□ Altered	
Facial hypomimia	□ Yes	□ No	
Floppy	□ Yes	□ No	

Psychomotor development

Sitting alone: age (months)
Self-standing: age (months)
Walk alone: age (months)
Social smile □ Yes age (months) □ No

Familiarity

□ Yes □ Maternal line □ Paternal line □ Brothers/sisters
□ No

Data collection

Mechanical ventilation	□ Yes (age:)	□ No
Vision problems	□ Yes (age:)	□ No
Hypoacusia	□ Yes	□ No
Academic success	□ Yes	□ No
Cognitive impairment	□ Yes	□ No
Loss of walking	□ Yes (age:)	□ No

First year of life

Neurological examination

Hypotonia	□ Yes	□ No
Facial mimicry	□ Normal	□ Absent
Antigravity limb movements	□ Possible	□ Impossible
Hospital admission	□ Yes Clinical report:	□ No

After the first year of life

Neurological examination

See *FSHD clinical form*

Lagophtalmos	□ Yes	□ No
Gowers	□ Yes	□ No
Waddling gate	□ Yes	□ No
Steppage	□ Yes	□ No
Scapular winging	□ Yes	□ No

Instrumental tests

Blood tests	CK UI/L		
Electrophysiological tests	EMG/ENG	□ Yes □ No	Results:
	BAERs	□ Yes □ No	Results:
	EEG	□ Yes □ No	Results:
Neuroradiological tests	Cerebral MRI	□ Yes □ No	Results:
	Muscular MRI	□ Yes □ No	Results:
Eye examination	Fundus oculi	□ Yes □ No	Results:
	ERG	□ Yes □ No	Results:
	Fluorescein angiography	□ Yes □ No	Results:
Ear examination	Audiometry	□ Yes □ No	Results:

Appendix 3

GGTACCAGCAGGTGGGCCGCCTACTGCGCACGCGCGGGTTTGCGGGCAGCCGCCTGGGCTGTGGGAGCAGCCCGG
GCAGAGCTCTCCTGCCTCTCCACCAGCCCACCCCGCCGCCTGACCGCCCCCTCCCCACCCCCACCCCCCACCCCC
GGAAAACGCGTCGTCCCCTGGGCTGGGTGGAGACCCCCGTCCCGCGAAACACCGGGCCCCGCGCAGCGTCCGGGC
CTGACACCGCTCCGGCGGCTCGCCTCCTCTGCGCCCCCGCGCCACCGTCGCCCGCCCGCCCGGGCCCCTGCAGCC
GCCCAGGTGCCAGCACGGAGCGCCTGGCGGCGGAACGCAGACCCCAGGCCCGGCGCACACCGGGGACGCTGAGCG
TTCCAGGCGGGAGGGAAGGCGGGCAGAGATGGAGAGAGGAACGGGAGACCTAGAGGGGCGGAAGGACGGGCGGAG
GGACGTTAGGAGGGAGGGAGGGAGGCAGGGAGGCAGGGAGGAACGGAGGGAAAGACAGAGCGACGCAGGGACTGG
GGGCGGGCGGGAGGGAGCCGGGGACGGACGGGGGGAGGAAGGCAGGGAGGAAAAGCGGTCCTCGGCCTCCGGGAG
TAGCGGGACCCCCGCCCTCCGGGAAAACGGTCAGCGTCCGGCGCGGGCTGAGGGCTGGGCCCACAGCCGCCGCGC
CGGCCGGCGCGGCACCCATTCGCCCCGGTTCCGTGGCCCAGGGAGTGGGCGGTTTCCTCCGGGACAAAAGACCGG
GACTCGGGTTGCCGTCGGGTTTTCACCCGCGCGGTTCACAGACCGCACATCCCCAGGCTGAGCCCTGCAACGCGG
CGCGAGGCCGACAGCCCCGGCCACGGAGGAGCCACACGCAGGACGACGGAGGCGTGATTTTGGTTTCCGCGTGGC
TTTGCCCTCCGCAAGGCGGCCTGTTGCTCACGTCTCTCCGGCCCCCGAAAGGCTGGCCATGCCGACTGTTTGCTC
CCGGAGCTCTGCGGGCACCCGGAAACATGCAGGGAAGGGTGCAAGCCCGGCATGGTGCCTTCGCTCTCCTTGCCA
GGTTCCAAACCGGCCACACTGCAGACTCCCCACGTTGCCGCACGCGGGAATCCATCGTCAGGCCATCACGCCGGG
GAGGCATCTCCTCTCTGGGGTCTCGCTCTGGTCTTCTACGTGGAAATGAACGAGAGCCACACGCCTGCGTGTGCG
AGACCGTCCCGGCAACGGCGACGCCCACAGGCATTGCCTCCTTCACGGAGAGAGGGCCTGGCACACTCAAGACTC
CCACGGAGGTTCAGTTCCACACTCCCCTCCACCCTCCCAGGCTGGTTTCTCCCTGCTGCCGACGCGTGGGAGCCC
AGAGAGCGGCTTCCCGTTCCCGCGGGATCCCTGGAGAGGTCCGGAGAGCCGGCCCCCGAAACGCGCCCCCCTCCC
CCCTCCCCCCTCTCCCCGTTCCTCTTCGTCTCTCCGGCCCCACCACCACCACCGCCACCACGCCCTCCCCCACCA
CCCCCCCCCCCCCCACCACCACCACCACCACCCCGCCGGCC**GG**CCCCAGGCCTCGACGCCCTGGGTCCCTTCCGG
GGTGGGGCGGGCTGTCCCAGGGGGGCTCACCGCCATTCATGAAGGGGTGGAGCCTGCCTGCCTGTGGGCCTTTAC
AAGGGCGGCTGGCTGGGTGGCTGGCTGTCCGGGCAGGCCCCCTGGCTGCACCTGCCGCAGTGCACAGTCCGGCTG
AGGTGCACGGGAGCCCGCCGGCCTCTCTCTGCCCGCGTCCGTCCGTGAAATTCCGGCCGGGGCTCACCGCGATGG
CCCTCCCGACACCCTCGGACAGCACCCTCCCCGCGGAAGCCCGGGGACGAGGACGGCGACGGAGACTCGTTTGGA
CCCCGAGCCAAAGCGAGGCCCTGCGAGCCTGCTTTGAGCGGAACCCGTACCCGGGCATCGCCACCAGAGAACGGC
TGGCCCAGGCCATCGGCATTCCGGAGCCCAGGGTCCAGATTTGGTTTCAGAATGAGAGGTCACGCCAGCTGAGGC
AGCACCGGCGGGAATCTCGGCCCTGGCCCGGGAGACGCGGCCCGCCAGAAGGCCGGCGAAAGCGGACCGCCGTCA
CCGGATCCCAGACCGCCCTGCTCCTCCGAGCCTTTGAGAAGGATCGCTTTCCAGGCATCGCCGCCCGGGAGGAGC
TGGCCAGAGAGACGGGCCTCCCGGAGTCCAGGATTCAGATCTGGTTTCAGAATCGAAGGGCCAGGCACCCGGGAC
AGGGTGGCAGGGCGCCCGCGCAGGCAGGCGGCCTGTGCAGCGCGGCCCCCGGCGGGGGTCACCCTGCTCCCTCGT
GGGTCGCCTTCGCCCACACCGGCGCGTGGGGAACGGGGCTTCCCGCACCCCACGTGCCCTGCGCGCCTGGGGCTC
TCCCACAGGGGGCTTTCGTGAGCCAGGCAGCGAGGGCCGCCCCCGCGCTGCAGCCCAGCCAGGCCGCGCCGGCAG
AGGGGATCTCCCAACCTGCCCCGGCGCGCGGGGATTTGGCCTACGCCGCCCCGGCTCCTCCGGACGGGGCGCTCT
CCCACCCTCAGGCTCCTCGGTGGCCTCCGCACCCGGGCAAAAGCCGGGAGGACCGGGACCCGCAGCGCGACGGCC
TGCCGGGCCCCTGCGCGGTGGCACAGCCTGGGCCCGCTCAAGCGGGGCCGCAGGGCCAAGGGGTGCTTGCGCCAC
CCACGTCCCAGGGGAGTCCGTGGTGGGGCTGGGGCCGGGGTCCCCAGGTCGCCGGGGCGGCGTGGGAACCCCAAG
CCGGGGCAGCTCCACCTCCCCAGCCCGCGCCCCCGGACGCCTCCGCCTCCGCGCGGCAGGGGCAGATGCAAGGCA
TCCCGGCGCCCTCCCAGGCGCTCCAGGAGCCGGCGCCCTGGTCTGCACTCCCCTGCGGCCTGCTGCTGGATGAGC
TCCTGGCGAGCCCGGAGTTTCTGCAGCAGGCGCAACCTCTCCTAGAAACGGAGGCCCCGGGGGAGCTGGAGGCCT
CGGAAGAGGCCGCCTCGCTGGAAGCACCCCTCAGCGAGGAAGAATACCGGGCTCTGCTGGAGGAGCTTTAGGACG
CGGGGTTGGGACGGGGTCGGGTGGTTCGGGGCAGGGCGGTGGCCTCTCTTTCGCGGGGAACACCTGGCTGGCTAC
GGAGGGGCGTGTCTCCGCCCCGCCCCCTCCACCGGGCTGACCGGCCTGGGATTCCTGCCTTCTAGGTCTAGGCCC
GGTGAGAGACTCCACACCGCGGAGAACTGCCATTCTTTCCTGGGCATCCCGGGGATCCCAGAGCCGGCCCAGGTA
CC

ANALYZED SEQUENCES OF D4Z4 REPEAT

AGATCT= BglII

GGCCGGCC = FseI (methylation sensitive)

TACGTG = BsaAI (methylation sensitive)

GGTACC = KpnI

CGGCCG = EagI (methylation sensitive)

CCATTCATGAAGGGGTGGAGCCTGCCT = DBE

Yellow: PRIMERS DR1

Blue: PRIMERS DR2

Dark green: PRIMERS DR3

Underlined: DBE-R

References

- Agresti A, Rainaldi G, Lobbiani A, Magnani I, Di Lernia R, Meneveri R, Siccardi A G, Ginelli E. Chromosomal location by in situ hybridisation of the human Sau3A family of DNA repeats. Hum Genet. 1987;75:326-332.
- Agresti A, Meneveri R, Siccardi AG, Marozzi A, Corneo G, Gaudi S, Ginelli E. Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family). J Mol Biol. 1989;205(4):625-631.
- Alexiadis V, Ballestas ME, Sanchez C, Winokur S, Vedanarayanan V, Warren M, Ehrlich M. RNAPol-ChIP analysis of transcription from FSHD-linked tandem repeats and satellite DNA. Biochim. Biophys. Acta. 2007;1769:29-40.
- Becker PE. Dystrophia Musculorum Progressiva, Eine genetiche und klinische Untersuchung der Muskeldystrophien. Georg Thieme Verlag, Stuttgart. 1953.
- Beckers MC, Gabriels J, van der SM, De Vriese A, Frants RR, Collen D, Belayew A. Active genes in junk DNA? Characterization of DUX genes embedded within 3.3kb repente elements. Gene 2001;264:51-57.
- Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, van der Wielen M, Rasmussen K, Frants RR: The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter**.** Muscle Nerve Suppl 1995**;**S39-44.
- Bailey RO, Marzulo DC, Hans MB. Infantile facioscapulohumeral muscular dystrophy: new observations. Acta Neurol Scand 1986;74:51-8.
- Becker PE. Dystrophia musculorum progressiva. Eine genetische und klinische untersuchung der muskeldystrophien. Stuttgart: Georg Thieme Verlag, 1953.
- Block GJ, Ptek LM, Narayanan D, Amell AM, Moore JM, Rabaia NA, Tyler A, van der Maarel SM, Tawil R, Filippova GN, Miller DG. Asymmetric bidirectional transcription from the FSHD-causing D4Z4 array modulates DUX4 production. Plos One 2012;7(4):e35532.
- Brooke MH. A clinician’s view on neuromuscular disease. Williams and Wilkins Co., Baltimore. 1977.
- Brouwer OF, Padberg GW, Wijmenga C, Frants RR. Facioscapulohumeral muscular dystrophy in early childhood. Arch Neurol. 1994;51(4):387-394.
- Buckingham M and Relaix F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. Ann Rev Cell Dev Biol. 2007;23:645-673.

- Butz M, Koch MC, Müller-Felber W, Lemmers RJ, van der Maarel SM, Schreiber H. Facioscapulohumeral muscular dystrophy. Phenotype-genotype correlation in patients with borderline D4Z4 repeat numbers. J Neurol. 2003 Aug;250(8):932-7.
- Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, & Gabellini D. A long ncRNA links copy number variation to a Polycomb/Trithorax epigenetic switch in FSHD muscular dystrophy. Cell 2012;149:819-831.
- Casasnovas C, Cano LM, Albertí A, Céspedes M, Rigo, G. Charcot-Marie-tooth disease. Foot Ankle Spec. 2008;1:350-354.
- Caruso N, Herberth B, Bartoli M, Puppo F, Dumonceaux J, Zimmermann A, Denadai S, Lebossé M, Roche S, Geng L, Magdinier F, Attarian S, Bernard R, Maina F, Levy N, Helmbacher F. Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy**.** PLoS Genet. 2013 Jun;9(6):e1003550.
- Chahwan R, Wontakal SN, Roa S. The multidimensional nature of epigenetic information and its role in disease. Discov Med. 2011;11:233-243.
- Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, Armour JA, Hewitt JE. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2007 Aug;81(2):264-79.
- Chung CS, Morton NE. Discrimination of genetic entities in muscular dystrophy. Am J Hum Genet 1959;11:339-59.
- D'Antona G, Brocca L, Pansarasa O, Rinaldi C, Tupler R, Bottinelli R. () Structural and functional alterations of muscle fibres in the novel mouse model of facioscapulohumeral muscular dystrophy**.** J Physiol 2007;584:997-1009.
- Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, Felicetti L. 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. 1995;3(3):155-67.
- de Greef JC, Lemmers RJ, van Engelen BG, Sacconi S, Venance SL, Frants RR, Tawil R, van der Maarel SM. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum Mutat. 2009 Oct;30(10):1449-59.
- de Greef JC, Lemmers RJ, Camaño P, Day JW, Sacconi S, Dunand M, van Engelen BG, Kiuru-Enari S, Padberg GW, Rosa AL, Desnuelle C, Spuler S, Tarnopolsky M,

Venance SL, Frants RR, van der Maarel SM, Tawil R. Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology. 2010 Oct 26;75(17):1548-54.

- Deidda G, Cacurri S, Piazzo N, Felicetti L. Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). JMed Genet. 1996;33(5):361-365.
- Deschamps J and Meijlink F Mammalian homeobox genes in normal development and neoplasia. Crit Rev Oncog. 1992;3:117-173.
- Ding H, Beckers MC, Plaisance S, Marynen P, Collen D, Belayew A. Characterization of a double homeodomain protein (DUX1) encoder by a cDNA homologous to 3.3kb dispersed repeat elements. Hum Mol Genet. 1998;(7):1681-1694.
- Dixit M, Ansseau E, Tassin A, Winokur S, Shi R, Qian H, Sauvage S, Mattéotti C, van Acker AM, Leo O, Figlewicz D, Barro M, Laoudj-Chenivesse D, Belayew A, Coppée F, Chen YW. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc Natl Acad Sci U S A. 2007 Nov 13;104(46):18157-62.
- Kim JD, Hinz AK, Bergmann A, Huang JM, Ovcharenko I, Stubbs L, Kim J.Genome Res. Identification of clustered YY1 binding sites in imprinting control regions. 2006 July;16(7):901-11. Epub 2006 Jun 7.
- Dorobek M, Kabzińska D. A severe case of facioscapulohumeral muscular dystrophy (**FSHD**) with some uncommon clinical features and a short 4q35 fragment. Eur J Paediatr Neurol. 2004;8(6):313-6.
- Duchenne G. Album de photographies pathologiques. 1862.
- Duchenne G. Revue photographique des hopitaux de Paris.1869.
- Engel AG, Franzini-Armstrong C. Miology. McGraw-Hill 2004 (Third Edition).
- Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. Curr Opin Cell Biol. 2003;15:172-183.
- Fitzsimons RB. Facioscapulohumeral muscular dystrophy. Curr Opin Neurol. 1999;12(5):501-511.
- Flanigan, K. M. Facioscapulohumeral muscular dystrophy and scapulohumeral syndrome. In: Engel AG, Franzini-Armstrong C (eds) Myology. New York: McGraw-Hill. 2004;123–34.
- Funakoshi M, Goto K, Arahata K.Epilepsy and mental retardation in a subset of early onset 4q35-facioscapulohumeral muscular dystrophy. Neurology 1998;50:1791-94.

- Gabellini D, Green MR, Tupler R. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell. 2002 Aug 9;110(3):339-48.
- Gabriels J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM,Padberg GW, Frants RR, Hewitt JE, Collen D, Belayew A. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene. 1999;236(1):25–32.
- Galluzzi G, Deidda G, Cacurri S, Colantoni L, Piazzo N, Vigneti E, Ricci E, Servidei S, Merico B, Pachì A, Brambati B, Mangiola F, Tonali P, Felicetti L. Molecular analysis of 4q35 rearrangements in fascioscapulohumeral muscular dystrophy (FSHD): application to family studies for a correct genetic advice and a reliable prenatal diagnosis of the disease. Neuromuscul Disord. 1999 May;9(3):190-8.
- Griggs RC, Tawil R, Storvick D, Mendell JR, Altherr MR. Genetics of facioscapulohumeral muscular dystrophy: new mutations in sporadic cases. Neurology 1993;43:2369-2372.
- Hanel ML, Wuebbles RD, & Jones PL. Muscular dystrophy candidate gene FRG1 is critical for muscle development. Dev Dyn 2009;238:1502-1512.
- Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM ().CTCF mediates methylation-sensitive enhancer blocking activity at the H19/Igf2
- locus. Nature. 2000;405:486-489.
- Hartweck LM, Anderson LJ, Lemmers RJ, Dandapat A, Toso EA, Dalton JC, Tawil R, Day JW, van der Maarel SM, Kyba M. A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology. 2013 Jan 22;80(4):392-9.
- Jacobsen SJ, Diala ES, Dorsey BV, Rising MB, Graveline R, Falls K, Schultz P, Hogan C, Rediker K, D'Amico C, et al. A clinically homogeneous group of families with facioscapulohumeral (Landouzy-Déjérine) muscular dystrophy: linkage analysis of six autosomes. Am J Hum Genet. 1990 Sep;47(3):376-88.
- Jardine PE, Koch MC, Lunt PW, Maynard J, Bathke KD, Harper PS, Upadhyaya M. De novo facioscapulohumeral muscular dystrophy defined by DNA probe p13E- 11 (D4F104S1). Arch Dis Child. 1994;71(3):221-227.
- Jiang G, Yang F, Van Overveld PG, Vedanarayanan V, van der Maarel SM, Ehrlich M. Testing the position-effect variegation hypothesis for facioscapulohumeral

muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Hum Mol Genet. 2003;12:2909-2921.

- Jones TI, Chen JC, Rahimov F, Homma S, Arashiro P, Beermann ML, King OD, Miller JB, Kunkel LM, Emerson CP Jr, Wagner KR, Jones PL. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Hum Mol Genet. 2012 Oct 15;21(20):4419-30. Epub 2012 Jul 13.
- Kanagawa M, Toda T. The genetic and molecular basis of muscular dystrophy: roles of cell-matrix linkage in the pathogenesis. J Hum Genet. 2006;11:915-926.
- Kim J, Kollhoff A, Bergmann A, Stubbs L. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. Hum Mol Genet. 2003;12:233-245.
- Kondo T, Bobek MP, Kuick R, Lamb B, Zhu X, Narayan A, Bourc'his D, Viegas-Pequignot E, Ehrlich M, Hanash SM. Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum Mol Genet. 2000;9:597-604.
- Krom YD, Thijssen PE, Young JM, den Hamer, B, Balog, J, Yao Z, Maves L, Snider L, Knopp P, Zammit PS, Rijkers T, van Engelen BG, Padberg GW, Frants RR, Tawil R, Tapscott SJ & van der Maarel S.M. Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. PLoS Genet 2013;9(4):e1003415.
- Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Matteotti C, Arias C, Corona ED, Nunez NG, Leo O, Wattiez R, Figlewicz D, Laoudj-Chenivesse D, Belayew A, Coppee F, Rosa AL. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 2007;17:611-623.
- Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, van Deutekom JC, Francis F, Sharpe PT, Hofker M, et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet. 1994 Aug;3(8);1287-95.
- Lamperti C, Fabbri G, Vercelli L, D'Amico R, Frusciante R, Bonifazi E, Fiorillo C, Borsato C, Cao M, Servida M, et al. A standardized clinical evaluation of patients affected by Facioscapulohumeral Muscular Dystrophy: the FSHD disability score. Muscle Nerve 2010;42:213-217.

- Lana E, Mégarbané A, Tourrière H, Sarda P, Lefranc G, Claustres M, De Sario A. Eur J Hum Genet. DNA replication is altered in Immunodeficiency Centromeric instability Facial anomalies (ICF) cells carrying DNMT3B mutations. 2012 Oct;20(10):1044-50.
- Landouzy L, Dejerine J. De la myopathie atrophique progressive. Rev Med Francaise 1885;5:81.
- Lecky BR, MacKenzie JM, Read AP, Wilcox DE. X-linked and FSH dystrophies in one family. Neuromuscul Disord. 1991;4:275-278.
- Lemmers RJLF, van der Maarel SM, van Deutekom JCT, van der Wielen MJR, Deidda G, Dauwerse HG, Hewitt J, Hofker M, Bakker E, Padberg GW, Frants RR(Interand intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. Hum Mol Genet. 1998;7(8):1207-1214.
- Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet. 2002 Oct;32(2):235-6.
- Lemmers RJ, Van Overveld PG, Sandkuijl LA, Vrieling H, Padberg GW, Frants RR, van der Maarel SM. Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy.Am J Hum Genet. 2004 Jul;75(1):44-53 (a).
- Lemmers RJ, Wohlgemuth M, Frants RR, Padberg GW, Morava E, van der Maarel SM. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2004 Dec;75(6):1124-30 (b).
- Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, de Knijff P, Padberg GW, Frants RR, van der Maarel SM. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2007 Nov;81(5):884-94.
- Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, Snider L, Straasheijm KR, van Ommen GJ, Padberg GW, Miller DG, Tapscott SJ, Tawil R, Frants RR, van der Maarel SM. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 2010 Sep 24;329(5999):1650-3.
- Lemmers RJ, Tawil R, Petek, LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR, Krom YD, Klooster R, Sun Y, den Dunnen JT, Helmer Q, Donlin-Smith CM, Padberg GW, van Engelen BG, de Greef J,

Aartsma-Rus AM, Frants RR, de Visser M, Desnuelle C, Sacconi S, Filippova GN, Bert B, Bamshad MJ, Tapscott SJ, Miller DG, van der Maarel S. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet 2012;44:1370-74.

- Liu Q, Jones TI, Tang V, Brieher WM, & Jones PL. (). Facioscapulohumeral muscular dystrophy region gene-1 (FRG-1) is an actin-bundling protein associated with muscle-attachment sites. J Cell Sci 2010;123:1116-1123.
- Lunt PW, Compston DA, Harper PS. Estimation of age dependent penetrance in facioscapulohumeral muscular dystrophy by minimising ascertainment bias. J Med Genet. 1989 Dec;26(12):755-60.
- Lunt PW, Jardine PE, Koch MC, Maynard J, Osborn M, Williams M, Harper PS, Upadhyaya M. 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. 1995 May;4(5):951-8 (a).
- Lunt PW, Jardine PE, Koch M, Maynard J, Osborn M, Williams M, Harper PS, Upadhyaya M. Phenotypic-genotypic correlation will assist genetic counseling in 4q35-facioscapulohumeral muscular dystrophy. Muscle Nerve. 1995;2:S103-9 (b).
- Lyle R, Wright TJ, Clark LN, Hewitt JE. 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. 1995 Aug 10;28(3):389-97.
- Matsamura T, Goto K, Yamanaka G, Lee JH, Zhang C, Hayashi YK, Arahata K: Chromosome 4q;10q traslocations; comparison with different ethnic populations and FSHD patients. Neurol 2002, Aug 20;2:7.
- Meneveri R, Agresti A, Marozzi A, Saccone S, Rocchi M, Archidiacono N, Corneo G, Della VG, Ginelli E. Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. Gene. 1993;123(2):227-234.
- Miura K, Kumagai T, Matsumoto A, Iriyama E, Watanabe K, Goto K, Arahata K. Two cases of chromosome 4q35-linked early onset facioscapulohumeral muscular dystrophy with mental retardation and epilepsy. Neuropediatrics. 1998 Oct;29(5):239-41.

- Mostacciuolo ML, Pastorello E, Vazza G, Miorin M, Angelini C, Tomelleri G, Galluzzi G, Trevisan CP. Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. Clin Genet. 2009 Jun;75(6):550-5.
- Nakagawa M, Matsuzaki T, Higuchi I, Fukunaga H, Inui T, Nagamitsu S, Yamada H, Arimura K, Osame M. Facioscapulohumeral muscular dystrophy: clinical diversity and genetic abnormalities in Japanese patients. Intern Med. 1997;36(5):333-339.
- Neguembor MV, Xynos A, Onorati MC, Caccia R, Bortolanza S, Godio C, Pistoni M, Corona DF, Schotta G, Gabellini D. FSHD muscular dystrophy region gene 1 binds Suv4-20h1 histone methyltransferase and impairs myogenesis. J Mol Cell Biol. 2013 Oct;5(5):294-307
- Okinaga A, Matsuoka T, Umeda J, Yanagihara I, Inui K, Nagai T, Okada S. Early-onset facioscapulohumeral muscular dystrophy: two case reports. Brain Dev. 1997 Dec;19(8):563-7.
- Ottaviani A, Schluth-Bolard C, Rival-Gervier S, Boussouar A, Rondier D, Foerster AM, Morere J, Bauwens S, Gazzo S, Callet-Bauchu E, Gilson E, Magdinier Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. EMBO J. 2009 Aug 19;28(16):2428-36.
- Osborne RJ, Welle S, Venance SL, Thornton CA, Tawil R. Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. Neurology. 2007 Feb 20;68(8):569-77.
- Padberg GW. Facioscapulohumeral disease. MD thesis, University of Leiden, 1982.
- Padberg GW, Lunt PW, Koch M, Fardeau M. Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul Disord. 1991;1(4):231-4.
- Pearson. Two new pedigrees of muscular dystrophy. Ann Eugen 1933;5:179-191.
- Pistoni M, Shiue L, Cline MS, Bortolanza S, Neguembor MV, Xynos A, Ares M. Jr, & Gabellini D. Rbfox1 downregulation and altered calpain 3 splicing by FRG1 in a mouse model of Facioscapulohumeral muscular dystrophy (FSHD). PLoS Genet. 2013;9(1):e1003186.
- Reilich P, Schramm N, Schoser B, Schneiderat P, Strigl-Pill N, Müller-Höcker J, Kress W, Ferbert A, Rudnik-Schöneborn S, Noth J, Lochmüller H,Weis J, Walter MC. Facioscapulohumeral muscular dystrophy presenting with unusual phenotypes

and atypical morphological features of vacuolar myopathy. J Neurol. 2010 ;7:1108-1118.

- Ricci E, Galluzzi G, Deidda G, Cacurri S, Colantoni L, Merico B, Piazzo N, Servidei S, Vigneti E, Pasceri V, Silvestri G, et al. 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. 1999;45(6):751-757.
- Robertson KD, Wolffe AP. DNA methylation in health and disease. Nat Rev Genet. 2000 Oct;1(1):11-9. Review.
- Sakellariou P, Kekou K, Fryssira H, Sofocleous C, Manta P, Panousopoulou A, Gounaris K, Kanavakis E. Mutation spectrum and phenotypic manifestation in FSHD Greek patients. Neuromuscul Disord. 2012 Apr;22(4):339-49.
- Sarfarazi M, Upadhyaya M, Padberg G, Pericak-Vance M, Siddique T, Lucotte G, Lunt P. An exclusion map for facioscapulohumeral (Landouzy-Déjérine) disease. J Med Genet. 1989 Aug;26(8):481-4.
- Sekimata M, Murakami-Sekimata A, Homma Y. CpG methylation prevents YY1-mediated transcriptional activation of the vimentin promoter. Biochem Biophys Res Commun. 2011 Nov 4;414(4):767-72.
- Sancisi V, Germinario E, Esposito A, Morini E, Peron S, Moggio M, Tomelleri G, Danieli-Betto D, Tupler R. Altered Tnnt3 characterizes selective weakness of fast fibers in mice overexpressing FSHD region gene 1 (FRG1). Am J Physiol Regul Integr Comp Physiol. 2014 Jan 15;306(2):R124-37.
- Scionti I, Fabbri G, Fiorillo C, Ricci G, Greco F, D'Amico R, Termanini A, Vercelli L, Tomelleri G, Cao M, Santoro L, Percesepe A, Tupler R. Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. J Med Genet. 2012 Mar;49(3):171-8 (a).
- Scionti I, Greco F, Ricci G, Govi M, Arashiro P, Vercelli L, Berardinelli A, Angelini C, Antonini G, Cao M, Di Muzio A, Moggio M, Morandi L, Ricci E, Rodolico C, Ruggiero L, Santoro L, Siciliano G, Tomelleri G, Trevisan CP, Galluzzi G, Wright W, Zatz M, Tupler R. Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. Am J Hum Genet. 2012 Apr 6;90(4):628-35 (b).

- Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der Maarel SM, Tapscott SJ, Miller DG. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. 2010 Oct 28;6(10):e1001181.
- Takahashi M, Asai N, Iwashita T, Murakami H, Ito S. Mechanisms of development of multiple endocrine neoplasia type 2 and Hirschsprung's disease by ret mutations. Recent Results Cancer Res.1998;154:229-236.
- Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, Figlewicz D. Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. Ann Neurol. 1996 Jun;39(6):744-8.
- Tawil R, van der Maarel S, Padberg GW, van Engelen BG. 171st ENMC international workshop: Standards of care and management of facioscapulohumeral muscular dystrophy. Neuromuscul Disord. 2010 Jul;20(7):471-5.
- Tyler FH, Stephens FE. Studies in disorders of muscle: II: clinical manifestations and hineritance of facioscapulohumeral dystrophy in a large family. Ann Int Med. 1950;32:640-660.
- Tonini MM, Passos-Bueno MR, Cerqueira A, Matioli SR, Pavanello R, Zatz M. Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromuscul Disord. 2004 Jan;14(1):33-8.
- Tsien F, Sun B, Hopkins NE, Vedanarayanan V, Figlewicz D, Winokur S, Ehrlich M Methylation of the FSHD syndrome-linked subtelomeric repeat in normal and FSHD cell cultures and tissues. Mol Genet Metab. 2001;74:322-331.
- Tupler R, Berardinelli A, Barbierato L, Frants R, Hewitt JE, Lanzi G, Maraschio P, Tiepolo L. Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. J Med Genet. 1996 May;33(5):366-70.
- Tupler R, Barbierato L, Memmi M, Sewry CA, De Grandis D, Maraschio P, Tiepolo L, Ferlini A. Identical de novo mutation at the D4F104S1 locus in monozygotic male twins affected by facioscapulohumeral muscular dystrophy (FSHD) with different clinical expression. J Med Genet. 1998 Sep;35(9):778-83.
- Tyler FH, Stephens FE: Studies in disorders of muscle: II. Clinical manifestations and inheritance of facioscapulohumeral dystrophy in a large family. Ann Intern Med 1950; 32:640-660.

- Upadhyaya M, Lunt PW, Sarfarazi M, Broadhead W, Daniels J, Owen M, Harper PS. DNA marker applicable to presymptomatic and prenatal diagnosis of facioscapulohumeral disease. Lancet. 1990 Nov 24;336(8726):1320-1.
- Upadhyaya M, Lunt PW, Sarfarazi M, Broadhead W, Daniels J, Owen M, Harper PS. A closely linked DNA marker for facioscapulohumeral disease on chromosome 4q. J Med Genet. 1991 Oct;28(10):665-71.
- Upadhyaya M, Jardine P, Maynard J, Farnham J, Sarfarazi M, Wijmenga C, Hewitt JE, Frants R, Harper PS, Lunt PW. Molecular analysis of British facioscapulohumeral dystrophy families for 4q DNA rearrangements. Hum Mol Genet. 1993 Jul;2(7):981-7.
- Upadhyaya M, Maynard J, Rogers MT, Lunt PW, Jardine P, Ravine D, Harper PS. Improved molecular diagnosis of facioscapulohumeral muscular dystrophy (FSHD): validation of the differential double digestion for FSHD. J Med Genet. 1997 Jun;34(6):476-9.
- van Deutekom JC, Wijmenga C, van Tienhoven EA, Gruter AM, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet. 1993 Dec;2(12):2037-42.
- van Deutekom JC, Bakker E, Lemmers RJ, van der Wielen MJ, Bik E, Hofker MH, Padberg GW, Frants RR. 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. 1996 Dec;5(12):1997-2003 (a).
- van Deutekom JC, Lemmers RJ, Grewal PK, van Geel M, Romberg S, Dauwerse HG, Wright TJ, Padberg GW, Hofker MH, Hewitt JE, Frants RR. Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. Hum Mol Genet. 1996 May;5(5):581-90 (b).
- van Geel M, Dickson MC, Beck AF, Bolland DJ, Frants RR, van der Maarel SM, de Jong PJ, Hewitt JE. Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics. 2002 Feb;79(2):210-7.
- van Overveld PGM, Lemmers RJ, Deidda G, Sandkuijl L, Padberg GW, Frants RR, van der Maarel SM. Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum. Mol. Genet. 2000;9:2879-2884

- van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet. 2003 Dec;35(4):315-7.
- van Overveld PG, Enthoven L, Ricci E, Rossi M, Felicetti L, Jeanpierre M, Winokur ST, Frants RR, Padberg GW, van der Maarel SM. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. AnnNeurol. 2005;58:569-576.
- Vitelli F, Villanova M, Malandrini A, Bruttini M, Piccini M, Merlini L, Guazzi G, Renieri A. Inheritance of a 38-kb fragment in apparently sporadic facioscapulohumeral muscular dystrophy. Muscle Nerve. 1999;22:1437-1441.
- Weiffenbach B, Bagley R, Falls K, Hyser C, Storvick D, Jacobsen SJ, Schultz P, Mendell J, Willems van Dijk K, Milner EC, et al. Linkage analyses of five chromosome 4 markers localizes the facioscapulohumeral muscular dystrophy (FSHD) gene to distal 4q35. Am J Hum Genet 1992;51:416-423.
- Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, et al. The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Res 1994;2: 225-234.
- Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, Padberg GW. Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet. 1990 Sep 15;336(8716):651-3.
- Wijmenga C, Padberg GW, Moerer P, Wiegant J, Liem L, Brouwer OF, Milner EC, Weber JL, van Ommen GB, Sandkuyl LA, et al. Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics. 1991 Apr;9(4):570-5.
- Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet. 1992 Sep;2(1):26-30.
- Wijmenga C, van Deutekom JC, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR. Pulsed-field gel electrophoresis of the D4F104S1 locus reveals the

size and the parental origin of the facioscapulohumeral muscular dystrophy (FSHD)-associated deletions. Genomics. 1994 Jan;1:19(1):21-6.

- Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, Tapscott SJ, van der Maarel SM, Hayashi Y, Flanigan KM. Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Hum Mol Genet. 2003 Nov 15;12(22):2895-907.
- Zatz M, Marie SK, Passos-Bueno MR, Vainzof M, Campiotto S, Cerqueira A, Wijmenga C, Padberg G, Frants R. High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. Am J Hum Genet. 1995 Jan;56(1):99-105.
- Zatz M, Marie SK, Cerqueira A, Vainzof M, Pavanello RC, Passos-Bueno MR. The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. Am J Med Genet. 1998 May 1;77(2):155-61.
- Zeng W, de Greef JC, Chen YY, Chien R, Kong X, Gregson HC, Winokur ST, Pyle A, Robertson KD, Schmiesing JA, Kimonis VE, Balog J, Frants RR, Ball AR Jr, Lock LF, Donovan PJ, van der Maarel SM, Yokomori K. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet. 2009 Jul;5(7):e1000559.
- Zhang XY, Loflin PT, Gehrke CW, Andrews PA, Ehrlich M (). Hypermethylation of human DNA sequences in embryonal carcinoma cells and somatic tissues but not in sperm. Nucleic Acids Res. 1987;15(22):9429-9449.
- Yamanaka G, Goto K, Hayashi YK, Miyajima T, Hoshika A, and Arahata K. Clinical and genetical features of Japanese early-onset facioscapulohumeral muscular dystrophy. No To Hattatsu. 2002;34(4):318-324.
- Yao Y, Yang M, & Seto E. 2001. Regulation of transcription factor YY1 by acetylation and deacetylation. Mol Cell Biol 2001;21:5979-5991.
- Young JM, Whiddon JL, Yao Z, Kasinathan , Sinder L, Geng LN, Balog J, Tawil R, van der Maarel SM, Tapscott SJ. DUX4 binding to retroelements creates promoters thata are active in FHSD muscle and testis. Plos Genet 2013;9(11):e1003947.

Time for some acknowledgments

The long list of people I would like to thank must start from my mentor and supervisor, Professor Tupler. Thank You for this exciting, nearly four years-long PhD journey. Thank You for the invitation to come to University of Modena and Reggio Emilia, and thank You for the invitation to stay.

Many hugs I would like to dedicate to my colleagues, I can say firmly my friends who gave me a warm welcome to Miogen lab. Thank you Monica, Jessica, Margherita, Antonella, Giulia, Valeria, Fabiano, Eustachio for all these wonderful days we spent together. Grazie davvero!

Special thanks I devote to you, Benedetta and Virginia, I can't imagine my Italian years without your loving words and big-hearted support. Vi voglio tanto bene!

At the end I thank my family and Marco. Without you Ada, Katarina, Gruja, Maja, Milos, Marko all this would never be possible. You always believed in me, cheered me up, made me laugh and loved me in the most generous way. I could never thank you enough mom, you have a special place in my heart. Marco thank you forever that you made Italia my home.