UNIVERSITY OF MODENA AND REGGIO EMILIA

Ph.D. School in Molecular and Regenerative Medicine

XXXIII CYCLE

The Italian National Registry for Facioscapulohumeral muscular dystrophy: a tool for advanced diagnosis and prognosis in neuromuscular diseases

Ph.D. Candidate:

Floriana Maria Napoli

Supervisor:

Prof.ssa Rossella Tupler

Ph.D. School Director:

Prof. Michele De Luca

Accademic Year 2019/2020

Summary

ABSTRACT

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common muscular myopathy with a prevalence of 1 in 20,000. So far, FSHD is considered an autosomal dominant disease with incomplete penetrance. The classic FSHD phenotype is characterized by onset in the first or second decade of life with progressive facial, shoulder girdle, and pectoral muscle weakness and atrophy, often asymmetric. The disease has been associated (95%) with reduced number (≤ 10 units) of tandemrepeats (D4Z4) in the subtelomeric region of chromosome 4q35. However, genotype–phenotype studies have shown that D4Z4 alleles with 4-10 repeat units are also present in 3-5% of general healthy population. Moreover, a 5-10% of patients with FSHD carry D4Z4 alleles within the range of the general population (≥11 RU) and represent a second form of disease, called FSHD2. From 1992 to the present, numerous evidences have accumulated indicating that FSHD is not transmitted as a simple Mendelian trait, but must be considered a complex disease to whose development numerous factors contribute. From 2007 to 2020, the Miogen laboratory of the University of Modena and Reggio Emilia has collected 7550 subjects in the Italian National Register for FSHD (INRF) of 5965 molecularly diagnosed families and 1585 single cases. Thanks to the Italian Clinical Network for FSHD, it was possible to collect standardized clinical information using the Comprehensive Clinical Evaluation Form (CCEF), from 2013 with prototype and from 2016 with actual version. It is a tool designed to describe the clinical heterogeneity of FSHD: in fact, the diagnosis and genetic counseling are further complicated by a variable phenotypic spectrum observed in subjects carrying the molecular defect, which can express with extremely different characteristics. Based on these premises, we analyzed the families collected from 1993 to 2020, to obtain an epidemiological context of the phenotypic spectrum. Our analysis re-evaluated penetrance and highlighted clinical differences between carriers of the D4Z4 small-sized alleles (DRA) and their families. In particular, we investigated family trios composed by the affected subject and its healthy parents in order to identify alternative genetic mechanisms by exploiting the WGS/WES technology. Moreover, we examine criteria for the preimplantation genetic diagnosis (PGD), testing feasibility use of microsatellite D4S1523 and two VNTR (D4S139 and D4S163), proximal to D4Z4, studying recombination in the segregation of three generations families with phase. These observations have the power to uncover additional factors involved in the risk of developing FSHD and to clarify the role of the reduction of D4Z4 elements in the 4q35 locus in the different myopathic phenotypes. Our work demonstrates the value of centralizing molecular and clinical diagnosis of rare disease and aims at the reconsideration of criteria for the diagnosis of the FSHD. This work represents the implementation of precision medicine, which by integrating clinical and genetic elements can create a phenotypic ontology that could distinguish a patient from others with similar Mendelian conditions and lay the basis for personalized treatments.

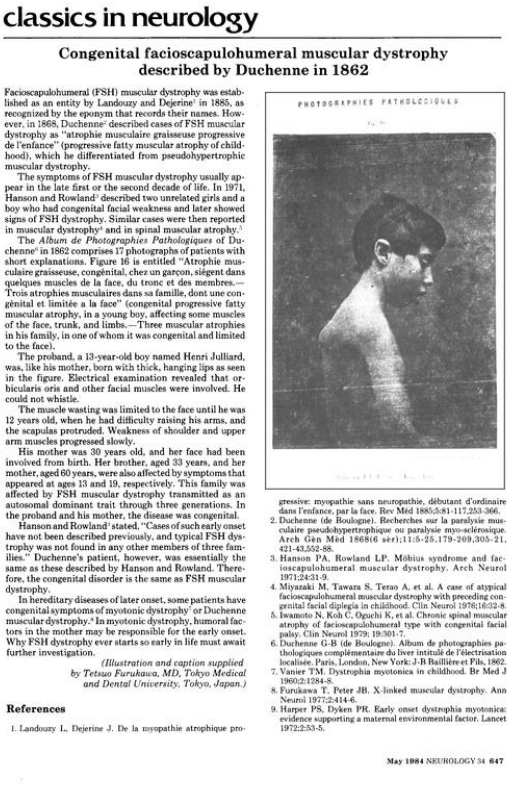
SINTESI

La distrofia muscolare facioscapolo-omerale (FSHD) è la terza miopatia muscolare più comune con una prevalenza di 1 su 20.000. Finora la FSHD è stata considerata una malattia autosomica dominante, con penetranza incompleta. Il fenotipo classico della FSHD è caratterizzato dall'esordio nella prima o seconda decade di vita, progressiva debolezza e atrofia dei muscoli facciali, del cingolo scapolare e pettorale, spesso con asimmetria. La malattia è associata (95%) ad un numero ridotto (≤ 10 Unità) di ripetizioni in tandem (D4Z4) nella regione subtelomerica del cromosoma 4q35. Tuttavia, studi genotipo-fenotipo hanno dimostrato che gli alleli D4Z4 con 4-10 unità ripetute sono presenti anche nel 3-5% della popolazione sana. Inoltre, il 5-10% dei pazienti con FSHD è portatore di alleli D4Z4 nel range della popolazione generale (≥11 RU) e rappresenta una seconda forma di malattia, chiamata FSHD2. Dal 1992 ad oggi abbiamo numerose evidenze che indicano che la FSHD non si trasmette come un semplice tratto mendeliano, ma deve essere considerata una malattia complessa al cui sviluppo contribuiscono numerosi fattori. Dal 2007 al 2020, il laboratorio Miogen dell'Università di Modena e Reggio Emilia ha raccolto nel Registro Nazionale Italiano per FSHD (INRF) 7550 soggetti, di cui 5965 famiglie e 1585 casi singoli. Grazie al Network clinico Italiano per la FSHD, è stato possibile raccogliere informazioni cliniche standardizzate utilizzando la scheda di valutazione CCEF (Comprehensive Clinical Evaluation Form), il cui prototipo è stato redatto nel 2013 e dal 2016 è utilizzata la versione attuale. Si tratta di uno strumento pensato per descrivere l'eterogeneità clinica della FSHD: infatti la diagnosi e la consulenza genetica sono ulteriormente complicate da uno spettro fenotipico variabile, osservato nei soggetti portatori del difetto molecolare, che può esprimersi con caratteristiche estremamente diverse. Sulla base di queste premesse, abbiamo analizzato le famiglie raccolte dal 1993 al 2020 per ottenere un contesto epidemiologico dello spettro fenotipico. Abbiamo rivalutato la penetranza ed evidenziato differenze cliniche tra i portatori degli alleli corti in D4Z4 (DRA) e nelle loro famiglie. Sfruttando la tecnologia WGS/WES abbiamo studiato trii familiari, composti da soggetto affetto e dai rispettivi genitori sani, al fine di identificare meccanismi genetici alternativi. Inoltre abbiamo testato la fattibilità dell'uso del microsatellite D4S1523 e due VNTR (D4S139 e D4S163), prossimali a D4Z4, per la diagnosi genetica preimpianto (PGD), studiando la ricombinazione nella segregazione di famiglie da tre generazioni. Questo studio ci da la possibilità di scoprire ulteriori fattori coinvolti nel rischio di sviluppare la FSHD e di chiarire il ruolo della riduzione degli elementi ripetuti D4Z4 nel locus 4q35 nei diversi fenotipi miopatici. Il nostro lavoro vuole dimostrare il valore della della diagnosi molecolare e clinica centralizzata nelle malattie rare e riconsiderare i criteri per la diagnosi della FSHD. Questo lavoro rappresenta l'implementazione della medicina di precisione, che integrando elementi clinici e genetici può creare un'ontologia fenotipica in grado di distinguere un paziente da altri, con condizioni mendeliane simili, e gettare le basi per trattamenti personalizzati.

1. INTRODUCTION

1.1 History

Facioscapolohumeral muscular dystrophy (**FSHD**, OMIM #158900) is the third most common form of hereditary myopathy with a reported prevalence 1 in 20.000 [Mostacciuolo et al., 2009]. The first description of a patient with FSHD appears in medical literature in an autopsy report by Jean Cruveilhier in 1852. The first clinical report is represented by photography of 17 FSHD affected patients published by neurologist Duchenne de Boulogne in his "*Album de Photographies pathologiques*" in 1862. Later the same patient's family was described in the "*Revue photographique des hopitaux de Paris"* in 1869 [Duchenne 1862, Duchenne 1869]. He was a pioneer in medical photography, and using localized faradization as a clinical diagnostic and prognostic tool, he demonstrated how the muscles of the face generate the different expressions and the effect of injuries characterized various neuromuscular diseases. He described cases of FSHD as progressive fatty muscular atrophy of childhood, differentiated from pseudohypertrophic muscular dystrophy. He summarized the disease like as follows: "it is congenital, usually appear in the late first or the second decade of life, affect some muscles of the face, trunk and limbs, and could be limited to the face anticipating phenotypic variability. Some patients were born with thick, hanging lips; some could not whistle. Weakness of shoulder and upper arm muscles progressed slowly". In 1885, Landouzy and Dejerine presented a detailed description of patient's difficulty in raising their arms and the protuding scapulae, the diffuse muscular atrophy and the involvement of distal limbs in addition to facial weakness [Landouzy and Dejerine, 1885]; in fact, the first name assigned to FSHD disease was "Landouzy-Dejerine form of muscular myopathy". Reports of large pedigrees suggested an autosomal-dominant inheritance model for this disorder [Pearson, 1933]. In 1950, Tyler and Stephens provided a detailed description of a large six-generation family affected by FSHD, reporting the progressivity of the disorder in the affected individuals and brachioradialis involvement of the



classics in neurology

Congenital facioscapulohumeral muscular dystrophy described by Duchenne in 1862

Facioscapulohumeral (FSH) muscular dystrophy was established as an entity by Landouzy and Dejerine[1] in 1885, as recognized by the eponym that records their names. However, in 1868, Duchenne[2] described cases of FSH muscular dystrophy as "atrophie musculaire graisseuse progressive de l'enfance" (progressive fatty muscular atrophy of childhood), which he differentiated from pseudohypertrophic muscular dystrophy.

The symptoms of FSH muscular dystrophy usually appear in the late first or the second decade of life. In 1971, Hanson and Rowland[3] described two unrelated girls and a boy who had congenital facial weakness and later showed signs of FSH dystrophy. Similar cases were then reported in muscular dystrophy[4] and in spinal muscular atrophy.[5]

The *Album de Photographies Pathologiques* of Duchenne[2] in 1862 comprises 17 photographs of patients with short explanations. Figure 16 is entitled "Atrophie musculaire graisseuse, congénital, chez un garçon, siégent dans quelques muscles de la face, du tronc et des membres.—Trois atrophies musculaires dans sa famille, dont une congénital et limitée a la face" (congenital progressive fatty muscular atrophy, in a young boy, affecting some muscles of the face, trunk, and limbs.—Three muscular atrophies in his family, in one of whom it was congenital and limited to the face).

The proband, a 13-year-old boy named Henri Julliard, was, like his mother, born with thick, hanging lips as seen in the figure. Electrical examination revealed that orbicularis oris and other facial muscles were involved. He could not whistle.

The muscle wasting was limited to the face until he was 12 years old, when he had difficulty raising his arms, and the scapulas protruded. Weakness of shoulder and upper arm muscles progressed slowly.

His mother was 30 years old, and her face had been involved from birth. Her brother, aged 33 years, and her mother, aged 60 years, were also affected by symptoms that appeared at ages 13 and 19, respectively. This family was affected by FSH muscular dystrophy transmitted as an autosomal dominant trait through three generations. In the proband and his mother, the disease was congenital.

Hanson and Rowland[3] stated, "Cases of such early onset have not been described previously, and typical FSH dystrophy was not found in any other members of three families." Duchenne's patient, however, was essentially the same as these described by Hanson and Rowland. Therefore, the congenital disorder is the same as FSH muscular dystrophy.

In hereditary diseases of later onset, some patients have congenital symptoms of myotonic dystrophy[7] or Duchenne muscular dystrophy.[8] In myotonic dystrophy, humoral factors in the mother may be responsible for the early onset. Why FSH dystrophy ever starts so early in life must await further investigation.

(Illustration and caption supplied by Tetsuo Furukawa, MD, Tokyo Medical and Dental University, Tokyo, Japan.)

PHOTOGRAPHIES PATHOLOGIQUES

References

1. Landouzy L, Dejerine J. De la myopathie atrophique progressive: myopathie sans neuropathie, débutant d'ordinaire dans l'enfance, par la face. Rev Méd 1885;5:81-117,253-366.
2. Duchenne (de Boulogne). Recherches sur la paralysie musculaire pseudohypertrophique ou paralysie myo-sclérosique. Arch Gén Méd 1868(6 sér);11:5-25,179-209,305-21, 421-43,552-88.
3. Hanson PA, Rowland LP. Möbius syndrome and facioscapulohumeral muscular dystrophy. Arch Neurol 1971;24:31-9.
4. Miyazaki M, Tawara S, Terao A, et al. A case of atypical facioscapulohumeral muscular dystrophy with preceding congenital facial diplegia in childhood. Clin Neurol 1976;16:32-8.
5. Iwamoto N, Koh C, Oguchi K, et al. Chronic spinal muscular atrophy of facioscapulohumeral type with congenital facial palsy. Clin Neurol 1979; 19:301-7.
6. Duchenne G-B (de Boulogne). Album de photographies pathologiques complémentaire du liver intitulé de l'électrisation localisée. Paris, London, New York: J-B Baillière et Fils, 1862.
7. Vanier TM. Dystrophia myotonica in childhood. Br Med J 1960;2:1284-8.
8. Furukawa T, Peter JB. X-linked muscular dystrophy. Ann Neurol 1977;2:414-6.
9. Harper PS, Dyken PR. Early onset dystrophia myotonica: evidence supporting a maternal environmental factor. Lancet 1972;2:53-5.

Figure 1. Duchenne manuscript about FSHD patients- Neurology 1984

subjects. They showed first results about the preservation of certain muscle fibers in contrast to the adjacent fibers and also variables onset and FSHD symptoms. [Tyler and Stephens, 1950]. In 1954, Walton and Natrass identified the clinical criteria for FSHD diagnosis, showing the slow progression of the disease with asymmetric muscular weakness limited to specific muscles of the body. They also described 33% of "abortive cases", people with minimal sign of disease who were discovered because they had at least one affected child [Walton, 1954]. In 1982, through the clinical examination of a large cohort of 107 recruited subjects from 19 families (73 symptomatic patients), Padberg performed the first attempt to stratify the previously described phenotypes in six stages of severity. Stage 1 included the scapular fixators but also the anterior tibial, a muscle that had been seen frequently affected in FSHD patients; stage 6, on the other side, meant that the patient was totally dependent on wheelchair and on nursing care (6% of all cases). The main clinical features were: facial weakness was present in 10% of cases, shoulder weakness in 82% and foot extensor weakness in 8%, while the pelvic girdle and calf weakness was not observed in any patient. [Padberg 1982]. Studying with electromyography (EMG) and histological analyses the characteristic of 24 biopsies from 23 FSHD patients, Padberg described in his doctoral thesis the predominant features: variation of fiber diameter (92%), fiber hypertrophy (>100 microns) (33%), necrosis and phagocytosis (33%), small angulated fibers (25%), endomysial fibrosis (58%) and endomysial fat infiltration (42%). Also, moth-eaten fibers, which are a rather unspecific finding, reported also in systemic connective tissue disorders and chronic dystrophies, were observed in 11 biopsies (46%). Lunt and colleagues in 1989 reported data about inheritance: reproductive fitness, judged by offspring, was not impaired, and parents have a 50% chance to transmit the disease. The estimation of age dependent penetrance was fixed as > 95 % by age of 20, considering 30 Welsh families; life expectancy was not changed by the disease, despite the progressive disability. These studies also suggested risk subjects remaining asymptomatic by age of 21 can be reassured against developing a severe phenotype; however, it was remarked that there were no predictive tests for disease progression. The reported observations constituted the basis for linkage analysis in FSHD families [Lunt et al, 1989].

1.2 Clinical diagnostic criteria and clinical features

In 1991, the International Consortium established four clinical and genetic criteria for FSHD diagnosis [Padberg et al, 1991]: 1) onset of the disease in facial or shoulder girdle muscles, but the majority of subjects refer weakness of scapular girdle muscle; sparing of the extra-ocular, pharyngeal and lingual muscles and the myocardium; facial weakness in more than 50% of the affected family members, including affecting eye closure (orbicularis oculi) and peri-oral muscles (orbicularis oris); (3) autosomal dominant inheritance in familial cases; (4) evidence of

myopathic disease in EMG and muscle biopsy in at least one affected member without biopsy features specific to alternative diagnoses. Creatine kinase level in blood can be slightly increased, but with no correlation with disease severity and Lunt had excluded the possibility of using it as presymptomatic test already [Lunt et al, 1991]. Others typical feature of FSHD are considered the asymmetric muscle involvement, progression and extension of weakness to other muscular groups: truncal and lower extremity muscles. The weakening of the abdominal muscles results in lumbar lordosis and in the appearing of the Beevor's sign, a typical FSHD signature. Weakness of tibialis anterior muscle is characterized by difficulty in lifting up feet at walk (dropped foot), involving abdominal and foot extensor muscles at an early stage. The deltoid muscles remain unaffected for a long period of time and often have a particular pattern of atrophy, i.e., partial and proximal. In general, the pelvic girdle weakness and upper arm weakness may occur at any time after the onset of shoulder girdle weakness. For these observation, the following definitions were introduced: (1) "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 skeletal muscle. It was reported that the clinically recognizable age of onset is often very variable and the mean age of onset is in the second decade. Onset before the age of 5 years has been rarely reported [Brouwer et al, 1994]. 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 and rarely there can be long periods of apparent arrest of progression.

The diagnosis is improbable in the case of a recessive pattern of inheritance and sporadic cases are also reported, although their frequency is unknown. Studies on FSHD subjects performed with electromyography (EMG) and histological analyses revealed non-specific myopathic changes associated, in some cases, with neurogenic and/or inflammatory aspects [Lin et al 1991]. In a specific work, Dorobek and colleagues demonstrated that there was no correlation between clinical features and electrophysiological findings [Dorobek et al, 2013]. The most advanced changes were found in *tibialis anterior* and deltoid muscles, but overall some of these changes were unspecific for myopathy and the degree of their intensity differed between muscles. More recently, magnetic resonance imaging of the involved muscles (MRI) show pattern similar to inflammatory myopathies, as it correlates with biological perturbations such as selective inflammation, neoangiogenesis, lipid metabolism and adipokine production [Tasca et al, 2012].

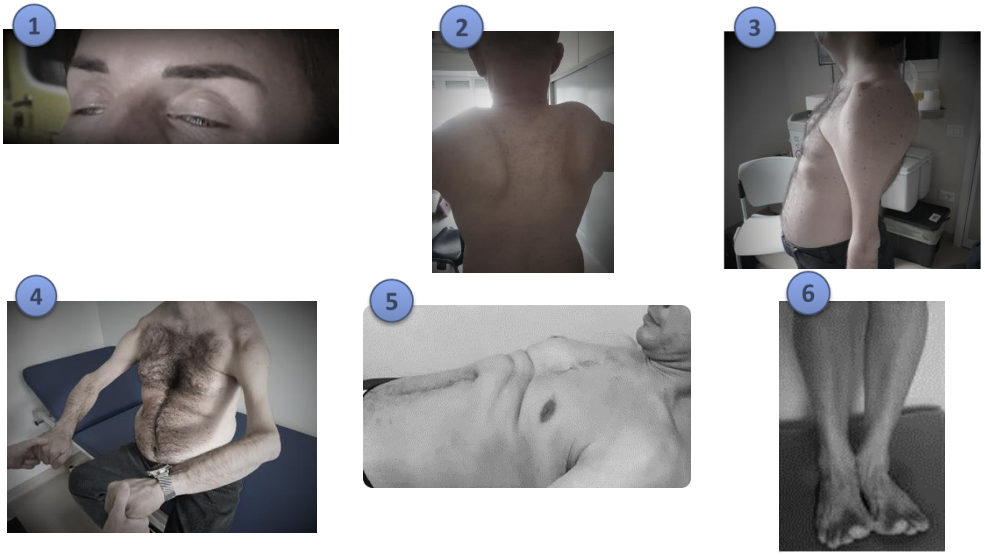


Figure 2. Muscles affected primarily: 1) Facial muscle (orbicularis oculi, oris and zygomaticus. 2) Shoulder girdle muscles (trapezius, serratus anterior, teres major, latissimus dorsi, rhomboids). 3) Upper limbs muscles (biceps, triceps, deltoid. 4) Pectoralis muscles. 5) Abdominal muscles 6) Tibialis anterior. Source: Photos collected with patients' consent during the clinical examination in Modena, Policlinico Hospital

By contrast, exclusion criteria are the following: (1) onset in the pelvic girdle muscles suggests alternative diagnoses, although subsequent pelvic girdle involvement is common in FSHD; (2) extra-ocular, masticatory, pharyngeal and lingual muscle weakness is not part of the disease; (3) regression of symptoms and signs does not occur, and would exclude the diagnosis; (4) severe and diffuse contractures exclude the diagnosis of FSHD; (5) cardiomyopathy is not part of the disease. Symmetric weakness and atrophy at presentation is unusual and can necessitate caution before accepting the diagnosis as FSHD. Neck extensor, intrinsic hand and *triceps surae* muscle weakness is described as uncommon features of disease, but occasionally observed within families; (6) persistent high CK values above five times the upper limit.

1.3 Identification of the molecular defect responsible for development of FSHD

The need of a presymptomatic diagnostic test gave the impetus to many studies aimed at mapping the putative genetic locus of FSHD [Lunt et al 1989]. The International Consortium for FSHD linkage has guided the first attempt of finding a genetic trait shared by FSHD affected patients in 1988, creating an exclusion map for FSHD [Sarfarazi et al, 1989; Jacobsen et al, 1990]. Two years later, positional mapping performed over ten multi-generation Dutch families assigned the FSHD locus to chromosome 4, associated with microsatellite D4S171 [Wijmenga et al, 1990]. This result was confirmed using additional polymorphic markers in other families [Upadhyaya et al., 1990; Upadhyaya et al., 1991]. Subsequently, Wijmenga and coworkers showed that the D4S139, which is a Variable Number Tandem Repeat Structure (VNTR) locus on 4q telomeric region, is the most

closely marker linked to FSHD. Two-point linkage analysis performed in nine informative families mapped it in the chromosomal sub-telomeric position 4q35, showing a maximum combined lod score (in linkage) [Wijmenga et al, 1991]. D4S139 has been proven to be located close to the telomeric region of the long arm of chromosome 4 by FISH (fluorescent in situ hybridization), thus firmly establishing the location of the FSHD gene in the subtelomeric region of chromosome 4 [Wijmenga et al, 1991]. In order to obtain a more detailed position of the FSHD locus with other four markers on 4q35 region, D4S171-F11-D4S163-D4S139, the International Consortium extended the study for linkage analysis to 65 families with 504 affected and 559 unaffected. For each marker, a mean of 648 meiosis were informatives, achieving the identification of the two closest markers to the FSHD locus: D4S139 and D4S163 [Sarfarazi et al., 1992]. The genetic region responsible for FSHD was defined to be the distal position to the D4S139 locus and the cosmid clone p13E was mapped to the same distal region. Its subclone p13E-11 detected by Southern hybridization EcoRI digested DNA fragments, usually larger than 28 kilobases (kb); the dimension of these fragments was found from 20 Kb to more than 50 Kb in the majority of the healthy population (72%), but fragments from 14 Kb to 28 Kb were numerically enriched in the population of FSHD affected individuals [Wright et al, 1993]. Wjimenga and coworkers detected also de novo DNA rearrangements characterized by short (14-28 kb) EcoRI fragments in five out of six cases with negative familial history; the one missing fragment was not detectable. [Wijmenga et al, 1992]. In the work of Wimenga the presence in 8 healthy individuals involved in the study of p13E-11 alleles smaller than 28 kb was apparently considered insignificant and this incongruity rose still unsolved questions. In 1993, it has been confirmed that these rearrangements associated with FSHD correspond to deletions of integral numbers of tandemly arrayed of 3,3 Kb *Kpn*I restriction fragments, named **D4Z4** repeats [van Deutekom et al, 1993]. Notably, the p13E-11 probe detects two polymorphic loci and a specific fragment of 9.5 Kb on chromosome Y. By haplotype analysis one of the loci was assigned to the 4q35 region (D4F104S1) and the other to the 10q26 region (D4F104S2). Indeed, the p13E-11 probe detected at least two pairs of EcoRI alleles, one derived from the 4q and the other from 10q. These alleles sometimes overlap because have a 98% homology of arrangement of *KpnI* tandem repeat and flanking sequences, [van Deutekom et al, 1996]; homology is extended on both the centromeric (42 Kb) and telomeric adjacent regions [van Geel et al, 2002]. Recombination events between these 4q35 and 10q26 are found in 20-30% of the general population, resulting in translocation of arrays of the 4 on chromosome 10 or vice versa [van Deutekom et al, 1996; Rossi et al, 2007]. Despite the high homology, FSHD is associated exclusively with reduction of D4Z4 repeats unite on chromosome 4q [Bakker et al, 1995; van Deutekom et al, 1996; Matsumura et al, 2002].

1.3.1 Molecular diagnosis

The search for sequence divergence between 4q and 10q revealed a particular SNP in the 10q D4Z4 repeats that creates a *Bln*I restriction site, absent in the 4q35 D4Z4 repeats. This restriction site is located approximately 3 Kb upstream of KpnI site of the first D4Z4 repeat unit and then about 80 nucleotides upstream each KpnI site of the 10q26 D4Z4 unit repeats. This important difference allows discriminating a 4-type array to a 10-type array after an *Eco*RI/*Bln*I digestion [Deidda et al, 1996; Upadhyaya et al, 1997]. After the double digestion of genomic DNA, the p13E-11 alleles at 4q35 will result 3 Kb shorter than the EcoRI single digestion fragment. With a sensitivity of almost 95% and a specificity of almost 100% in DNA from patients carrying fragment smaller than 35 Kb, this method was officially validated and became the diagnostic criteria to diagnose FSHD [Upadhyaya et al, 1997]. Normal subjects carry more than 10 D4Z4 units originating from chromosome 4q, while array of 8 D4Z4 units or shorter are present in the majority of FSHD patients [Upadhyaya et al, 1993; Wijmenga et al, 1994; Lunt et al, 1995]. To complicate the interpretation of the results of p13E-11 Southern blot hybridization, translocation of either the complete repeat array, or of some repeats only can occur, leading to a 4q and 10q-type repeats mixture and double digestion *Eco*RI/*Bln*I is not sufficient. In these cases, other analyses can support diagnosis. In particular, Southern hybridization of NotI digested DNA allows the identification of D4Z4 arrays from chromosome 4q or 10q. This is because the distance between the NotI restriction site at 4q and 10q is differs. At 10q the distance between NotI sites gives rise to a fragment 100 Kb larger than the one from 4q. Moreover, the probe B31 specifically detects NotI fragments from chromosome 4 and the restriction pattern identified can be compared to the one detected by p13E-11 This approach makes possible to ascertain the chromosome location of each D4Z4 repeat array [Lemmers et al, 2003]. Notably, *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. All these observations suggest that the D4Z4 array is highly recombinogenic. 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].

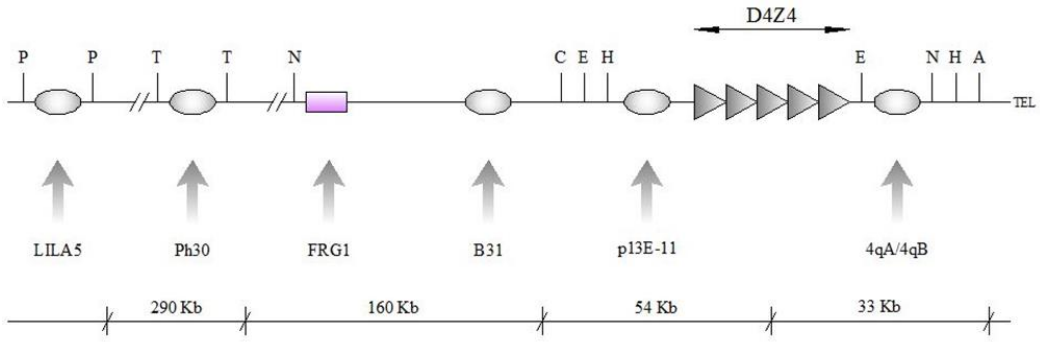


Figure 3. Restriction map of 4q35. From down the probes used for FSHD molecular diagnosis and respectively distance. The enzyme used are: P=PstI, T=TaqI, N=NotI, C=ClaI, E=EcoRI, H=Hind, A=AhdI.

1.3.2 Preimplantation Genetic diagnosis and test (PGD-PGT)

The great variability together to other several factors complicate FSHD molecular diagnosis. These do not hinder the request for prenatal genetic counseling and diagnosis (PND) from women, belonging to families that suffer from FSHD. In parallel, in the last few decades, a great improvement in genetic testing and counseling occurred to the point that almost all women have now become candidates for some sort of prenatal screening or diagnostic during pregnancy [Minkoff, 2014]. Our laboratory has performed forty-eight prenatal analyses from 2008 to 2019, ten couples required testing more than once, up to five times. PND (pre-natal diagnosis) can be performed trough chorion villus biopsy. The CVS can be carried out between 11 and 14 weeks of pregnancy, followed by the conventional southern blot used for normal diagnosis [Upadhyaya et al., 1999]. Because of the high homology between 4q35 and 10q26, the chromosomal origin of D4Z4 repeat arrays is based on EcoRI/BlnI and also XapI digestion, which selected only chromose 10q. PND is offered by a few numbers of specialized laboratories, in Italy two. Moreover, as this reference method requires large amounts of genomic DNA, the time for diagnosis may be prolonged for technical reasons, sometimes up to several weeks of diagnosis delay. An alternative method/option is the preimplantation genetic testing (PGT). PGT is an early form of prenatal genetic diagnosis where abnormal embryos are identified, thereby allowing transfer of genetically normal embryos. It consists in obtaining cellular biopsy samples from developing human embryos, i.e., trophectoderm biopsy of blastocyst, generated through in vitro fertilization, to evaluate their genetic composition. This technology has become an integral part of Assisted Reproductive Technology (ART) procedures. PGD (preimplantation genetic diagnosis) is applied to specific genetic abnormalities known to be present in one or both parents, with the purpose of preventing the birth of affected children [Brezina et al., 2012]. The principal advantage of PGD is that it avoids the problem of terminating a pregnancy as the genetic status of the embryo is revealed before the implantation. Indeed, there are guidelines for regulation and

standardization of diagnostic testing and to assist scientists interested in PGT in developing the best laboratory and clinical practice possible. Blastocyst stage biopsy, despite whole embryos, is also widely used and now thanks to the improving of single-cell genetic testing methods. The analysis can be performed on one cell with lower general risk, but more reduced time of analysis [Brezina et al., 2012; Traeger-Synodinos et al., 2017]. The molecular testing could be performed by using different techniques: inverse-PCR, long-distance PCR, southern blotting, direct DNA sequencing trough NGS, DHPLC, TTGE, MLPA, ACGH, PCR, capillary electrophoresis, ARMS-qPCR, linkage analysis, mini-sequencing, multiplex PCR, nested PCR, FISH, SNP microarray [Chen et al., 2018; Dahdouh et al., 2015]. Often, a direct multiplex PCR method is the preferred choice, becuse is sensitive and fast: permitting both to detect known mutation at several genomic loci simultaneously or to amplify informative polymorphic regions to perform linkage analysis in specific disease loci. Preimplantation genetic diagnosis (PGD) is a frequent request from FSHD families with several affected relatives. However, in FSHD, the D4Z4 repeat array can only be detected indirectly using nearby polymorphic markers, because millions of cells are needed to perform southern blot genetic test and this approach is not applicable to single cell methodologies used to perform PGD. The impossibility of conducting a direct test dictate further considerations. In the case of FSHD, PGD can be performed using polymorphic markers that are in linkage with the D4Z4 reduced allele. This requires a family study for the identification of the associated markers and possible recombination events must be taken into account. These conditions have been considered in this thesis in order to provide the best preconceptional counseling to couples. Barat-Houari and colleagues in 2010 propose a newly designed strategy for FSHD molecular preimplantation diagnosis, based on a rapid one-step multiplex PCR protocol allowing robust segregation analysis of highly informative markers proximal to the D4Z4 locus. They are 4 microsatellite polymorphic loci, D4S2390, D4S1652, D4S2930 and D4S1523, in order to achieve an indirect diagnosis applicable on single cell, based on linkage analysis. Recombination rates were estimated from 144 meioses: 28 meioses from the 12 FSHD families, 64 from the 6 control families and 52 from 4 CEPH families. This strategy allowed to determine a recombination fraction 0=0.021 in the interval D4S2390- D4S1523 studying 22 pedigrees, showing high sensitivity and specificity and LOD score = 2.98. Nevertheless, the work suffers from a probable overestimation of the number of informative meiosis for linkage analysis, thus results presented in this work are not reliable to estimate the feasibility of PGD in FSHD. The same authors write that such study was primarily intended to test the accuracy of this PCR-based approach at the single-cell level and to estimate its feasibility in PGD [Barat-Houari et al., 2010]. Due to the poor number of markers available and the strong tendency to recombination of the locus, PGD is still a challenging goal to be achieved [Tawil et al., 2010].

1.3.3 Molecular signatures

The hypothesis that additional DNA sequences flanking the D4Z4 repeat array are necessary for disease development was very studied. In fact, since 1992 it was known that D4Z4 reduced arrays were present in the general population [Wijmenga et al, 1992]. In 2002, a polymorphic bi-allelic segment 10 kb of β-satellite DNA was identified directly distal to D4Z4 and existing in two allelic forms, 4qA and 4qB. [van Geel et al, 2002]. It is detectable as a HindIII restriction fragment and in healthy individuals an equal distribution of 4qA and 4qB (42% and 58% respectively) was observed on chromosome 4, on chromosome 10 only alleles of the 4qA type were described. In FSHD patients, 44 de novo cases and 36 unrelated familial cases, 4qA allele was found only in chromosomes 4 bearing the D4Z4 contraction, and never in those with the 4qB allele [Lemmers et al, 2002]. Thus, the authors supposed that, in addition to a contraction of D4ZA, other cis-acting elements on 4qA might be necessary for the development of FSHD. Another polymorphic site was found upstream the D4Z4 array, a Simple Sequence Length Polymorphism (SSLP), at both 4q35 and 10q26: 17 and 8 different SSLPs has been described for the 4q35 and 10q26 subtelomere respectively [Lemmers et al, 2007]. Out of 17, three SSLPs on chromosome 4 were found to be associated with FSHD: the rare variants 4A159 and 4A168, but also the common variant 4A161. Finally, a single nucleotide polymorphism (SNP), AT[T>C]AAA, has been found among the chromosomal region distal to the last D4Z4 repeat in the PLAM sequence of the 4qA alleles, that provides a Poly Adenylation Signal (PAS). All these polymorphisms, named 4A (159,161,168) PAS, has been proposed like the specific FSHD molecular signature [Lemmers et al, 2010]. Even though, further analysis showed that there is a little predictive value of the 4qA161PAS haplotype when no family history is available, this haplotype is as frequent as a common polymorphism, being carried by 1.3% of healthy subjects [Scionti et al., 2012]. Moreover, Ricci and colleagues in a wide phenotype-genotype study involving 530 subjects, pointed out that no 4q haplotype was exclusively associated with the presence of the disease [Ricci et al, 2013]. In general, FSHD genotype model characterized by reduction D4Z4 array, 4qA 159,161,168 and a SNP in pLAM sequence has been proposed because carries a polyadenylation signal stabilizing the mRNA of DUX4, the D4Z4 gene, allowing the expression of the double homeobox protein DUX4 between the most distal D4Z4 repeat of the array and adjacent sequence. The effect due to DUX4 overexpression has been proposed to be crucial for FSHD development, considering the role of homeobox genes in human embryogenesis and its expression restricted to testis and FSHD muscles.

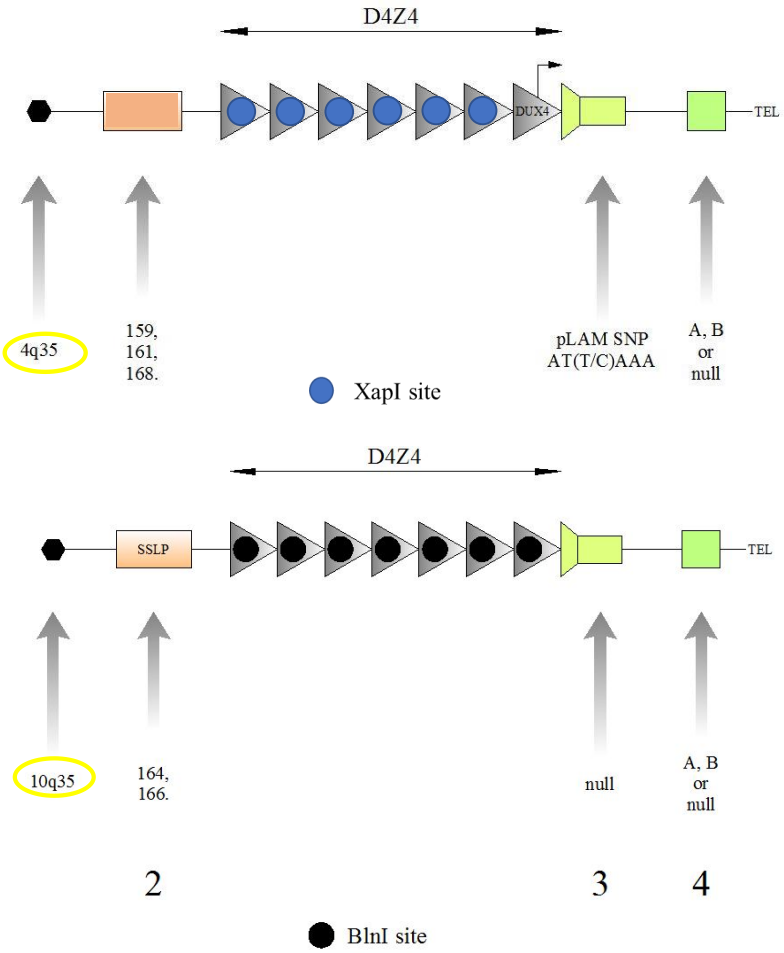


Figure 4. Polymorphisms at 4q35 and 10q26: from left the Simple Sequence Length Polymorphism, an 8 bp insertion/deletion, and two SNPs localized 3.5 kb proximal to D4Z4 and varies in length (in figure 159-168 bp). SNP AT(T/C)AAA in the pLAM region; a large sequence variation, termed 4qA o B that is distal to D4Z4. The chromosome that does not hybridize to A and B probes are termed "null".

1.4 The main hypothesis for FSHD pathogenesis

One of the main issues in FSHD is to connect the D4Z4 reduced member to muscle disease. The main hypothesis is that the reduction of the number of the repetitive elements below a certain threshold might result in a more open chromatin configuration of the region leading to transcriptional depression of 4q35 genes [Gabellini et al 2002, Lemmers et al., 2007; Lemmers et al., 2010; van der Maarel et al., 2011; Hewitt et al., 1994].

1.4.1 DUX4

Each 3,3 Kb D4Z4 unit, apart from containing the two homeobox sequences within the same open reading frame, is flanked by sequences typical of the heterochromatin domains, LSau motifs and hhspm3 motif. They are two classes of repetitive DNA; hhspm3 is a GC-rich low copy repeat [Zhang et al, 1987] and LSau [Agresti et al 1987, Agresti et al, 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]. Each D4Z4 unit contains a putative promoter and a single open reading frame (ORF) encoding a putative double homeobox gene, named *DUX4* [Gabriels et al, 1999; Hewitt et al, 1994; Lyle et al, 1995]. The chromosomes where the last copy of the *DUX4* gene has a third region immediately flanking and stabilizing the transcript owing to the presence of the poly(A) signal (PAS) are defined permissive. 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]. Although, for a long time, the functionality of the *DUX4* gene was questioned, after 2007, D4Z4 homologues and organization in array 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. [Clapp et al, 2007]. In 2007, it has been demonstrated *DUX4* protein overexpression induces cell death by apoptosis in myoblasts transfected with genomic fragments containing the *DUX4* gene. In particular, its overexpression may caspase 3/7 activation and alter emerin distribution at the nuclear envelope [Kowaljow et al, 2007]. Moreover, using cells transfected too, it is reported that aberrant *DUX4* expression triggers a deregulation cascade inhibiting muscle differentiation, sensitizing cells to oxidative stress, and inducing muscle atrophy [Bosnakovski et al, 2008]. More detailed analysis of expression shows that 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]. 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. 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]. Finally, large studies based on clinical evaluation demonstrate that 2% of healthy individuals from the general population bear one FSHD1-sized D4Z4 allele with the possibility of expressing the *DUX4* transcript [Scionti et al., 2012]. Moreover, in 2013 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 human, but not shows pathological phenotype.

1.4.2 Other candidate genes in the 4q35 region

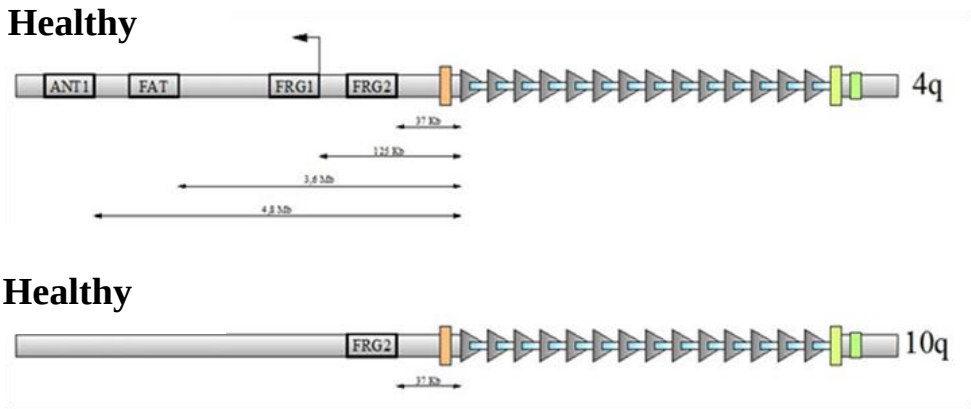


Figure 5. 4q35 genes and distance to D4Z4

ANT1 OR SLC25A4 GENE

Zooming off the region 4q35, positioned 4.8 Mb proximal to the D4Z4 array we found *ANT1* (Adenine Nucleotide Translocator 1) or *SLC25A4* gene. It is the muscle/heart/brain isoform of Ant, which is the prototypical member of the Mitochondrial carrier family (MCF) proteins. *ANT1* is involved in the response to oxidative stress and its loss of function, due to homozygous transport-defective or null mutations, cause cardiomyopathy, myopathy and lactic acidosis. Several missense mutations of *ANT1* cause Progressive External Ophtalmoplegia manifested by ptosis, ophthalmoplegia and skeletal muscle weakness with multiple mtDNA deletions. However, transgenic mice overexpressing *ANT1* in muscle were normal [Gabellini et al, 2006].

FRG1

FSHD Region Gene 1 (*FRG1*) is located 125 kb proximal to D4Z4 sequences toward the centromere [van Deutekom et al, 1996] and is highly conserved from vertebrates to invertebrate and has been shown to be crucial for muscle development [Hanel et al, 2009; Liu et al, 2010]. Nevertheless, the molecular function of *FRG1* has not yet been clarified, but some lines of evidence suggest that *FRG1* is involved in developing vasculature too. This could be in accord with the observation that, in addition to muscular symptoms, FSHD patients also could exhibit retinal vasculopathy as demonstrated in a small cohort [Fitzsimons et al., 1987] and recently in patients with large deletions [Hamel and Tawil, 2018; Statland et al., 2013]. Human *FRG1* has been analyzed endogenous in muscle cells during myogenesis and three subcellular pools of *FRG1*, nuclear, cytoplasmic and sarcomeric, have been identified. Thus, *FRG1*, an actin-bundling protein previously describe, could be critical for muscle development [Liu Q et al., 2010]. *FRG1* has been found to associate with RNA processing components and large-scale proteomic studies revealed that *FRG1* is associated with

human spliceosome complex (van Koningsbruggen et al., 2004; Rappsilber et al., 2002). Furthermore, mis-spliced mRNA transcripts were found in transgenic mice overexpressing FRG1 [Gabellini et al., 2006]. However, protein and mRNA expression profile studies failed to determine conclusively if *FRG1* is mis-regulated in FSHD. Separate lines of transgenic mice overexpressing the human form of the 4q35 genes *FRG1, FRG2 or ANT1* has been generated, in our laboratory, to test whether their overexpression may result in the development of the disease. This study showed that *FRG1* over-expression causes a dystrophic phenotype sharing several features with FSHD patients [Gabellini et al, 2006]. Indeed, *FRG1* transgenic mice showed spinal abnormality with no evidence of skeletal deformities suggesting that the abnormal phenotype was due to muscle weakness. Skeletal muscle from *FRG1* transgenic mice showed dystrophic histological and ultrastructural features characterized by increased fiber size variability and necrosis, number of central nuclei, and connective tissue. As in FSHD patients, *FRG1* overexpression induces a progressive myopathy in which sarcolemma is not damaged and no elevated creatine kinase levels are detected. *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]. 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]. 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]. Recently, our laboratory has discovered new roles and deepened some one already described in the literature; 1) *FRG1* is an RNA binding protein whose overexpression causes overt myopathy in transgenic mice; 2) postnatal characterization of transgenic mice confirms that *FRG1* overexpression causes the global perturbation of muscle maturation. This includes the anomalous expression of embryonic/neonatal proteins fundamental for muscle structures, contraction and metabolism 3) *FRG1* overexpressing mice show the deceleration of growth curve and reduction of muscle cross-sectional area before the appearance of dystrophic signs in absence of protein synthesis reduction 4) Moreover, from post-natal day 28 we detected a progressive inflammatory response in *FRG1* muscle with an involvement of macrophages, T-cells and fibroadipogenic progenitors (FAPs) [unpublished results]. Currently *FRG1* remains the first candidate gene for the development of the disease, but the complexity of the background of the human genome requires further studies for homogeneous results.

FRG2

FRG2 (FSHD Region Gene 2) maps 37 kb proximal to the D4Z4 repeat array. It is a gene of unknown function, expressed only in myoblasts of FSHD patients and its transcript is undetectable in normal individuals. It has not been considered an attractive FSHD candidate gene, as it is absent in some

FSHD patients with a proximally extended deletion (Lemmers et al. 1998; Lemmers et al. 2003]. Rijkers et al demonstrate in 2004 that *FRG2* is upregulated in cultures of differentiating myoblast from FSHD patients compared to healthy controls. Its expression in MyoD driven myogenesis in fibroblasts indicates that may play a role in the myogenic process. In addition, transient transfection experiments suggest that increasing numbers of D4Z4 units inhibit expression of *FRG2*. [Rijkers et al, 2004]. The interest in *FRG2* has been dropped because its over-expression in muscles of transgenic mice did not cause myopathy [Gabellini et al, 2006]. 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]. Our laboratory explored the role of *FRG2* and showed that in FSHD muscles the overexpression is inversely related to the number of D4Z4 repeats [Gabellini et al, 2002]. Indeed, *FRG2* is subject to reversible silencing appearing to be chromatin-mediated, because chromatin modifying agents and genotoxic agents induce their selective derepression. Consistent with this, our laboratory found that histone deacetylases (HDACs) 4 and 5, which regulate skeletal myogenesis, selectively regulate *FRG2* expression, and are recruited to D4Z4 repeats.

FAT1

FAT1 (FAT Atypical Cadherin 1) gene is located 3.6 Mb centromeric to the D4Z4 repeat array at 4q35 [Geng et al. 2012]. *FAT1*, a protocadherin, has been reported to be expressed in developing muscles and tendons [Smith et al. 2007] and to be regulated by muscle developmental genes such as Pax3, Lbx1, or Met [Vasyutina et al. 2005, Barber et al. 2002]. Caruso et al. report the finding that *FAT1*-deficient mice reproduce the highly selective muscular and non-muscular aspects of the clinical picture of FSHD. They show that *FAT1* is required during development to shape specific groups of shoulder and facial muscles by modulating the polarity of myoblast migration. 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. Moreover, either in presence or absence of D4Z4 contractions, mechanisms leading to tissue-specific deregulation of *FAT1* expression are associated with FSHD and may contribute to causing regional-specific muscle shape abnormalities that prefigure muscle degeneration in the adult. [Caruso et al, 2013].

DBE

Finally, the D4Z4 sequence contains a transcription repression element, named DBE (D4Z4 binding element), discovered after the analysis of the interaction between the locus and nuclear proteins. It able to recruit a multi-protein complex named DRC (D4Z4 recognition complex) in vitro and in vivo comprising of YY1, HMGB2 and nucleolin which mediates transcriptional repression of upstream

genes [Gabellini et al, 2002]. In FSHD muscle, the deletion of an integral number of repeats lead to a reduced number of DRCs in the region, provoking silencing removal and aberrant expression of nearby genes [Gabellini et al., 2002]. Indeed, 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 [Yao et al, 2001]. Recent studies showed that myoblasts from FSHD patients express the lncRNA called DBE-T (D4Z4 Binding Element Transcript) transcripted by sequence of a non-coding RNA, located in a proximal region of the repeat array. DBE-T could promote the recruitment of the Trithorax group protein AshlL, associated with transcriptionally active chromatin, to the D4Z4 locus thus inducing topological reorganization of the FSHD locus leading to de-repression of 4q35 genes, *FRG1, FRG2, ANT1, FAT1* [Cabianca et al 2012]. This discovery provides a possible mechanism causing disease. Unifying all the obeservations, another hypothesis is that the combination between inefficient chromatin silencing of the D4Z4 repeat and the permissive polymorphism that stabilize DUX4 transcript (4qAPAS), may determine an inappropriate expression of DUX4 protein with toxic effect. Despite several efforts in genetic and molecular research, at date, the molecular mechanism subtending FSHD is still to be determ.

1.4.3 FSHD2 and epigenetic hypothesis

Parallel studies on the genetic and phenotypic variability led to another definition that characterizes the disease in a second form: FSHD2, which represents subjects with > 10 U D4Z4, phenotypically indistinguishable from subjects of the classical form carried of reduced size. Scionti et al in 2012 defined 20% of FSHD probands do not carry a deleted < 35 Kb D4Z4 repeats array at 4q35, leading to increase difficulty in performing and interpreting prenatal diagnosis and use molecular data for prognosis. The definition of FSHD2 was proposed in 2010: a patient fulfilling all the clinical diagnostic criteria for FSHD, displaying normal sized D4Z4 arrays on both 4q alleles [Tawil et al, 2010]. Therefore, it has been suggested that similar epigenetic and molecular mechanism are involved. DNA methylation analyses were used in FSHD for the first time in 2001, when D4Z4 DNA methylation level was tested in FSHD lymphoblasts and somatic tissues (including skeletal muscle) at four CpG methylation-sensitive restriction sites (SmaI, MluI, SacII, EagI), without discriminating from chromosome 4q and 10q [Tsien et al, 2001]. In 2003 the methylation sensitive restriction sites BsAI and FseI in the most proximal repeat of the D4Z4 array has been shown to be representative for the entire array. Moreover, using this methodology it can be assessed the methylation level of the first D4Z4 repeat in both chromosomes 4, distinguishing from 10q D4Z4 repeat, when no translocation events are present. FSHD1 individuals showed low methylation levels in both sites (roughly 35% and 33% respectively), while FSHD2 subjects showed lower level of methylation at the BsaAI site (29%)

and strong demethylation at the FseI site (10%) [van Overveld et al, 2003]. In 2009, De Greef et al showed that the levels of D4Z4 hypomethylation were not correlated with phenotype severity, but parallelly presence of at least one permissive 4A161 haplotype on D4Z4 repeat may be two mechanisms of FSHD2 development [de Greef et al, 2009]. In 2010, the same groups have performed a cross-sectional study on 33 patients showing differences of almost 10 years than FSHD1 in median age at onset and a gender difference in disease severity in FSHD2 was not observed. The analysis about hereditary pattern showed a majority of sporadic cases (67%) and 2 seemed recessive in inheritance (sibling pairs). In 2012, whole-exome (WES) was performed in 19 different families where FSHD subjects showed a threshold of 25% of D4Z4 methylation at the *Fse*I restriction site: a new heterozygous mutation in SMCHD1 gene (*Structural Maintenance of Chromosomes flexible Hinge Domain containing 1*) was identified in 15/19 (79%). SMCHD1 is a highly conserved gene, positioned on chromosome 18p11.32 in human genome and it is a member of family proteins responsible of chromosome condensation, cohesion, recombination, DNA repair and epigenetic silencing of gene expression [Lemmers et al, 2012]. Supposing a role in D4Z4 hypomethylation, Sacconi et al selected 6 individuals from a cohort of 53 unrelated FSHD1 families based on D4Z4 methylation level <25% and the presence of 8-10 D4Z4 RU at 4q35. Three out of six were found to harbor a mutation in SMCHD and present severe clinical phenotype and 4q35 allele of 9 D4Z4 units. It was proposed that these mutations can exacerbate the phenotype in FSHD1 [Sacconi et al, 2013]. In 2015 the analysis of two cohort of FSHD patients 1) 55 with subjects carrying more than 11 D4Z4 repeats units (RU) on 4q and 2) 40 subjects with less than 11 D4Z4 RU revealed that SMCHD1 is mutated in 11 subjects in the first cohort and in 2 in the second. In all 13 SMCHD1 mutated subjects was found also a permissive 4qA161 allele with *Fse*I methylation level ≤25%. The co-segregation of these 3 elements was proposed at the basis of FSHD2 development [Larsen et al, 2015] and it is supposed that SMCHD1 mutations account for ~85% of FSHD2 cases [Daxinger et al, 2015]. Up to date over 80 potentially pathogenic variants have been found in FSHD2 patients (Leiden Open source Variation Database, LOVD database). These variants are not clustered in a specific region, instead they are positioned across the gene and suggest haploinsufficiency to be a candidate disease mechanism. Recent study conducted in patients with arhinia [Shaw et al, 2017], a rare developmental malformation causing absence of the nose and in several cases accompanied by other cranio-facial developmental defects, showed that 32 out of 38 probands display a rare missense variant in SMCHD1. However, none of the analyzed subjects showed any FSHD clinical feature but sodium-bisulfite sequencing methylation analysis on D4Z4 arrays on 4q35 and 10q26, performed in Bosma patients revealed that 74% of arhinia subjects carrying a SMCHD1 specific missense variant displayed D4Z4 hypomethylation pattern similar to FSHD2. A digenic inheritance pattern was

proposed for FSHD2, but it was also claimed that other modifier loci may affect D4Z4 chromatin structure, giving rise to a complex picture. Already in 2015, a study on 60 FSHD2 families identified 9 families and a total of 17 affected individuals without SMCHD1 mutations, despite an average low level of D4Z4 methylation [Lemmers et al, 2015]. In 2016 another study] pointed out that few FSHD2 cases carried heterozygous mutations in the DNMT3B (DNA methyltransferase 3 beta) gene, which is responsible for the establishment of the proper de novo cytosine methylation profile during mammalian development [van der Boogard et al, 2016. Parallelly, whether in FSHD1 and FSHD2 patients was observed loss of marks of unexpressed heterochromatin such as histone H3K9me3, many studies indicate that the D4Z4 locus displays a chromatin structure more similar to euchromatin [Zeng et al, 2009]. Indeed, histone H4 acetylation levels were found at the p13E-11 region immediately proximal to D4Z4. Interestingly in all the proposed model epigenetic changes such as methylation or histone modifications are used as an additional level of complexity, but these observations favored the hypothesis that this region might be more dynamic than expected. Recently, a large cohort study of 122 index cases and 110 relatives demonstrated that the mean D4Z4 methylation levels do not permit to discriminate between different clinical phenotypes and that D4Z4 methylation test has a low positive predictive value for the identification of the disease [Nikolic et al, 2020]. Moreover, an accurate review of literature shows lack of homogeneity in the design, techniques and phenotypic evaluation of the subjects analyzed in the different studies [Saksi et al, 2020].

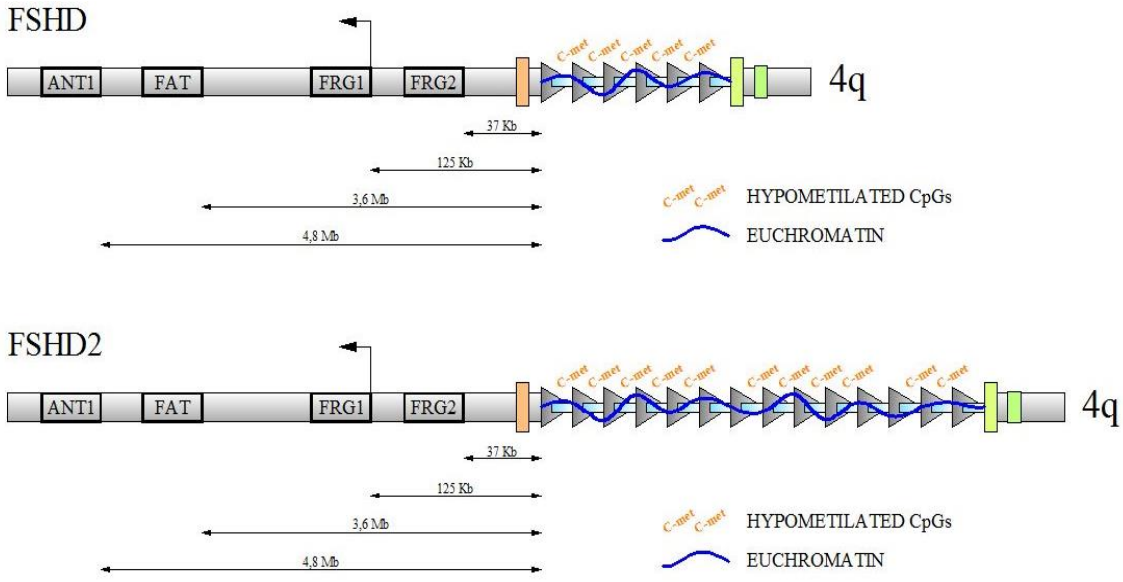


Figure 6. Model of the molecular mechanism of FSHD1 and FSHD2

1.5 Genetic epidemiology and clinical variability

1.5.1 Correlation phenotype-genotype

Since the discovery of the FSHD molecular defect, several genotype-phenotype studies were conducted in order to evaluate how the D4Z4 array and its size are correlated with the pathogenesis of FSHD and their impact on the phenotypic expression and severity. The study of Lunt and coworkers in 1995 [Lunt et al, 1995] were the first important report of genotype-phenoptype analysis FSHD patients. Results was obtained on 14 FSHD families and 25 clinically isolated cases, 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 significant correlation between proband age at onset and D4Z4 fragment size: smaller fragment sizes is associated at earlier age at onset and most severe phenotypes which also involved lower limb muscles. Interestingly, the very short alleles (< = 18 kb, 3 RU) occurred mostly in isolated cases, while "typical" families showed autosomal dominant inheritance. Moreover, in those families in which a D4Z4 allele with size >30 kb segregates, researchers reported a FSHD penetrance lower than the previous estimate of 95% above age 20. 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 in the correlation age onset and D4Z4 size. In 1999, Ricci et al confirmed how more severe phenotypes inversely correlated to D4Z4 fragment size in their group of 165 patients. 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 was observed in 100% of subjects with D4Z4 allele of 10 to 13 kb (1-2 RU), in 53% of patients with a fragment size of 16 to 20 kb (2-3), and in 19% of patients with a fragment size larger than 21 kb (>4). Also, the authors found 3 somatic mosaics and 7 carriers of D4Z4 fragment larger than 20 kb who were unaffected ("nonpenetrant carriers") [Ricci et al 1999]. A study conducted on 1-3 DRA carriers [Nikolic et al, 2016] confirmed that subjects with 1-3 U developed a more severe phenotype in comparison with patients carrying 4-8 DRA. The study has been conducted in FSHD families carrying 1-3 DRA, as well as in carriers of *de novo* 1-3 DRA rearrangements (the parents carry D4Z4 alleles of normal size and 1-3 DRA appears in proband for the first time). The study showed that not all carriers of 1-3 DRA are affected, and that the early onset of FSHD in the proband is not necessarily linked to a severe outcome of disease. Among the population of this study, the majority of cases presenting disease onset in the first decade of life are carriers of *de novo* rearrangements; among familial carriers' onset is equally distributed between the first and the second decade of life.

Confirmation that controls or healthy subjects carrie the reduced D4Z4 came from a van Overveld study in 2000, which showed the percentage of 3% in blood donors. In 2003, Butz and coworkers conducted a systematic study of 39 unrelated FSHD patients with 32-51 kb (7-13 RU) D4Z4 repeat and 102 healthy controls, in order to identify the molecular diagnostic cut-off point between FSHD cases and the control population. Consistently with previous studies a broad myopathic spectrum with four phenotypes (typical FSHD, facial-sparing FSHD, FSHD with atypical features, non-FSHD muscle disease) was found in the borderline region too. Three symptomatic subjects carried D4Z4 beyond diagnostic range (45-51 kb), of those 2 displayed a typical FSHD phenotype. The results of that study indicated that there was no definite D4Z4 diagnostic cut-off point separating FSHD, FSHD-like myopathies and controls. Typical FSHD phenotype was found in 24/39 subjects, facial-sparing FSHD in 6/39, atypical FSHD in 4/39, non-FSHD in 5/39, none above 41 kb. Therefore, the authors suggested the D4Z4 cut off should be 8 repeats or 35 kb [Butz et al, 2003]. All these observations suggest that, despite the presence of a correlation with D4Z4 size, other genetic and environmental factors might play a role in FSHD development. The numerous non-penetrant cases suggest also a different inheritance model, for instance a polygenic pattern. Indeed, before the discover of molecular signatures, large families with FSHD clinical diagnosis suggested autosomal dominant inheritance pattern, with almost complete penetrance. In the first analysis of inheritance, Lunt and colleagues in 1995 observed in FSHD families carrying of 30 kb and 34 kb fragment size, subjects unaffected above age 20 years, age with estimated penetrance of 95% in a previous study, hypothesizing that this percentage is lower in families with >30 kb. Zatz and colleagues in 1998 investigated a sample of 172 affected subjects from 52 families with fragment smaller than 35 kb, affirming that penetrance was 85% for patients until age 30. Moreover, FSHD affects males more severely and those women are more often asymptomatic then men, as already indicated by other groups. In the same study, 3 families presented affected siblings born from asymptomatic parents suggesting a recessive pattern of inheritance. Years later, the percentage of individuals, related to FSHD patients carrying a DRA, which remained asymptomatic or minimally affected was estimated to 20% [Tonini et al. 2004]. The most recent work, in 2012, reports the clinical and genetic analysis of 133 individuals from 71 Greek families, revealing almost 50% of asymptomatic relatives carrying a contracted 4q allele, older than 30 years and the percentage of unaffected carriers was lower among males. Moreover, clinical phenotypes were discordant in two female monozygotic twins with de novo 13 kb deletion; this finding was coherent with other previously reported sporadic cases of male monozygotic twins [Sakellariou et al. 2012]. Tupler and colleagues in 1998 already have described two monozygotic twins with different phenotype, one being almost asymptomatic and the other severely affected, with a de novo fragment of about 23 kb, which originated from a mutational event occuring during paternal

gametogenesis, as demonstrated by haplotype segregation studies. In 2013, Ricci et al have described a study on a large Italian cohort of patients carrying alleles of 1-8 D4Z4 units (167 index cases and 363 relatives) according to a standardized clinical score. They showed that among carriers of a 1-3 units D4Z4 reduced allele (1-3 DRA) a high percentage of individuals are affected, while this percentage lowers among carriers of 4-6 units DRA and in these families the diagnosis of FSHD was reported only in one generation; in general, 32,2 % of relatives were not affected. They demonstrated the phenotype and penetrance were influenced by the degree of relationship with the proband and male relatives has a mean score significantly higher than females. Moreover, developing motor impairment was evaluated in correlation with D4Z4 size and age at onset: FSHD penetrance is almost complete (96,2%) by age 60 only for carriers of DRA with 1-3 U. [Ricci et al, 2013]. This was the first investigation that provided a clear insight into a great number of FSHD families. In 2020, Ruggiero et al conducted a large genotype-phenotype study in borderline subjects to provide guidance for FSHD diagnosis. The clinical characterization of 134 index cases revealed their phenotypic heterogeneity; seventy-three (54.5%) displayed the classic FSHD phenotype, 9 subjects (6.7%) did not show motor impairment, whereas 20.1% present incomplete phenotype with facial weakness or scapular girdle weakness, 18.7% of index cases show more complex phenotypes with atypical clinical features. The study demonstrates that however there is a natural tendency to look for a relationship between number of repeat units and clinical severity, this approach has presented several flaws through years. The detection of borderline DRA should be considered as a genetic susceptibility condition rather than a diagnostic marker. The large cohort study confirmed that the diagnostic value of borderline DRA is poor and, therefore, the recurrence risk cannot be estimated. Finally, several studies conducted on FSHD families have reported patients presenting “double trouble” phenotype, where FSHD specific features are concomitant with symptoms characteristic for other myopathies or independent genetic mutations. Some examples are: Beckery dystrophy/FSHD, Duchenne/FSHD, LMGD24FSHD, multiple sclerosis/FSHD, myasthenia gravis/FSHD etc. This suggests the possibility of a synergistic effect of simultaneous mutations in modulating the clinical expression and not exclude the involvement of D4Z4 reduced array in other myopathies [Lecky et al, 1991; McGonigal et al, 2002; Chuenkongkaew et al, 2005; Filosto et al, 2008; Korngut et al, 2008; Rudnik-Schöneborn et al, 2008; Ricci et al, 2012; Masciullo et al, 2013].

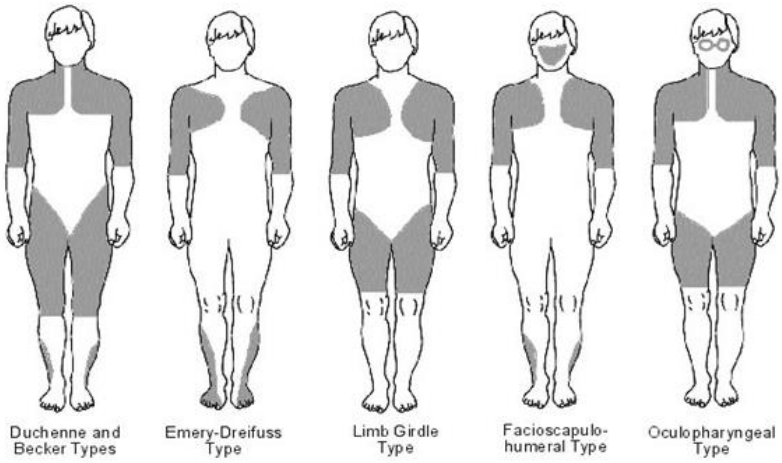


Figure 7. Main area of muscle weakness in different types of dystrophies

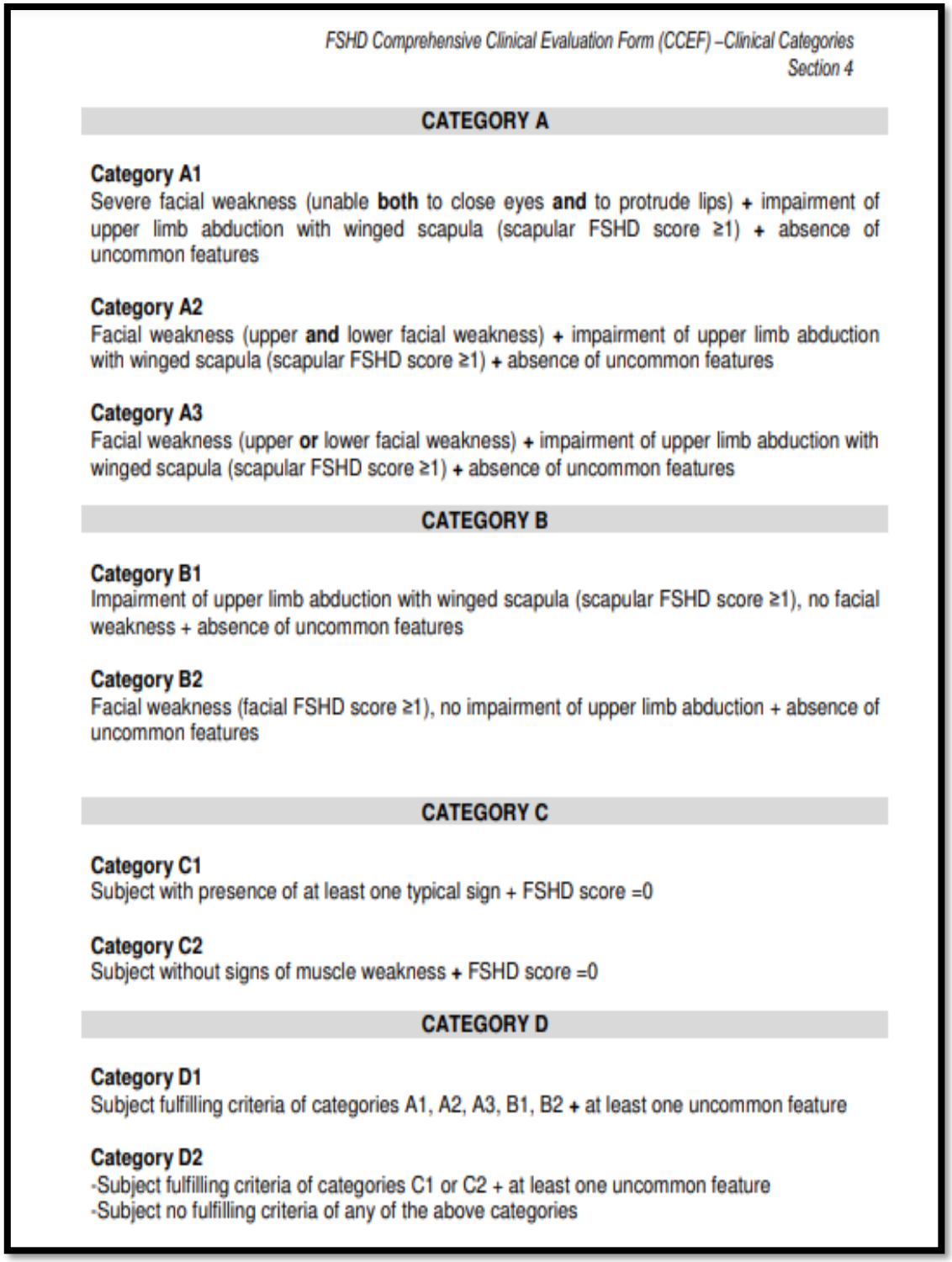
1.5.2 The Italian National Registry for FSHD (INFR) and CCEF tool: a prototype

The Italian Clinical Network for FSHD (ICNF) was born in Modena in 2007 and is composed by a diagnostic laboratory (Miogen Lab) at University of Modena and Reggio Emilia, and by 14 clinical centers, with expertise in diagnosis and management of neuromuscular disorder, across all of Italy. In the same year, ICNF established the Italian National Registry for FSHD (INFR) with the purpose of collected DNA samples to conduct molecular diagnosis of FSHD, anamnestic data and main clinical information of patients and their families. Dedicated specific software and a website for data management has been designed registrating patients visited from 1993. The neuromuscular Clinical Centers are the following: Department of Neurology, IRCCS Fondazione Ospedale Maggiore Policlinico, University of Milan (Prof. Maurizio Moggio) IRCCS Foundation, C. Besta Neurological Institute at Milan (Dr Lorenzo Maggi), Department of Neuroscience, University of Turin (Prof. Tiziana Mongini), Department of Neurosciences, University of Padova (Prof. Elena Pegoraro), IRCCS S. Camillo at Venice (Prof. Corrado Angelini), IRCCS "C. Modino" Foundation, University of Pavia (Dr Angela Berardinelli), Department of Neurological Sciences and Vision, University of Verona (Dr. Gaetano Vattemi), University Hospital "Spedali Civili" of Brescia (Dr. Massimiliano Filosto), Department of Clinical and Experimental Medicine, University of Pisa (Prof. Gabriele Siciliano), Department of Neurology, S. Andrea Hospital. "Sapienza" University of Rome (Prof. Giovanni Antonini), Center for Neuromuscular Disease, University "G. d'Annunzio of Chieti (Dr. Antonio Di Muzio), Department of Neurological Sciences, "Federico II" University of Naples (Prof. Lucio Santoro), Department of Neurosciences, Psychiatry and Anaesthesiology University of Messina (Prof. Carmelo Rodolico), ASL8 University of Cagliari (Prof. Giovanni Marrosu).

The Italian Clinical Network centers

From 2007 to 2016 each subject was examined by a neurologist using a standardized protocol that allow to define numerically the clinical severity of the disease, the "score". The evaluation procedure allows to assess the strength and the function of six muscular groups specifically affected in FSHD belonging to 1) 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). [Lamperti et al. 2010]. 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. Until that moment, the Italian Network for FSHD successfully used the FSHD clinical score as a tool in genotype/phenotype correlation and genetic counseling. All this work has permitted to observe the large variety of phenotypes and the need to distinguish them despite some common characteristics. In 2016 a new tool was created to facilitate the description of the combination of different features among and within FSHD families: the Comprehensive Clinical Evaluation Form (CCEF). It has been designed to provide a novel, specific, inter-rater reproducible method to describe the clinical features of a FSHD patient [Ricci et al, 2016]. The first section of CCEF describes the patient's clinical history, disability, medical conditions and particular habits. The second section, the *Evaluation Form* evaluates onset, trigger events, segmental muscles involvement using the Medical Research Council scale too and includes also information about the presence of typical and atypical signs for FSHD. The third section, the *Evaluation Scale* translates muscle weakness into FSHD score, maintained from first clinical evaluation protocol. The fourth section (the *Clinical Diagnostic Form*) both the score and the muscle weakness are summarized, leading to the classification of patients in *Clinical Categories*, which are described in the fourth section of CCEF. Category A defines typical FSHD phenotypes (presenting facial and scapular girdle weakness), through the three sub-categories A1, A2 and A3 based on severity of facial weakness. Category A1 is defined to distinguish some infantile forms or more severe

phenotypes from the vast majority of patients in which it was observed a milder facial involvement (categories A2 and A3) [Nikolic et al, 2016]. A1 patients are characterized by an early and prominent weakness of *orbicularis oculi* and *oris* with facial diplegia and dysarthria. So, this distinction should facilitate the identification of a specific clinical group deserving *ad hoc* studies. Category B includes the incomplete phenotypes with only scapular girdle weakness (impossibility in upper limbs abduction and winged scapula) or only facial weakness, respectively in the sub-categories B1 and B2. These categories are identified because, for instance, an isolated scapular girdle muscle weakness can be observed in FSHD relatives, but it can be also related to other myopathic disorders or nerve injuries. Category C differentiate non penetrant carriers, for subjects presenting signs suggestive for FSHD and the healthy subjects with no signs of muscle weakness, respectively in the sub-categories C1 and C2. The last category, D, comprises FSHD patients presenting additional myopathic signs from or non-muscular features in the sub-category D1 or non-FSHD patients (not fulfilling the diagnostic criteria for FSHD) in category D2. Atypical features were chosen based on evidences from the literature [Ricci et al, 2014]. This category may facilitate the discovery of factors that contribute to the disease expression or identify those subjects who are wrongly considered FSHD because of a diagnostic bias due to the random finding of DRA. The categories outlined by the CCEF may assist diagnosis, genetic counseling and natural history studies. Furthermore, the CCEF categories could support selection of patients in randomized clinical trials [Ricci et al, 2016]. In 2020, the first 5-year clinical follow-up study from the Italian National Registry for FSHD was conducted in 246 FSHD1 patients, divided between index cases and carrier relatives. The disease progression was measured as a variation of the FSHD score, calculated with CCEF tool. Disease worsened in 79.4% (112/141) of index cases versus 38.1% (40/105) of carrier relatives and advanced more rapidly in index cases, with FSHD score 2.3 versus 1.2 respectively. The 79.1% (38/48) of asymptomatic carriers remained asymptomatic. The highest FSHD score (1.7) was found in subject with facial and scapular weakness at baseline (category A), whereas in subjects with incomplete phenotype (facial or scapular weakness, category B) had lower FSHD score (0.6) $p < 0.0001$. The study has confirmed the strong prognostic effect of the CCEF categories in FSHD1 patients.



FSHD Comprehensive Clinical Evaluation Form (CCEF) –Clinical Categories
Section 4

CATEGORY A

Category A1
Severe facial weakness (unable **both** to close eyes **and** to protrude lips) + impairment of upper limb abduction with winged scapula (scapular FSHD score ≥1) + absence of uncommon features

Category A2
Facial weakness (upper **and** lower facial weakness) + impairment of upper limb abduction with winged scapula (scapular FSHD score ≥1) + absence of uncommon features

Category A3
Facial weakness (upper **or** lower facial weakness) + impairment of upper limb abduction with winged scapula (scapular FSHD score ≥1) + absence of uncommon features

CATEGORY B

Category B1
Impairment of upper limb abduction with winged scapula (scapular FSHD score ≥1), no facial weakness + absence of uncommon features

Category B2
Facial weakness (facial FSHD score ≥1), no impairment of upper limb abduction + absence of uncommon features

CATEGORY C

Category C1
Subject with presence of at least one typical sign + FSHD score =0

Category C2
Subject without signs of muscle weakness + FSHD score =0

CATEGORY D

Category D1
Subject fulfilling criteria of categories A1, A2, A3, B1, B2 + at least one uncommon feature

Category D2
-Subject fulfilling criteria of categories C1 or C2 + at least one uncommon feature
-Subject no fulfilling criteria of any of the above categories

Figure 8. Description of FSHD clinical categories from the Clinical Comprehensive Evaluation Form (CCEF)- Ricci et al, 2016

AIM

This study aims to show the value of centralizing molecular and clinical diagnosis of rare diseases, in order to distinguish a patient from others with similar Mendelian conditions and lay the basis for personalized treatments. For this, it is necessary to obtain big data about epidemiological context and standardrized phenotypic spectrum. The present study collected molecular and clinical informations on 7550 subjects registered in the Italian National Register for FSHD (INRF), established by the Italian Clinical Network for FSHD (ICNF) consisting of diagnostic laboratory Miogen Lab (University of Modena and Reggio Emilia) and 14 clinical centers of all Italy. In the present work the whole cohort (1993-2020) has been divided in 4 different sets based on the evolution of the data collection, which started with results of molecular analysis (1993) till the creation of Comprehensive Clinical Evaluation Form (CCEF). The work analyzes the outcomes and the benefits of this approach in clinical practice and in genetic counseling and how the Italian National Registry for FSHD can be considered a prototype for the study and collection of big data in rare diseases. The study evaluated penetrance and highlighted clinical differences between carriers of the D4Z4 small-sized alleles (DRA) and also compared to subjects who carry D4Z4 alleles of normal size to clarify the role of the reduction of D4Z4 elements. For an integrated approach aimed to identify alternative genetic mechanisms, family trios with a pattern of inheritance compatible to an autosomal recessive trait were studied. Considering important and efficient impact of using new generation tools for neuromuscular diseases study, WGS technology has been exploited. In the context of genetic counseling, we evaluated relevant aspects of prenatal diagnosis. In particular, we reconsidered the feasibility of preimplantation genetic diagnosis (PGD) testing microsatellite D4S1523 and two VNTR proximal to D4Z4, studying recombination in three generations families with phase. This work represents the implementation of precision medicine, which by integrating clinical and genetic elements can create a phenotypic ontology for FSHD.

2. MATERIAL AND METHODS

2.1 Subject recruitment and analysis

The study was performed on the families accrued through the Italian National Registry for FSHD (INRF), established in 2007 by Italian Clinical Network for FSHD (ICNF). The cohort selected includes all subjects, probands and relatives, registered from 2007 to December 2020 considering that the first people are visited and diagnosticated already in 1993. For all subjects has been performed molecular diagnosis of FSHD at Miogen Lab (University of Modena and Reggio Emilia). The whole cohort is divided in 4 sets that covered different years of cohort, respectively 1993-2006, 2007-2013, 2014-2016, 2017-2020. The data collected for each subject include diagnostic analysis, specifying rearrangements and peculiarity, clinical evaluation from score to CCEF by neurologist with long standing expertise in diagnosis and management of neuromuscular disorders. The families have been descripted, including penetrance and specific characteristic inheritance. For each set of data, an application was created consisting of several spreadsheets capable of grouping and analyzing the data according to multiple logics that vary from time to time, depending on the objectives and purposes of the analysis. This application has been designed to be able to highlight, intersect and relate to each other as much data and significant information as possible contained in the databases. This means represented a tool which, going alongside the analysis and study of traditional data, made it possible to verify, control and enrich the results of the latter. According to Declaration of Helsinki, informed consent was obtained during the compilation of score evaluation and the CCEF for research purpose and publication.

2.2 Standard molecular test for FSHD

- Lymphocytes are isolated from whole blood: 15 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 1 X 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 and 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 40ul. Avoiding formation of air bubbles is recommended.
- An equal volume of 1% Low Melt Agarose in PBS 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 (Bio Rad). Plugs are placed at 4°C in order to allow the agarose plugs to solidify for 20-30 minutes. Each well, in Bio Rad's disposable plug mold, holds approximately 80 ul of samples and contains 2x10 cells.
- DNA isolation in agarose blocks: plugs are gently transferred into a 6 well 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 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 (40u g/ml with TE PH 8.0 - 1ml/plug), are incubated at 50°C for 30 minutes.
- PMSF solution is removed, plugs are washed for 5 minutes with distilled sterile water 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. 2 of plug (corresponding to 4,0x10° 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 and XapI: 6U/ul (final volume: 200 ul), overnight at 37°C. 1 plug is digested. Binl: 0,25U/ul (final volume: 200 ul), 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 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 is 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 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 ug/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 nalgene tube. An equal volume of phenol equilibrated with 0,1M Tris-HCl (pH 8.0) is added and mixed by gentle agitation for 10 minutes, at room temperature. To separate the aqueous phase, the sample is centrifuged for 10 minutes at 4500 g.
- 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: isoami lic alcohol (24 25:1) is added to the aqueous phase. Samples are then mixed for 10 minutes, centrifuged for 10 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 0,4% agarose gel electrophoresis (LGE): 14 ug 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 Blnl 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 ug 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 64 hours.
- After electrophoresis, the gel is stained with ethidium bromide, photographed under UV light. 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 PFGE and 0,4% 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, 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 capillary transfer ZetaProbe GT Nylon membrane (Bio Rad) 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 (3 mm) 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 16 hours for PFGE, 8 hours for 0.4% agarose gel.
- 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 1 h and it can be used for hybridization.
- Probe labeling: Add 25 ng of probe p13E-11 (D4F104S1) and water to a final volume of 11 ul to a clean tube. Denature DNA by heating it to 100°C for 10 minutes and chill quickly in ice.
- Add 4 ul High Prime DNA labelling Kit (Roche), dNTP (less dATP) and 2ul [α^{32} P] dATP at 37°C for 12 minutes. Stop reaction with 0,02 M EDTA
- Remove non-incorporated deoxyribonucleoside-triphosphates applying the mixture on a Quick Spin Column provided by the kit.
- Hybridization Membrane is incubated in 10 ml of pre-warmed pre hybridization solution "PerfectHyb Plus hybridation buffer" (Sigma) for 10 minutes at 68°C in a roller bottle.
- Add salmon sperm DNA to the radio-active labeled probe that is denatured by heating at 100°C for 5 minutes and chilled quickly in ice.
- Discard the pre-warmed pre hybridization solution and add probe (1-3X109 cpm/ml to membrane in warmed pre-hybridization solution and incubated for 4h at 68°C.
- Hybridization solution is removed and membrane is washed twice with solution 2X SSC with 0,1% SDS for 30 minutes, at room temperature.
- 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 autoradio graphic image.
- To obtain the autoradiograph, soak the X-ray film for a few minutes in developer solution and then in a fixer solution. Wash the film in water
- The autoradiograph will display the RFLP obtained by hybridization of sample DNA and the radioactive probe.

2.3 Whole Genome Sequencing analysis (WGS)

2.3.1 Next-generation sequencing approaches for the diagnosis

Next generation sequencing technologies have changed the way to approach the analysis of genomes. The ability to provide, at low cost, millions of DNA sequences per run, and reduction of times

furnitured by these tools have had repercussions in various disciplinary areas, such as cancer research, metagenomics, animal research and organic selection. Indeed, the introduction of next generation sequencing technologies has revolutionized the approach to complex and genetically heterogeneous disorders, including skeletal muscle disorders, made sequencing of large genes possible in patients and has associated the clinical phenotypes with already known disease genes [Savarese et al, 2020]. Three different NGS strategies for DNA sequencing have been developed: a targeted resequencing of specific regions/genes of interest, a whole exome sequencing (WES) targeting the entire set of human exons (about 2% of the human genome) and a whole genome sequencing (WGS) that is theoretically able to investigate the entire human genome. Exome and genome sequencing will probably soon be adopted as a first-tier test for all genetically heterogeneous conditions and genetic data may subsequently be available to clinicians during the first evaluation of a patient [Nigro et al, 2016] Bioinformatics in this development process was fundamental. Thanks multiple algorithms have been developed to it, based on different techniques, to optimize analyzes on increasingly vast amounts of data and provide a unique sequence output. Few tens to several hundred bases, they come combined together, usually into a FASTQ, a type of file used in biology to memorize genetic sequences and related quality scores, namely the score that the algorithm assigns to the sequence and that comes then used to choose the best match against the reference genome. At this point, the bioinformatics analysis is divided into three steps: 1) alignment, attempt to align the reads to the reference genome, defined the variant calling 2) filtering and 3) annotation, which attempts to separate differences due to genetic errors from instrumental errors performed in analysis and give a definitive file [Dolled-Filhart et al 2013]. NGS applications are revelead useful to improve the diagnosis in myopathic patients, considering the huge phenotypic variability observed in muscular disorders, the large size of many of the associated genes and the absence of a clear-cut genotype–phenotype correlation [Nigro et al, 2014]. WGS will soon represent the ‘gold standard’ as the unique and comprehensive genetic test that replaces any other genetic analysis. Indeed, sample preparation is not biased by differences in capture, and PCR artifacts can also be reduced or canceled using PCR-free protocols. Moreover, WGS is also able to detect structural rearrangements [Pirooznia et al, 2015]. WGS has not been applied to a diagnostic setting so far: its cost and, above all, the challenges related to the interpretation of the molecular findings are still hampering its routine diagnostic use [Alyass et al, 2015]. More recently, large panels investigating hundreds of neuromuscular genes have been set up and they are now increasingly being used in the diagnostic setting. In fact, they include all relevant neuromuscular genes and potential strong candidate genes, it represents a cost-effective strategy and for this it is applicable to a large number of patients and unaffected relatives. About choose of samples, trio analysis (an affected proband with both parents) is always ideal beacuse increases the detection rate

and the accuracy of the test, improves variant calling, allows the rapid identification of de novo variants as well as of sequencing errors and, above all, provides phase and haplotype information [Savarese et al, 2014, Savarese et al, 2016, Sevy A et al, 2016, Vasli N et al 2012].

2.3.2 Families trio analysis

Trio analysis reduces false positive calls and enables the prioritisation of potential disease-causing variants, with efficiency of 5-10% increase compared to analyzing proband only. Total, 137 Trio (or subjects) have been identified with this pattern of inheritance in cohort 1993-2020, but not for all both parents have been analyzed. For WGS analysis, 4 trio are selected: they are composed to asymptomatic parents carrying or not of molecular defects (DRA) and probands/sons with classic phenotype (category A). Trio with somatic mosaicism are excluded. WGS analysis has been performed with Novaseq 6000 (Illumina) that leverage the proven technology of the NGS (Next-Generation Sequencing, NGS) Illumina. NovaSeq 6000 system has a scalable and effective processivity, and is suitable for practically every study need. Applications that require a large amount of data, such as the sequencing of the whole human genome (Whole-Genome Sequencing, WGS), extremely exome sequencing deep and tumor-normal profiling can be completed more cost-effectively. Users can choose and combine four types of flow cells (SP, S1, S2 or S4), process one or two flow cells simultaneously and choose between different reading lengths to easily customize the output and processivity of samples for each sequencing run. For our analysis has been used the NovaSeq S4 flow cell, which allows cost-effective sequencing for a range of applications, thus studies WGS are an option interesting and accessible for a larger number of laboratories. Combined with flow cells at lower output, the same tool allows you to execute methods with fewer data. The NovaSeq 6000 system offers outputs up to 6 Tb and 20 billion readings in less than two days. The major flexibility is offered by the NovaSeq Xp workflow which supports the loading single lanes to sequence different libraries in the flow cell. The multiplex analysis also reduces the amount of DNA input required relative to the flow of standard work. As regards the preparation of the libraries, were prepared at the CNR-IBIOM under directing dr. Tullo Apollonia. The Illumina DNA PCR-Free Library prep tagmentation protocol was followed, starting from 600 ng of nDNA, assayed with Qubit (fluorimetric assay). The concertation of the library pool x sequencing was 3nM. The bioinformatics analyzes are coordinated by Prof Pesole at University of Bari, Puglia and Phd Matteo Chiara at University of study of Milan, Lombardia. Variant calling was performed by the CoVaCS pipeline [Chiara et al 2018] The ANNOVAR software was employed for variant annotation [Wang et al 2010]. The following annotation resources were considered for the estimation of allele frequencies: ExAC (version 1.0 updated 27 February 2017) [Lek et al 2016], 1000 Genomes (phase 3) [1000 Genomes

Project Consortium et al, 2015], gnomAD (version 3.0, updated 15 August 2020) [Karczewski et al, 2020], dbSNP (build 151) [Sherry et al, 2001], Kaviar (version 160204-Public) [Glusman et al 2011]. RefSeq (release 200) [O'Leary et al 2016] was considered for genes and transcript annotations, ClinVar [Landrum et al 2014] (updated May 1rst 2021) for the annotation of disease-causing variants, and the dbNSPF [Liu et al 2016] (v4.2a, updated 10th July 2021) 25 database for the evaluation of the effect of nonsynonymous substitutions. DeNovoGear [Ramu et al 2013] was applied for the identification of potential deNovo mutations.

2.4 Study of D4S1523, D4S163, D4S139 markers for PGT

The classical linkage analysis of the parametric type (also known as model-based analysis) represents a classical statistical methodology of genetic analysis and is based on identification of markers (mostly microsatellite) along a chromosome that segregate together with gene that influences the phenotype of interest or the disease within families. In parametric analyzes it is possible to assume a Mendelian inheritance model dominant or recessive and therefore the gene frequency, penetrance and the phenocopy of each genotype. However, if more genes each are involved in complex traits in which potentially could have different parameters and therefore, it is impossible to define these values correctly. To work around this problem, it is possible to conduct an analysis of linkage without models, called non-parametric, which searches for the presence of common alleles between affected individuals of the analyzed sample consisting of nuclear families (ASP-Affected Sib Pairs) or multigenerational (APM-Affected Pedigree Member). The biological basis is the concept of recombination or crossing-over, thanks to which during the first meiotic phase occurs the physical exchange of genetic material between homologous chromosomes paternal and maternal. If a crossing-over occurs between two loci, then the alleles that the child receives from the parent are said "recombinants" as they are not longer identical to those transmitted by the parent himself. In practice, what is observed is the degree of independence transmission of a marker with respect to another, and this reflects the degree of recombination between those markers. The percentage of recombinant progeny correspond the recombination fraction, θ. According to this parameter, if two markers are independent (not in linkage), the recombination fraction θ is maximum ($\theta = 0.5$), otherwise the recombination fraction progressively decreases with increasing linkage state of the two markers under consideration. When the families are large and the recombinants can be counted, the analysis is simple. However, families with interesting illnesses are rarely large enough to achieve statistically significant results from a single family. In addition, the imperfect structure of human families means that often recombinants cannot be identified without ambiguity. An alternative method that applies in these cases is the computer generated LOD Score calculation. The LOD Score is the logarithm to

base 10 of the ratio between the probability that two loci are associated, as a function of θ, and those that are not with θ = 0.5. Being a function of the recombination fraction, the LOD scores are calculated for a range of values of ϑ and the maximum estimated Z value. The overall probability of linkage in a set of families is the product of the probabilities in each individual family. The significance threshold of linkage is fixed for LOD score ≥ 3 and, since the value of LOD score is the logarithm to base 10 of the probability in favor of linkage, corresponds to the 1000: 1 probability in favor of linkage [Morton, 1955]. If two loci are found far from each other, then the LOD scores will be mostly negative. In a parametric linkage analysis, studied family trees (pedigrees) include several generations, ideally from three generations, and different affected individuals from which it is possible to deduce the inheritance model of the disease locus. Only families with informative meiosis, i.e., subjects heterozygous for a marker, and phase-know will be analysed.

$$\text{LOD score } Z(\theta) = \Sigma \log_{10} = \frac{L \text{ (family se } \theta = X)}{L \text{ (family se } \theta = 1/2)}$$

$$H = 1 - \sum_{i=1}^{n} p_i^2$$

Figure 9. LOD score formula as a function of θ (recombination fraction). L: probability that the data observed within a family occur for a given assumed value of θ. (down) H represent heterozygosity, pi^2 is the frequency of homozygous subjects for the allele i and n is total number of alleles identified.

To test the feasibility of using D4S1523 microsatellite and the linkage of two VNTR D4S139 and D4S163, proximal to D4Z4, for Preimplantation Genetic Diagnosis for FSHD, we individuated 12 three-generation families from our Register of FSHD selecting 86 individuals overall. To determine the allele phase, segregation at 4q35 is controlled on subjects of the second generation and successively in the third generation. 1 family was discarded for unclear data, while a second family was rejected as it hampered MLINK program correct functioning. 5 families presented a complete pedigree (all grandparents, all parents and at least one nephew), 3 families missed one parent in the second generation, one family missed one grandparent in the first generation and one family missed one grandparent in the first generation and one parent in second generation. The selection criteria consist of presence of grandparents, presence of both parents and at least one grandchild. De novos are excluded because it would be impossible to define haplotypes.

The microsatellite D4S1523 and the two VNTRs D4S163 and D4S139, proximal to D4Z4, has been studied by capillary electrophoresis and southern blotting, PCR in order to obtain information about their recombinational behavior. Genomic DNA of 75 subjects from three generation FSHD families has been extracted from whole blood, followed as previously described. It has been analyzed the

segregation of these 3 polymorphic markers in 26 fully informative meiosis (phase-know). Information obtained from this approach may be useful in the context of pre-natal analysis (as PGD or PND) as it may provide a wider set of evaluable elements, offering patients the possibility to obtain an earlier and more accurate prediction in genetic counseling.

2.4.1 Microsatellite D4S1523

- PCR: Polymerase chain reaction was performed using Phusion (Thermo Fisher scientific). The Phusion polymerase reaction mix (50 ul) contains:

  -Genomic DNA: 100ng

  -5X Buffer: 10 ul

  -dNTPs 10 mM: 1 ul

  -Forward primer 1 uM

  -Reverse primer 1 uM

  -GoTaq polymerase: 0.5 ul

  -H2O to 50 ul

The Phusion polymerase reaction mix (50 ul) contains:

-Genomic DNA/ PCR product: 50ng

-5X Buffer: 10 ul

-dNTPs 10 mM: 1 ul

-Forward primer 1 uM

-Reverse primer 1 uM

-Phusion polymerase: 0.3 ul

-H20 to 50 ul

To obtain higher specificity for D4S1523 marker analysis, we set up a nested PCR approach using high-fidelity polymerase (Phusion) and a touchdown thermal protocol consisting of:

-Initial denaturation: 94°C for 2 minutes

- X15: -Denaturation: 94°C for one minute

-Annealing: from 69.5°C to 62.5°C for 1 minute with temperature decreasing of 0.5°C per cycle

-Elongation: 72°C for 1 minute

- X20: -Denaturation: 94°C for one minute

-Annealing: 62°C for 1 minute

-Elongation: 72°C for 1 minute

- Final elongation: 72°C for 10 minutes

-4°C for infinite hold

The first PCR cycle was performed starting from genomic DNA to obtain an amplicon of about 1 Kb. This amplicon was subsequently used as template for the nested PCR cycle obtaining an amplicon of about 125 bp. The nested forward primer was labelled with a FAM green fluorophore to make the amplification product detectable by capillary electrophoresis.
Primer sequence of D4S1523 external FW is 5'- GATTACAAATGTGGGAATGAAGGG-3'; D4S1523 external RW is 5'- TGTGACAGAGACAGAGACATAAAG-3'; D4S1523 nested FW is /56-FAMN/5'-AACAACAGCTCACAAAGGAAAG-3'; D4S1523 nested RW is 5'-CTCTCTCTCTCTCTCTCTCATCTATC-3'. To purify PCR products, both QIAquick PCR purification kit (QIAGEN) and QIAquick Gel Extraction Kit (QIAGEN) were used. These kits were used to purify amplicons eliminating non incorporated nucleotides, enzymes and buffer traces and agarose, taking advantage of a chromatographic column. Direct purification of DNA from the PCR mix was performed with QIAquick PCR purification kit:

- Purification DNA from PCR: Add 5 volumes of PB buffer to 1 volume of the PCR reaction and mix.
- To bind DNA, apply the sample to a QIAquick column and centrifugate 1 minute at 13000 RPM.
- Discard the flow-through and place the QIAquick column back in the same collection tube.
- To wash add 750 ul of PE buffer to QIAquick column and centrifuge for 1 minute at 13000 RPM. Then discard the flow-through.
- Centrifuge for 1 additional minute at 13000 RPM.
- Place QIAquick column into a clean 1 ml tube and add 50 ul of water to the center of the column then centrifuge for 1 minute at 13000 RPM.

Alternatively, DNA purification trough gel extraction was performed using QIAquick Gel Extraction Kit:

- Purification DNA from Gel: Excise the DNA fragment from the agarose gel and remove extra agarose to minimize the dimension of the gel slice.
- Collect the slice into a clean tube and weight it. Then add 3 volumes of QG buffer to 1 volume of gel (100 mg of gel is equal to 100 ul).

- Incubate at 50°C for 10 minutes (until agarose is completely dissolved). To help dissolve gel, mix by vortexing every 2/3 minutes.
- Add one volume of isopropanol to the sample and mix.
- To bind DNA, apply the sample to the QIA quick column and centrifuge for 1 minute at 13000 RPM.
- Discard the flow-through and place the QIAquick column back into the same tube.
- Add 500 ul of QG buffer to QIAquick column and centrifuge or 1 minute at 13000 RPM. Discard the flow-through.
- Wash by adding 750 ul of PE buffer to QIAquick column and centrifuge for 1 minute at 13000 RPM.
- Discard the flow-through and centrifuge for an additional 1 minute at 13000 RPM.
- Place the QIAquick column into a clean 1.5 ml tube and add 50 ul of water to the center of the column. Centrifuge for 1 minute at 13000 RPM.

Capillary electrophoresis was performed starting from FAM-labeled PCR product on Applied Biosystems® 3130xl Genetic Analysers. This technique allows the resolution of DNA fragments with a small difference in number of nucleotides.

- Capillary electrophoresis: 1 ul of FAM-labelled PCR and an 0.5 ul of GeneScanTM 500 ROXTM dye Size Standard marker (Thermo Fisher) are distributed in the instrument wells with formamide. The PCR product must be diluted to a concentration suitable to obtain a fluorescence intensity which is detectable by the machine.
- The machine needles aspirate the vial content and pump it into a silica capillary filled with a polymer and start the run.
- Files in fsa format are produced and can be analyzed by Peak ScannerTM Software.

2.4.2 VNTR D4S163 and D4S139

- Digestion of genomic DNA: Analysis of RFLPs detected by probes D4S163 and D4S139 genomic DNA was digested with PstI retriction enzyme. The reaction was settled as follows:

  -Genomic DNA: 5 ug

  -10X Buffer: 2.5 ul

  -Spermidine (100 mM): 3 ul

  -PstI: 25 U

  -H20 up to final volume of 25 ul

  -The reaction was performed at 37°C for 4 hours.

- Southern blotting: DNA fragments obtained by PstI enzyme digestion were resolved by electrophoresis on 0,8% agarose gel in TBE 1X. The following steps are the same as those previously described for 0,4% agarose gel. The ladder used are: Lambda/Hind Ladder (2-23

kb, Promega), 1 kb ladder (1-10 kb, Promega), Lambda Eco/Hind (1-21 kb, Promega). To detect D4S139 and D4S163, it has been respectively used PH30 and LILA probe. The pre-hybridization, probe labelling, hybridization and probe signal detection were performed under condition previously described for p13E-11 probe. Different conditions are only:

- Probe labelling: Probe PH30 and LILA are labelled with 32P-DATP using High Prime DNA labelling Kit (Roche) at 37,5°C for 12 minutes.
- Hybridization: Analyse 40 ul of purified labelled probe with a beta counter and aliquot a volume of labelled probe to obtain 20 to 25 million of cpm.

3. RESULTS

3.1 Molecular features of cohort 1993-2020

The studied population consisted of 7550 subjects, representative of 1366 probands of families, 4224 relatives, 1585 single cases, 375 controls. Probands of our cohort are 2951 and represent the Italian population. Probands carrying deletion repeat array (DRA) are 56,2% (1660), while 42,1% (1781) are heterozygote relatives (fig. 10). Demographics and molecular data are given in Table 1. Previous smaller investigations of the Dutch population suggested that 3% of healthy subjects had D4Z4 alleles of 35–38 kb in size [van Overveld et al, 2000]. In the present study we collected 446 controls from the general Italian population, of which 375 spouses from families and 152 healthy volunteers. We found 25 subjects with 11-41 in control population, so this percentage is 6,6% in Italian population. Subjects with "incomplete analysis" did not have suitable samples to complete the analysis and were not considered. The subjects of 2021, included in 1993-2020 cohort, without analysis, were excluded from all counts.

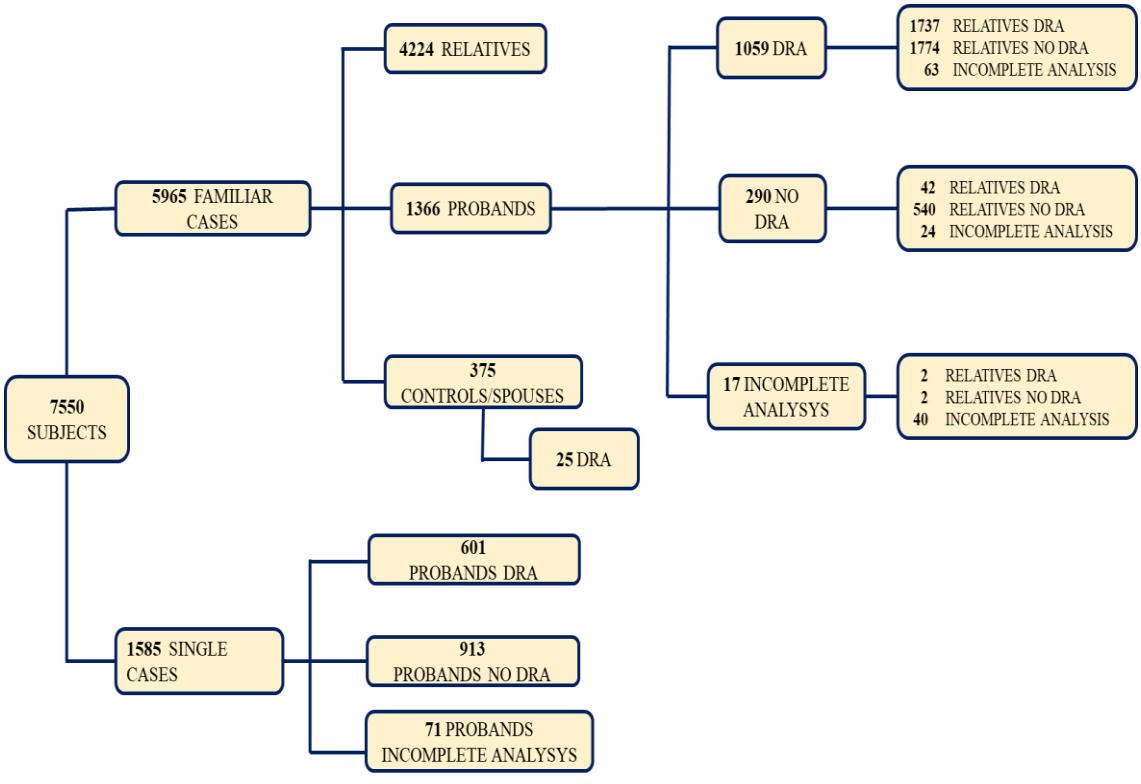


Figura 10. Molecular analysis of the 1993-2020 cohort

	PROBANDS (N. 2951)	RELATIVES (N. 4224)
DRA (N.)	1660 (56,2%)	1781 (42,1%)
Age at evaluation	44,6	48,4
Men (N.)	984 (59,2%)	807 (45,3%)
DRA SIZE (RU)		
1-3	219 (13,1%)	96 (5,3%)
4-5	393 (23,6%)	474 (26,6%)
6	333 (20%)	375 (21%)
7-8	446 (26,8%)	580 (32,5%)
9-10	269 (16,2%)	256 (14,3%)

Table 1. Demographics and molecular data of probands and relatives

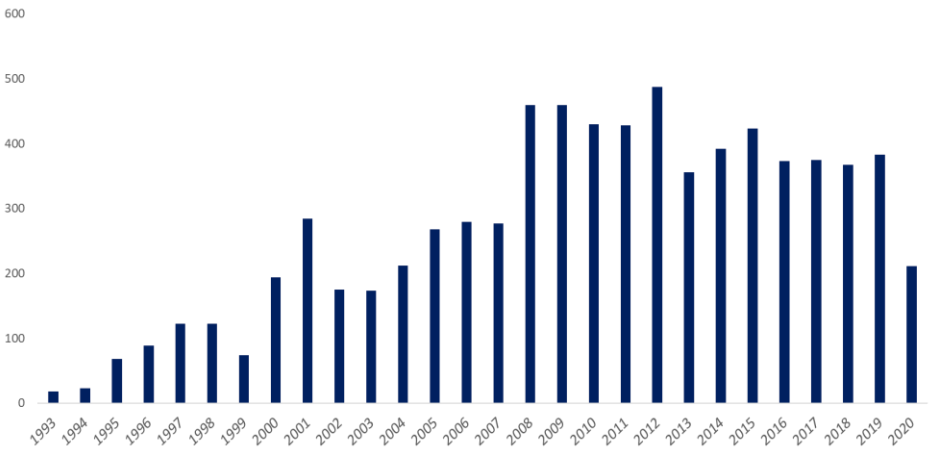


Figura 11. Registered persons for each year of the cohort 1993-2020.

Figure 11 shows the number of subjects analyzed yearly since 1993. Starting from 2007, with the establishment Italian National Registry for FSHD (INRF), the number has increased by about 200 subjects and then stabilized in following years. In 2020, due to SARS Cov-2 pandemic, the number of examined families has drastically decreased, mainly resctricted only to probands. The whole cohort has been divided in 4 sets that covered respectively the timeframe 1993-2006 (1), 2007-2013 (2), 2014-2016 (3), 2017-2020 (4) cohorts. In the first cohort, 1993-2006, an average of 150 people per year were registered; since 2007 the average has risen to 400 people per year. As part of the activity of the ICNF analysis was extended to families. Figure 12 shows an average of 100 relatives registered per year, belonging to families with probands registered in previous cohorts. The number of relatives investigated in each cohort is shown in figure 13A. Figure 13B shows that two-generation families are the most represented in all cohorts. In some cases, data were collected from a large number of relatives, up to 33 people in one family, including 6 generation families. These results attest how families continue to be followed and studied over the years. Demographics and molecular data of 4 different sets are given in Table 2.

AVERAGE SUBJECTS 1993-2006	AVERAGE SUBJECTS 2007-2013	AVERAGE SUBJECTS 2014-2016	AVERAGE SUBJECTS 2017-2020
150	414	396	334

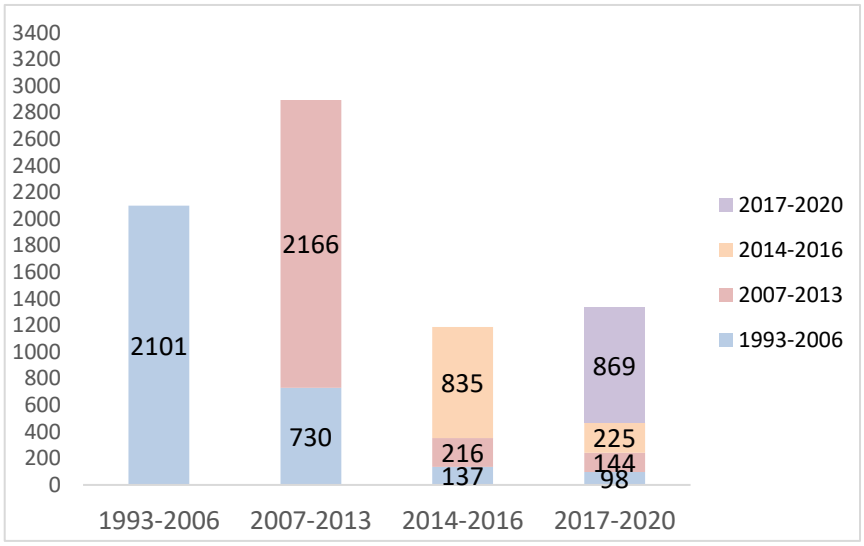


Figure. 12 Trend of the study of families in the different cohorts; the graph shows how many people of a given cohort are relatives of families registered in the previous cohorts.

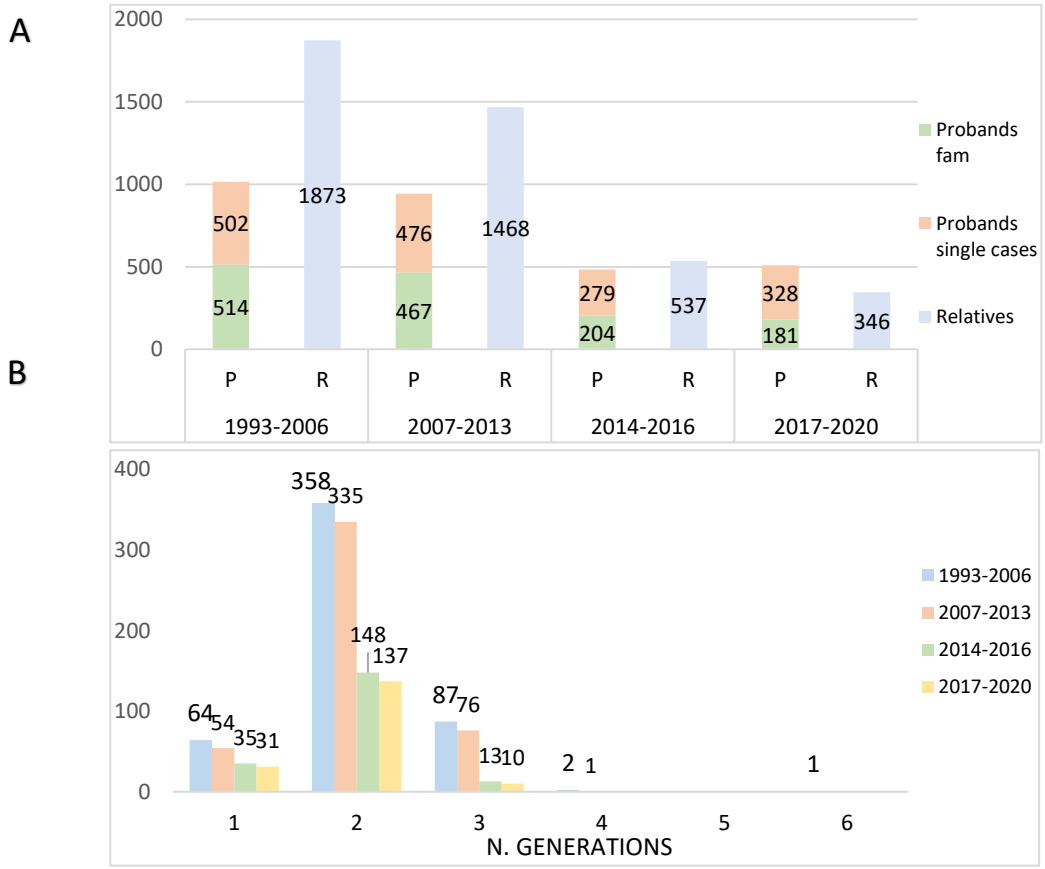


Figure 13.A Graphic rappresentation to probands and relative; B. number of families divided in generations in the 4 different cohorts.

COHORT DRA				
SET	PROBANDS SINGLE CASES	PPROBANDS WITH RELATIVES	RELATIVES DRA	RELATIVES NO DRA
1993-2006	162	415	810	777
2007-2013	153	362	598	598
2014-2016	118	147	203	212
2017-2020	168	135	128	140

COHORT NO DRA				
SET	PROBANDS SINGLE CASES	PROBANDS WITH RELATIVES	RELATIVES NO DRA	RELATIVES DRA
1993-2006	310	96	199	15
2007-2013	301	101	225	18
2014-2016	148	50	96	5
2017-2020	154	43	69	4

Table 2. Number of specific molecular diagnosis of 4 different sets or cohorts. Probands are divided in familiar case and single case. Relatives DRA and NO DRA listed refers respectively to proband DRA (up) and NO DRA (down). Relatives DRA of probands NO DRA have been identified in the "mixed families" and will be better described in the following chapters

The discovery of FSHD2 subjects, named NO DRA, between 2007-2013, led to an increase of analyses in patients with no family history and no known segregation. The recruitment has increased from 18 to 24% in probands of families up to 2020, while from 10% to 19% in relatives. About NO DRA single cases, the number is constant in the 4 different cohort and it is three times greater than families. Thanks to combination of PFGE and gel 0.4%, we can describe the distribution of alleles within the collected population (fig 14A). The majority of probands and relatives carry alleles between 30 and 38 kb (7-9 U). Notably, 70% of the shortest alleles (11-19 kb, 1-3 U) result from of a de novo mutation. The percentage of de novo subjects in the whole DRA cohort is 8% (fig 15). Pulsed-Field Gel Electrophoresis (PFGE) revealed a high percentage of translocation of D4Z4 repeats between chromosome 4q and 10q. Considering 2863 probands and 375 healthy controls with molecular analysis, we identified 364 subjects carrying translocation of 4q chromosome on 10 and 10q chromosome on 4. This data establishes the percentage of translocation in our cohort is 11,2 %, differently as reported 20-30%. The most frequent recombination profiles are 4/3 with 188 subjects (51% tot translocated) and 155 4/1 (42%), followed by 16 subjects with 4/4 profile (4,4%) and 4 subjects with 4/0 profile (1%) and 1 subject with double translocation. If we calculated the percentage of 4 singles alleles distinctly to translocate, we have 2,81%.

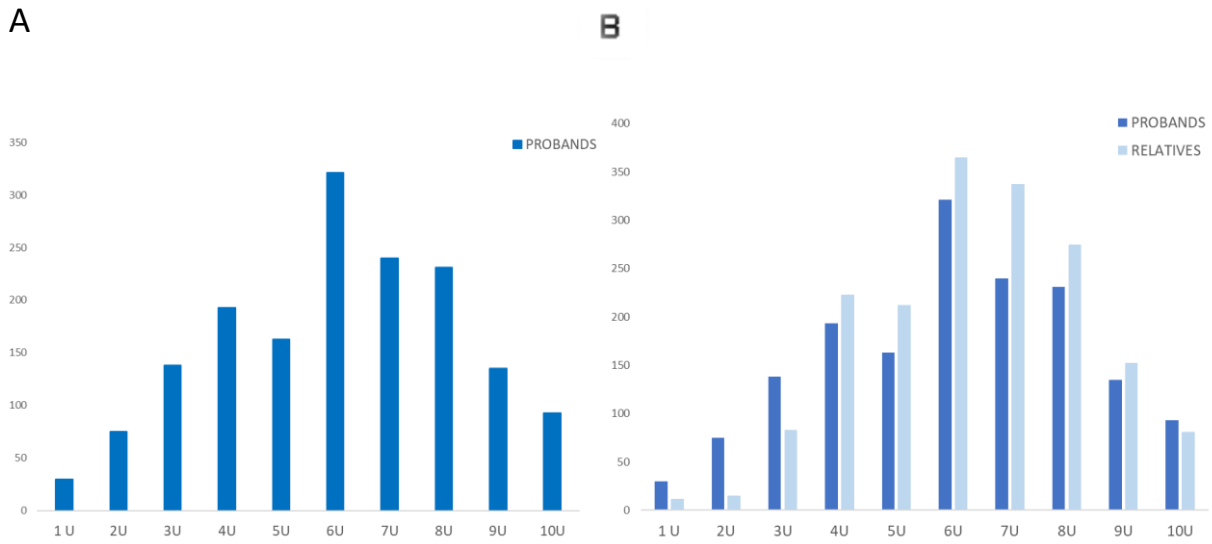


Figure 14. (A) Allele distribution in probands of whole cohort 1993-2020 (dimensions are defined in kb). (B) Allele distibution in proband and their relatives of whole cohort 1993-2020

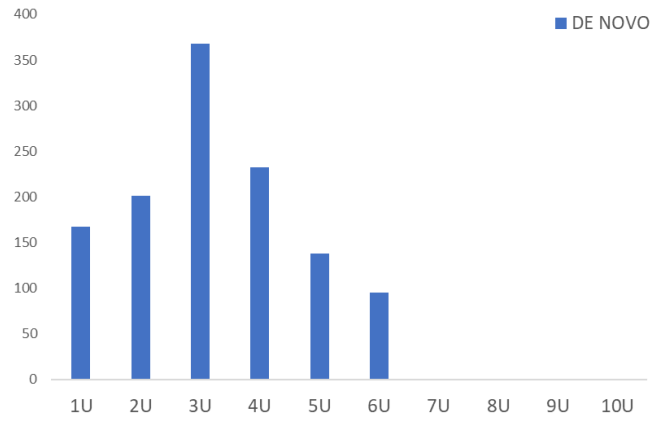


Figure 15. Allele distribution in probands de novo

3.2 Standard clinical evaluation

In 2007, the ICNF introduced a standardized methodology to evaluate the degree of muscle impairment in carrier of DRA. The clinical protocol was designed to define the involvement of muscle groups specific for FSHD independently [Lamperti et al, 2010]. A score for each muscle group was attributed to define the degree of muscle impairment, if any. The sum of the assigned scores gives the FSHD score that translate the degree of disability into a number. The evaluation of 530 DRA carriers from 173 families demonstrated the reduced penetrance and the clinical variability present in these families. Remarkably, the study revealed that disease penetrance depends on the degree of kinship with the probands, being affected 75% of first-degree relatives versus 50% of second-to-fifth degree

relatives. Moreover, the FSHD score was significantly lower in relatives in comparison with probands. These results indicate that multiple elements contribute to disease expression, a feature of complex genetic diseases. On the basis of these data the ICNF designed the Comprehensive Clinical Evaluation Form (CCEF), which, besides defining disability with the FSHD score, permits a precise description of clinical phenotypes. Category A defines typical FSHD phenotypes (presenting facial and scapular girdle weakness), through the three sub-categories A1, A2 and A3 based on severity of facial weakness. Category B includes the incomplete phenotypes with only scapular girdle weakness (impossibility in upper limbs abduction and winged scapula) or only facial weakness, respectively in the sub-categories B1 and B2. Category C differentiate non penetrant carriers, for subjects presenting signs suggestive for FSHD and the healthy subjects with no signs of muscle weakness, respectively in the sub-categories C1 and C2. The last category, D, comprises FSHD patients presenting additional myopathic signs from or non-muscular features in the sub-category D1 or non-FSHD patients (not fulfilling the diagnostic criteria for FSHD) in category D2. In the period 2007-2013 the clinical evaluation assessed the FSHD score. From 2014 the use of the CCEF was introduced and both protocols were used in the period 2014-2016. In the period 2017-2020 the use of the CCEF was fully implemented [Ricci et al, 2016]. Currently, the INRF accrued 2122 subjects clinically assessed using the CCEF and 1783 subjects who were appraised using FSHD SCORE. Figure 16 shows the distribution of the clinical assessment. The FSHD score remains the most common assessment in the 2007-2013 cohort when it was first used as a clinical standard.

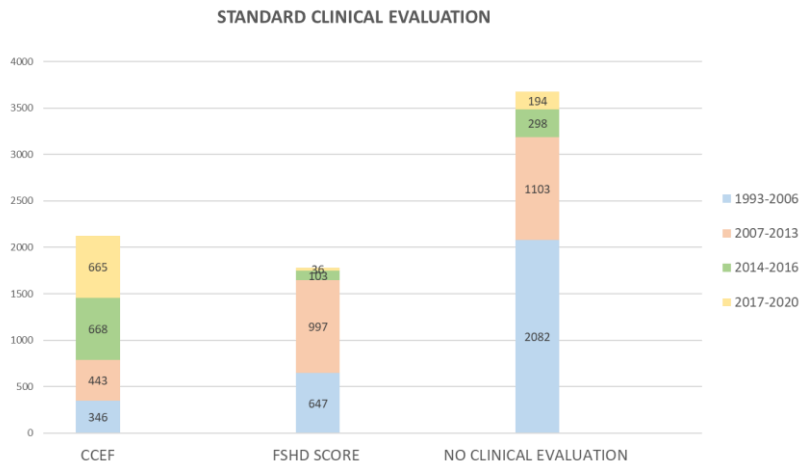


Figure 16. Standard clinical evaluation distribution in 4 different cohorts

3.2.1 FSHD SCORE

In figure 17A and 17B FSHD score is rappresented, from 0 to 15, in the probands and relatives divided in to DRA and NO DRA cohorts. We can see that the distribution of the score in DRA probands

follows a Gaussian, unlike relatives. Respect to the total number of each cohort, we observed similar percentages in the range score 3 - 6 between DRA and NO DRA probands. As expected, we have a higher percentage of DRA probands between scores 7 -10 and major NO DRA probands between low score (1-2). The average score for probands DRA is 6,4, while in relatives is 2,5. For NO DRA subjects, proband have a score of 3,8, relatives 1,4.

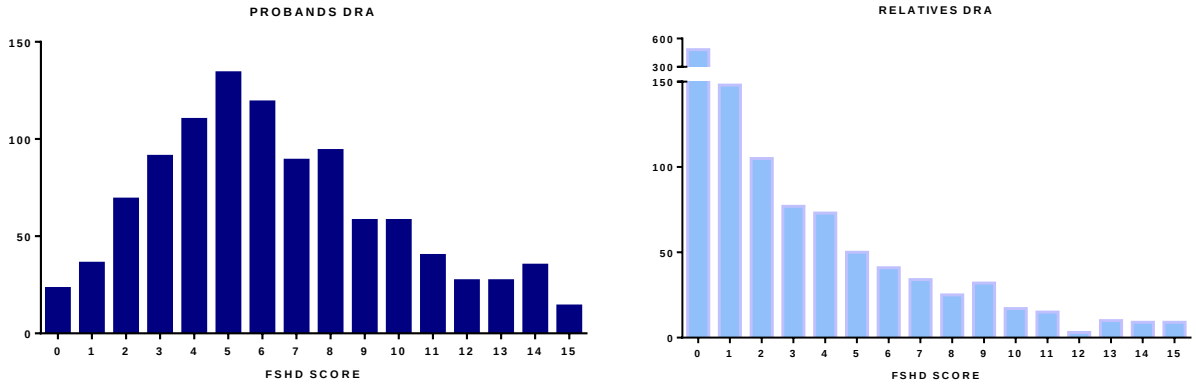


Figure 17.A. Score distribution between probands and relatives DRA

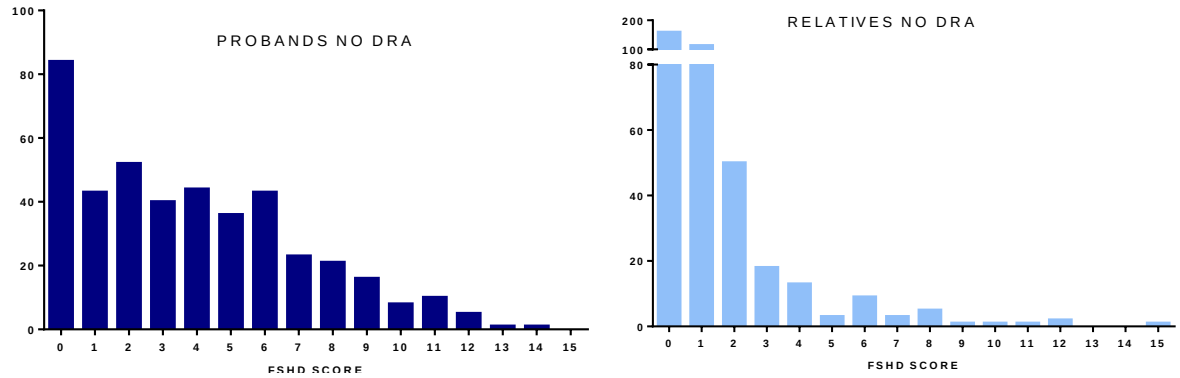


Figure 17.B. Score distribution between probands and relatives NO DRA

Figure 18 shows the different distribution of the FSHD score in probands and relatives grouped on the basis of the DRA size. The large number of relatives with no muscle impairment is in 4-10 DRA carriers, whereas no significant differences between probands and relatives were observed in carriers of 1-3 DRA. The average score for each range of size is calculated in all DRA probands and relatives. The difference between the two cohorts ranges from 3 to 4 score value. The same data has been obtained considering only the affected of both cohorts, and score value is lower of two points (see below).

DRA SIZE (RU)	PROBANDS (N. 1024)	RELATIVES (N. 1131)
TOTAL SCORE	6.4	2.5
1-3	9.1 (139)	6.4 (52)
4-5	6.9 (217)	3.8 (273)
6	5,9 (213)	2.8 (254)
7-8	5.8 (265)	1.8 (389)
9-10	5.3 (190)	0.8 (163)

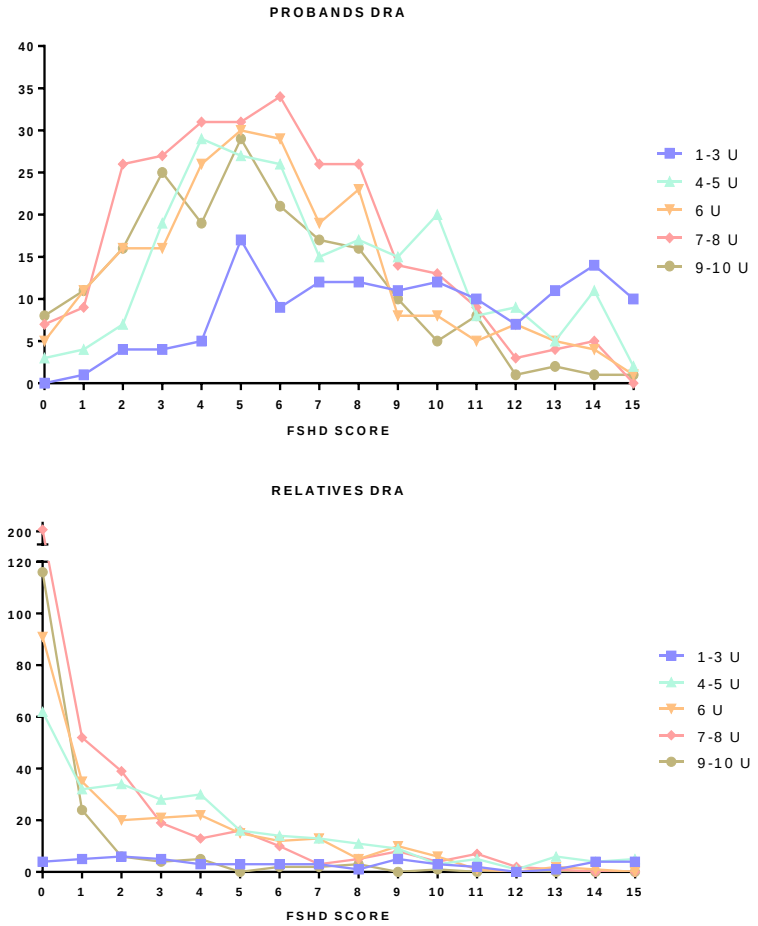


Figure 18. Distribution of FSHD SCORE compared to different size DRA in probands and relatives.

DRA SIZE (RU)	PROBANDS (N. 1001)	RELATIVES (N. 648)
AFFECTED SCORE	6.6	4,4
1-3	9.1 (139)	6.9 (48)
4-5	7.0 (214)	4.9 (211)
6	6.1 (208)	4.3 (163)
7-8	5.9 (258)	3.8 (179)
9-10	5.5 (182)	2.7 (47)

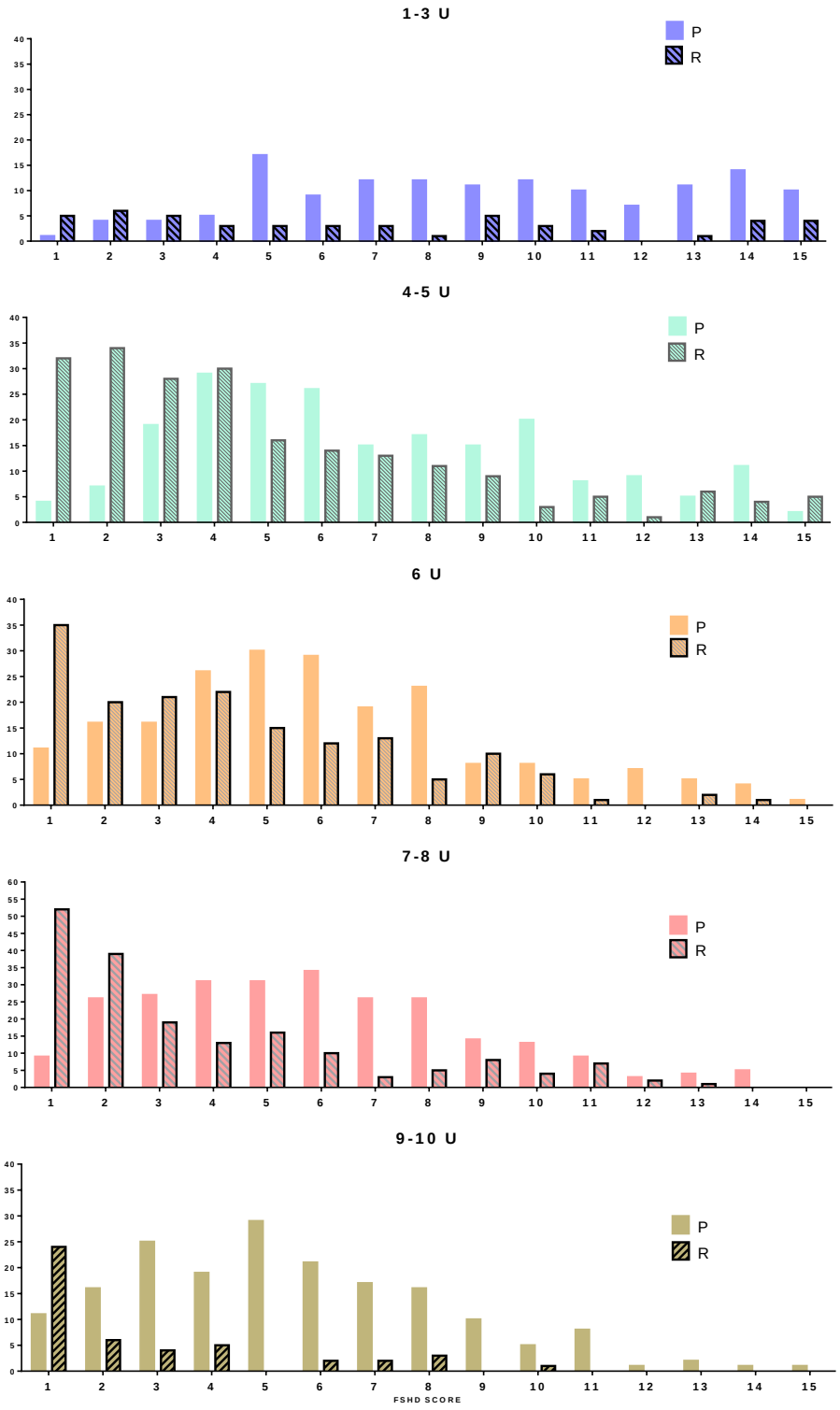


Figure 18.A. Distribution of FSHD SCORE compared to different size DRA in probands and relatives affected.

Figure 19A shows the distribution of FSHD score in probands carriers of different size DRA on the base of gender. Considering the numbers reported in the demographic table, we could think a greater distribution of the disease in males. The average score calculated in all DRA probands is 6.1 for male and 6.8 for femal; in all relatives is 2.6 for male and 2.5 for female. A t test has been performed and

there is no significant difference between male and female value score and specific size DRA ratios, but for the 6 U and 9-10 U there is a greater difference (p.0.0654 and p 0.0622). For this, the average score for each range of size is calculated on affected subjects of both cohorts: 6.3 for male and 6.9 for femal, in relatives is 4.3 for male and 4.5 for female. Females probands have a score greater than 1 compared to males in the size range 6 U and 9-10 U, as anticipated in the statistical test, even if it is not significant. For the other ranges there are no differences.

DRA cohorts	PROBANDS (N. 1024)	RELATIVES (N. 1131)
Men (N.)	603	518
Women (N.)	421	613

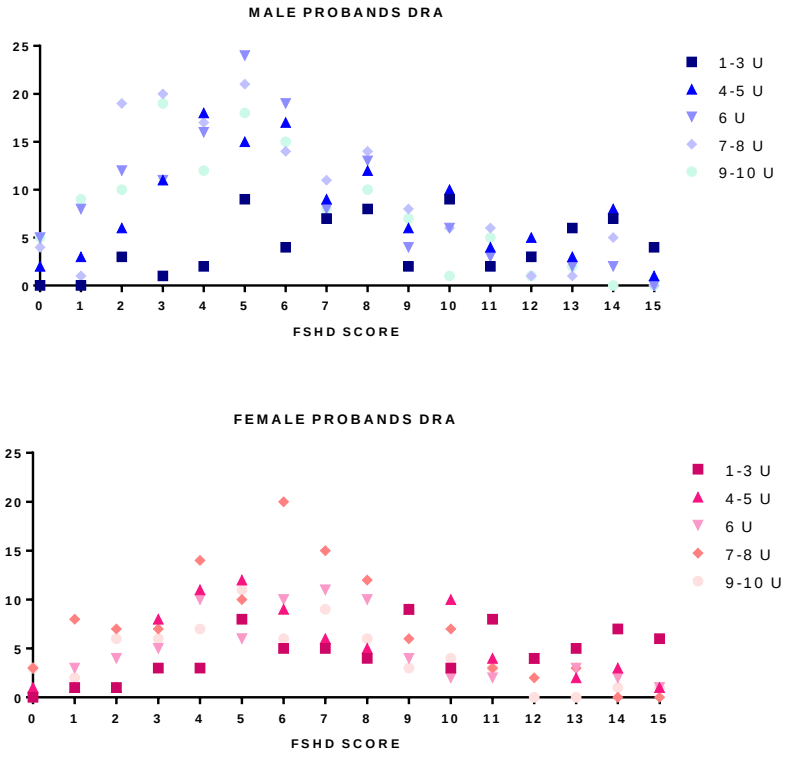


Figure 19A. Distribution of FSHD score in probands male (up) and female respect different size of DRA

In females affected relatives we found a score greater than 2 points in the 1-3 U range respect male, considering we observed 20 and 28 subjects respectively. In the same cohort we found 1 point score higher in the range 7-8 and 9-10 U compared to the affected males (Fig 19B). For NO DRA cohorts there is no differences between gender, neither between affected cohorts. All relatives NO DRA have been considered here. (Fig 19C).

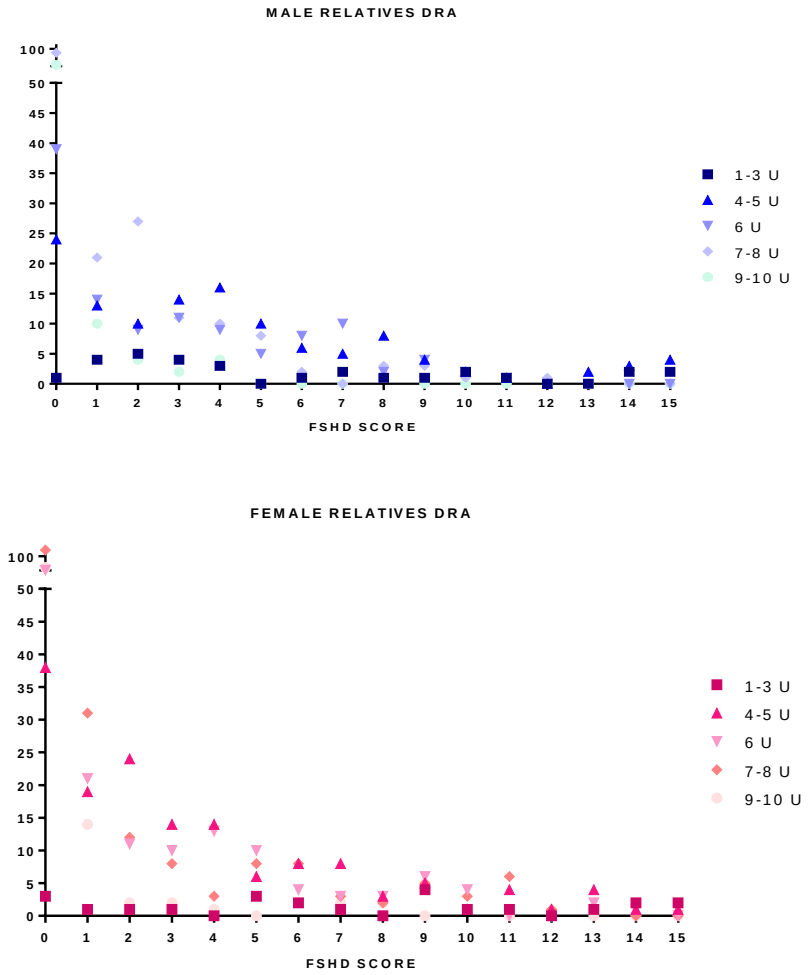


Figure 19 B. Distribuzion of FSHD score and CCEF score in relatives male (up) and female respect different size of DRA

NO DRA cohorts	PROBANDS (N. 428)	RELATIVES (N. 997)
Men (N.)	240	487
Women (N.)	188	510

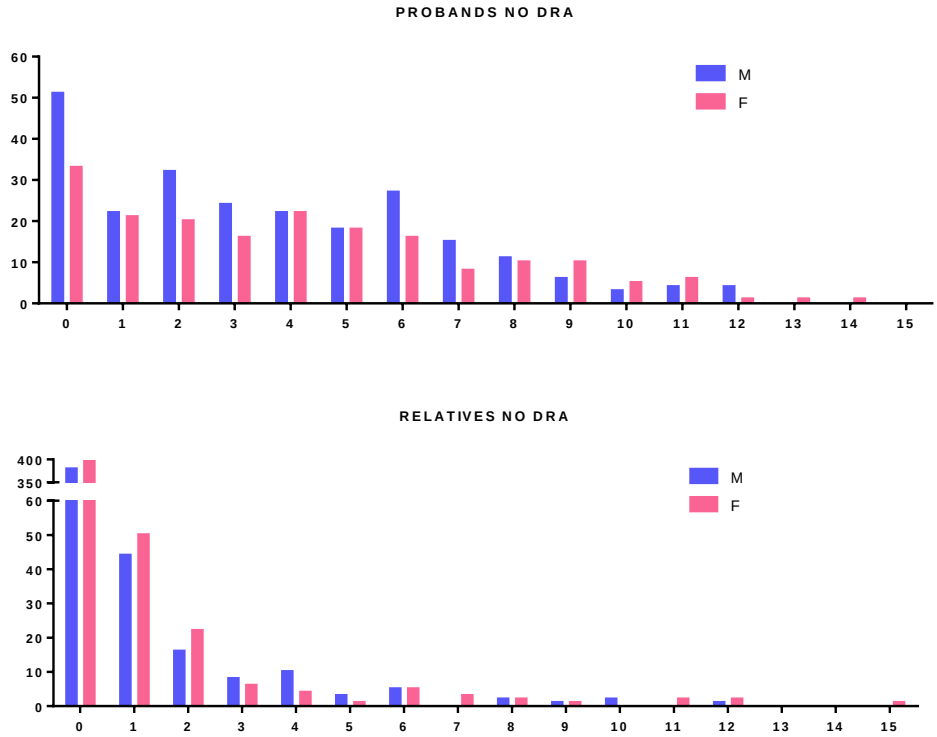


Figure 19C. Distribuzion of FSHD score in probands and relatives male and female NO DRA

Male and female FSHD score was related to age at evaluation and age at onset. Figure 20 shows the cohorts divided in DRA and NO DRA subjects. In general, in both cohorts the males get sick earlier and, in some cases, we observed a difference in the duration of the disease. In male DRAs in the range of score 8-15, the duration of the disease is longer than 5 years (mean) compared to females. In fact, as the score increases, the years of difference increase. In NO DRA subjects this is not observed, but we not have male with score >12. In addition, in male NO DRAs with scores between 1-2, there is a difference of 26 years in the duration of the disease. The average value is shown below.

	MALE	FEMALE
AGE AT EVALUATION DRA	44,3	47.7
ONSET AGE AT ONSET DRA	23,3	28,8
AGE AT EVALUATION NO DRA	50.3	51,9
ONSET AGE AT ONSET NO DRA	29,2	39,4

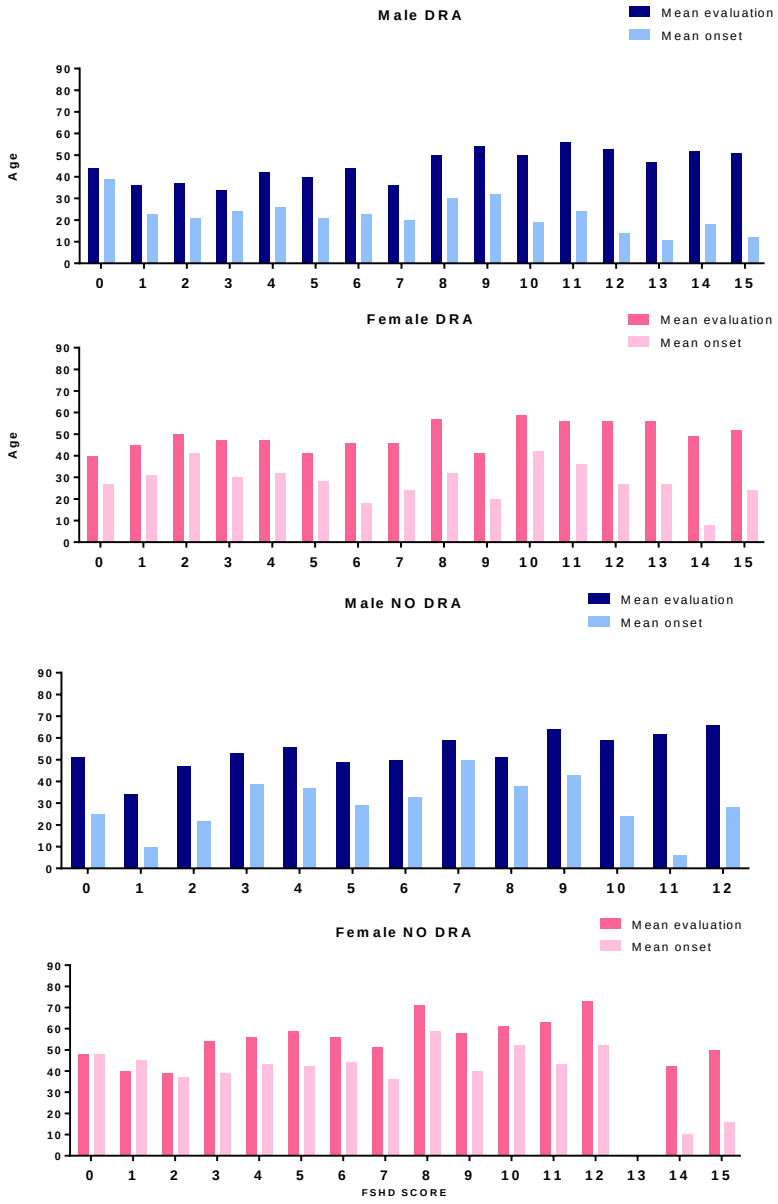


Figure 20. Age at evaluation and age onset in male and female in probands and relatives

3.2.1 PHENOTYPIC CATEGORIES

Figure 21A shows the distribution of CCEF clinical categories in probands carrying a DRA on the basis of the FSHD score. As expected, subjects with the classic FSHD phenotype (category A), are the most common among DRA probands and among the highest scores (54,08% vs 16,3% NO DRA). Notably subjects with uncommon features (category D1) are present in all the groups with FSHD score >1. However, the D1 percentage between probands cohorts is the same (29,5%-31,4%), the high percentage of D1 within NO DRA confirms that this phenotype is associated with a complex multifactorial process. These observations strengthen the link between the short allele with the classic phenotype despite the variability. Another observation is the percentage of B1 and B2 in DRA

probands (9% and 1,6% respectively) is the same observed in NO DRA (8,9% and 2,85% respectively) respect of total for each cohort. Using the CCEF, changes also molecular analysis evaluation. Indeed, the half of NO DRA probands not have any kind of phenotypic involvement (category C), or have another phenotype or disease. The percentage of D2 category is greater in NO DRA subjects (48 vs 16).

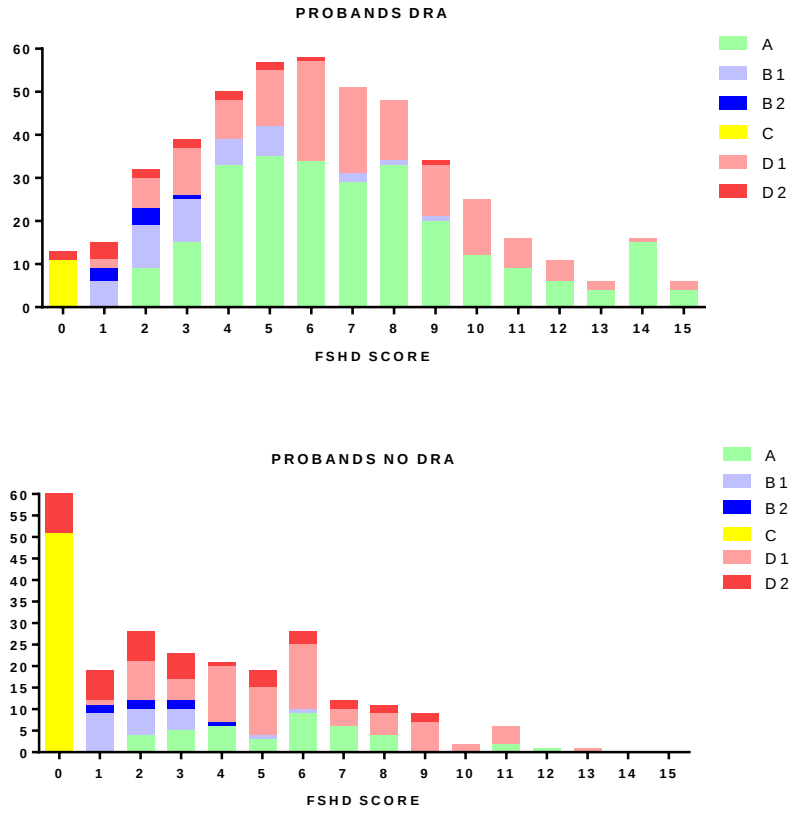


Figure 21. Categories distribution in DRA and NO DRA probands (up). Category A include A1, A2, A3; category C include C1 and C2

	PROBANDS DRA SCORE	PROBANDS NO DRA SCORE
A	7.1	5.6
B1	3.4	2.1
B2	1.8	2.3
D1	7	5.7
D2	2.9	2.7

Figure 22 shows the distribution of clinical categories in relatives considering the FSHD score and presence of DRA. The majority of relatives carrying DRA have no muscle impairment (43% with FSHD score 0) or very mild (22% with FSHD score 1-2). This is also evident in relatives carrying

D4Z4 alleles of normal size, even though it is expected. In this group 12 relative have been assessed as category A (2.1%) respect DRA relatives (26,5%). Moreover, complex phenotype D1 is more represented in DRA relatives (8% vs 2%), compared to what was observed in the probands. This suggest that the mechanisms lead to the involvement of atypical muscle groups could be different in two cohorts and should be investigated separately. B1 category is presented in 9,6% of DRA relatives and, like in probands, cover a highest range score differently to 5,2% of B1 NO DRA relatives, but we observed a difference in percentage in relatives. B2 percentages not change. Notably, the distribution of subjects presenting a phenotype different from FSHD (category D2) is similar between NO DRA and DRA relatives. D2 relatives NO DRA are 42 and 25 subjects carry DRA and this means affected by other muscular diseases could be carriers of DRA alleles but D4Z4 allele dimension are not involved.

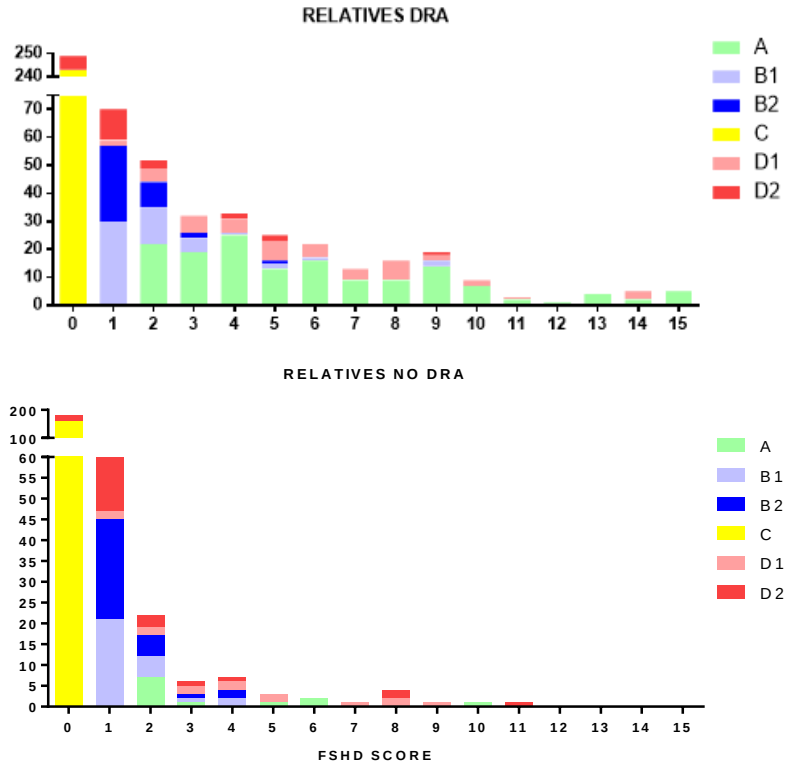


Figure 22. Distribution of categories in Relatives DRA (up) and NO DRA. Category A include A1, A2, A3; category C include C1 and C2.

	RELATIVES DRA SCORE	RELATIVES NO DRA SCORE
A	6	3.7
B1	2	1.4
B2	1.4	1.4
D1	5.9	4.4
D2	1.8	1.3

I used the data collected with CCEF on the onset and duration of the disease to understand the commonalities between the different categories and molecularly different cohorts. As described in table of figure 23, the average age at evaluation of DRA probands is 44,6, while of DRA relatives is 48,4. Not-carriers probands have an average age of 8 years older (53.1) and onset data show same trend. For NO DRA relatives the average age at evaluation is lower (48,3) respect DRA relatives, because mayority belong to DRA probands, which arrive before in visiting. I compared the average both age at evaluation of registered subjects and the onset age for each category comparing probands and relatives. For categories A1, A2, D1, probands shows dystrophyc signs earlier, indipendently to age at evaluation. For categories A3, B1 and B2, this ratio changes. Relatives onset is lower than of probands, but age at evaluation is the same. This can be explained because mayority of these relatives have probands with severe phenotyope and is early evaluated. Indeed, the difference with age at onset shows the phenotype stops. For these cathegory we could consider the proband data as actual one. NO DRA probands clearly fall ill almost 10 years later than DRAs. For A2 category, which is the most representative in the classical phenotype, there is a high difference between probands and relatives. For categories A3, B1 the trend is the same to DRA cohort, and curiously B2 probands show an onset 20 years later. Probably, the same situation descripted for DRA cohorts is verified, i.e., there may be that there is greater attention even to the small signs of disease already from the first visit.

	PROBANDS	RELATIVES
Age at evaluation DRA	44,6	48,4
Onset age at onset DRA	24,55	27,56
Age at evaluation NO DRA	53,1	48,3
Onset age at onset NO DRA	34,76	30,43

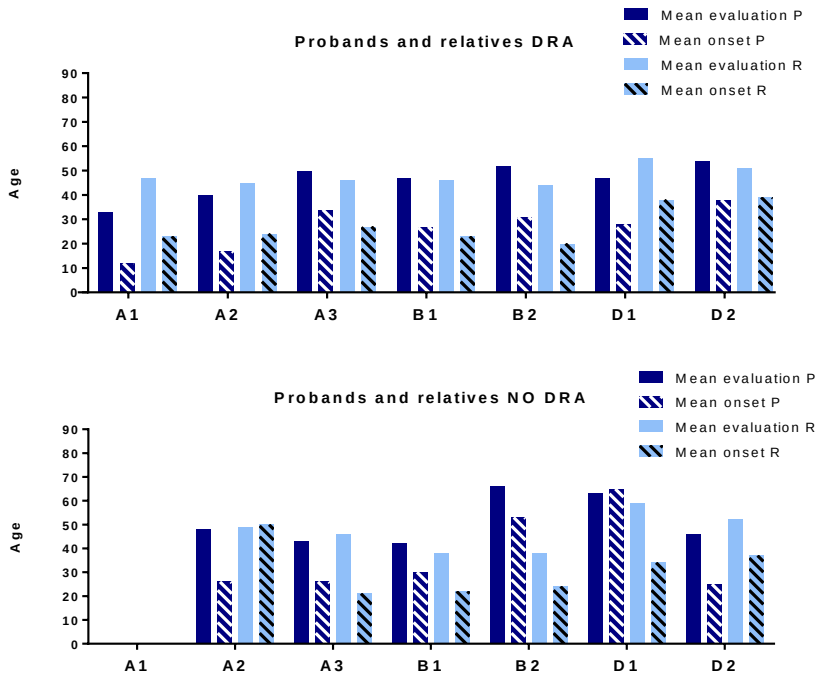


Figure 23. (up) Summary table of the average age at evaluation. (down) Graphic representation of the relationship between categories, average age and age of onset. On the Y axis we find the age ranges.

I crossed the age at evaluation and onset data with categories by comparing males and females. Generally, males get sick about 4 years earlier than females in DRA cohorts. For A1 and D1 categories the duration of disease in male is higher of 4 and 8 years respectively. This is in agreement with the previous observations shown in the report with the score evaluation. NO DRA males present dystrophy signs 12 years earlier. Same results of DRA cohorts are observed in male. For A3 category the duration of disease is higher of 8 years and 10 years for D1 category. Together to score evaluation results, these data could suggest gender factors are involvement in progression of disesase.

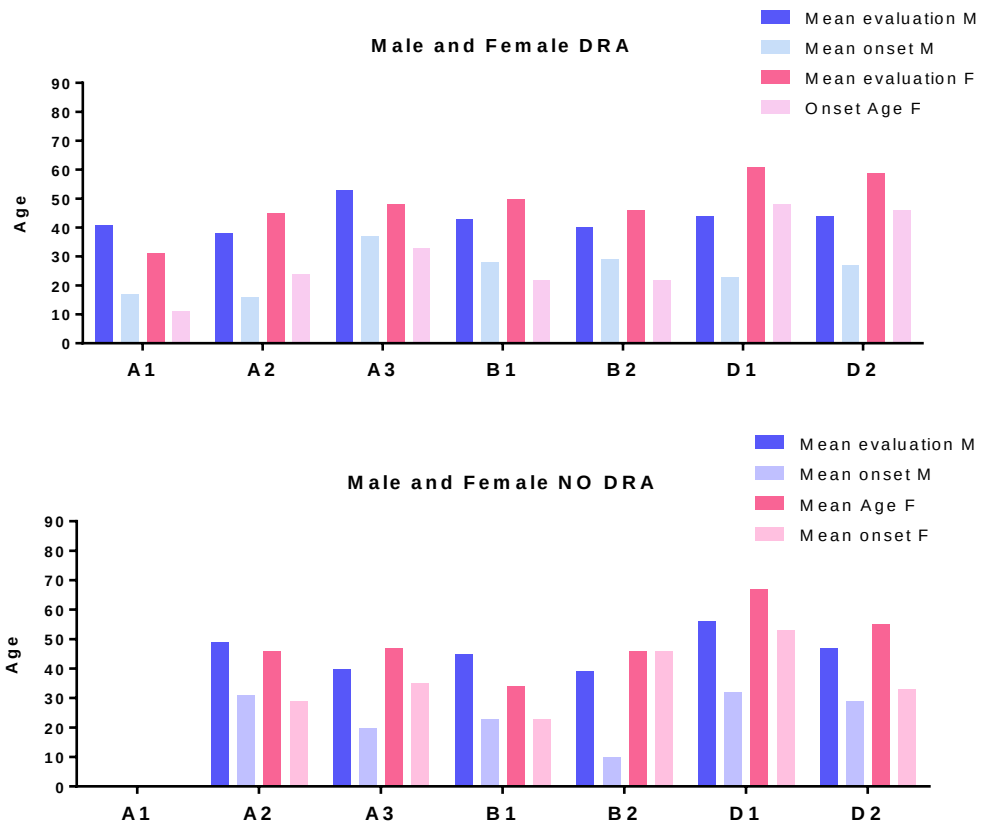


Figure 23. (up) Summary table of the average age and onset age in male and female. (down) Graphic representation of the relationship between categories, average age and age of onset in male and female. On the Y axis we find the age ranges.

3.3 PENETRANCE AND FAMILY ANALYSIS

Thanks to the Italian National Registy for FSHD and the standard clinical evaluation methods, we have the possibility to calculate the overall penetrance of DRA in relatives carrying a DRA. Out of 1166 clinically examined relatives carrying a DRA, 671 had FSHD score ≥ 1. Thus, the estimated penetrance is 57,5%. We also calculated the penetrance in individual families, those with 4 and more carriers. We identified 138 families, with a total number of 733 DRA carriers (138 probands and 595 relatives. To describe our population, the families are divided for carrier number and defined the range of penetrance. Out of 138 families, 23 (8.7%) have a complete penetrance and major families present less of 50% of penetrance. Other results are summarized in table 3. As previous descripted

[Ricci et al, 2016] penetrance decreases as the number of carriers analyzed per family increases. Moreover, as the number of families analyzed increases, we observe a great variability in penetrance.

DRA CARRIERS (N.)	61	24	24	11	6	5	2	1	2	2
FAMILIES	n/4	n/5	n/6	n/7	n/8	n/9	n/10	n/11	n/12	n/14
PENETRANCE (%)	20-100	20-100	10-100	40-100	10-80	30-90	20-60	90	50-70	40-50

Table 3. Rappresentation of families with 4 and more DRA carriers. The total number is 138, which are included 595 relatives. Families are divided by number of carriers; for each the total number and range of penetrance are marked.

We further investigated the clinical setting of families subdivided on the basis of the proband category. The study on the penetrance of families led me to investigate the probability for relatives to have a specific category respect that one of the proband, or to have a certain FSHD score. Therefore, the following graphs bring together probands categories with those one of their respective relatives. Figure 25 show the DRA cohort. Probands with classic phenotype A1, A2, A3 have 43% 25%, 27% probability respectively of having a DRA relative of C2 category. The following highest probability is to have A2 relatives and is 21% for A1 and A2 probands, 16% for A3. Moreover, A2 subjects have 14% of probability to having NO DRA relatives of D2 category (fig.26). B1 probands have 22% of probability of relatives with same phenotype, in addition to the 37% of having C2 family member. B2 probands do not have sufficient relatives with CCEF to be able to give a correct estimate, however have only C and D1 relatives. D1 probands show heterogeneous percentage of relatives having mild phenotype; moreover 13% are D1, followed by D2 subjects. D2 subjects have a little percentage to have relatives with A category and same observation have been found in NO DRA probands. In general, for relatives NO DRA percentages are very homogeneous, excluding C2. A2 and A3 probands have the probability of 11 and 13% respectively of having a relatives B1, B2. (fig. 25). Curiously, D1 subjects have a higher percentage of relatives A3 rather than D1 or D2, unlike that observed in DRA probands.

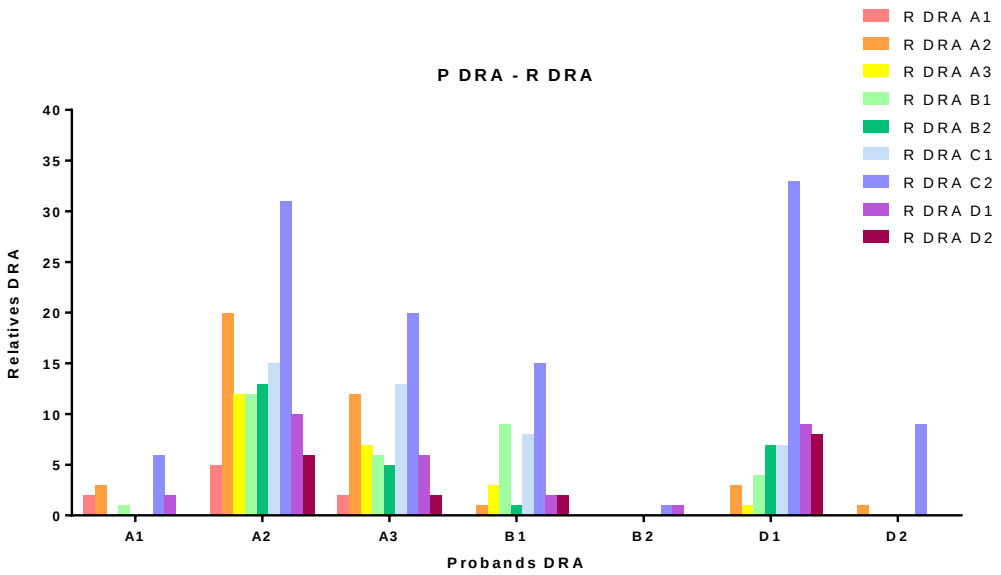


Figure 24. Relatioship between probands and respective relatives categories in DRA cohort

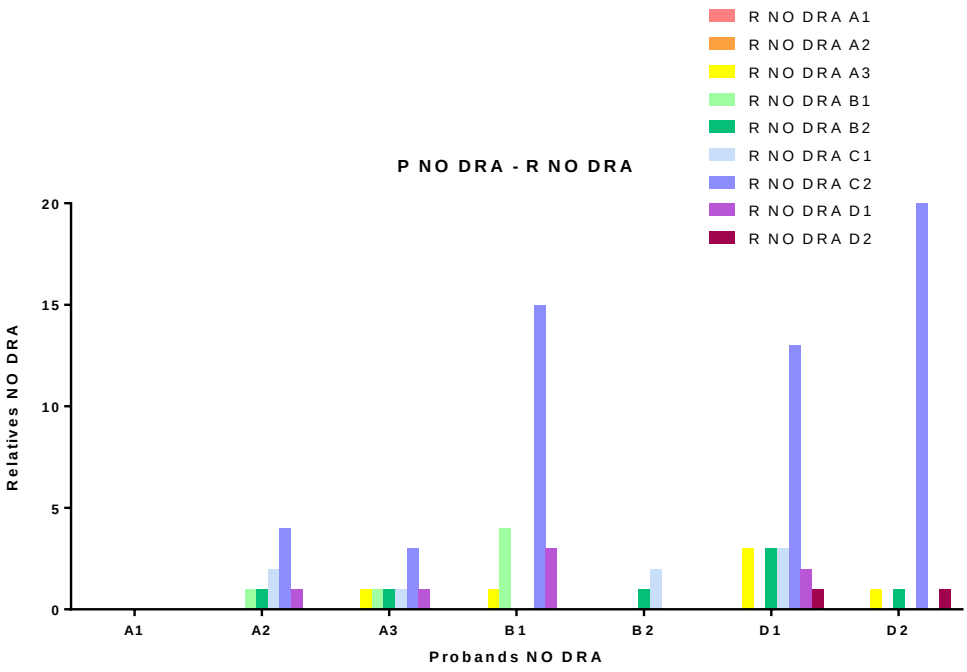


Figure 25. Relatioship between probands and respective relatives categories in NO DRA cohort

During analysis of the families, category and molecular variability between members was observed. I observed 47 DRA probands who have 89 relatives NO DRA with category A, B, D or score ≥ 1. I defined these "mixed families" and I selected 42 NO DRA relatives with clinical evaluation born from 1970 onwards, so they are aged between 50 and 1 year. This choice was made for two reasons: first, it has been reported that the average age at which complete penetrance is reached is 50 [Tawil

et al, 1993]; the older subjects may have clinical signs due to aging. Regarding the phenotype, 22 of the NO DRA subjects have CCEF evaluation: 2 are category A, 6 B1, 8 B2, 3 D1, 3 D2; 17 with a score between 1-2; 3 between score 3-6. At this point molecular data has been crossed to understand if there were other mixed families. We found 26 families with NO DRA proband, with 42 DRA relatives of all ages. Of these, only 12 families have phenotype in both cohorts, therefore I added them to the cohort of mixed families (data not show).

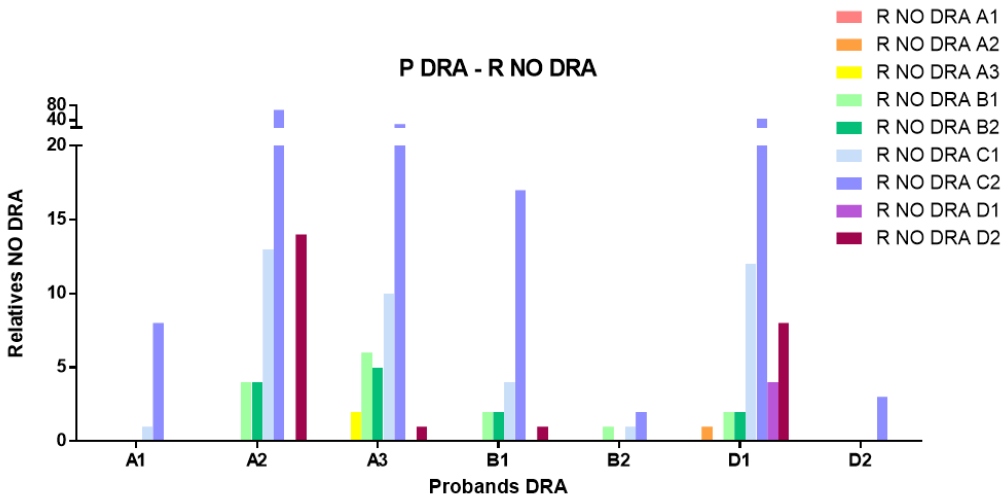


Figure 26. "Mixed families": relatioship between categories of probands DRA and respective relatives NO DRA.

The relationship between probands categories and the FSHD score in the DRA relatives confirms that more severe phenotype is associated with a high probability of having a relative with the same phenotype. For A2 probands, there is 35% probability that relatives have FSHD scores from 3 to 7. The same score range can involve relatives of A3 probands. D1 category has heterogeneous probability, but the highest score range remains between 1 and 2. Probably, this is because D1 is also representative of incomplete phenotypes, therefore for a more correct data it would necessary to distinguish the various subgroups of category with uncommon signs. The relationship between NO DRA proband and relatives with FSHD score evaluation is not very representative, but it seems that relatives have scores in the high range. It would be necessary to re-evaluate the relatives and study deeply the families.

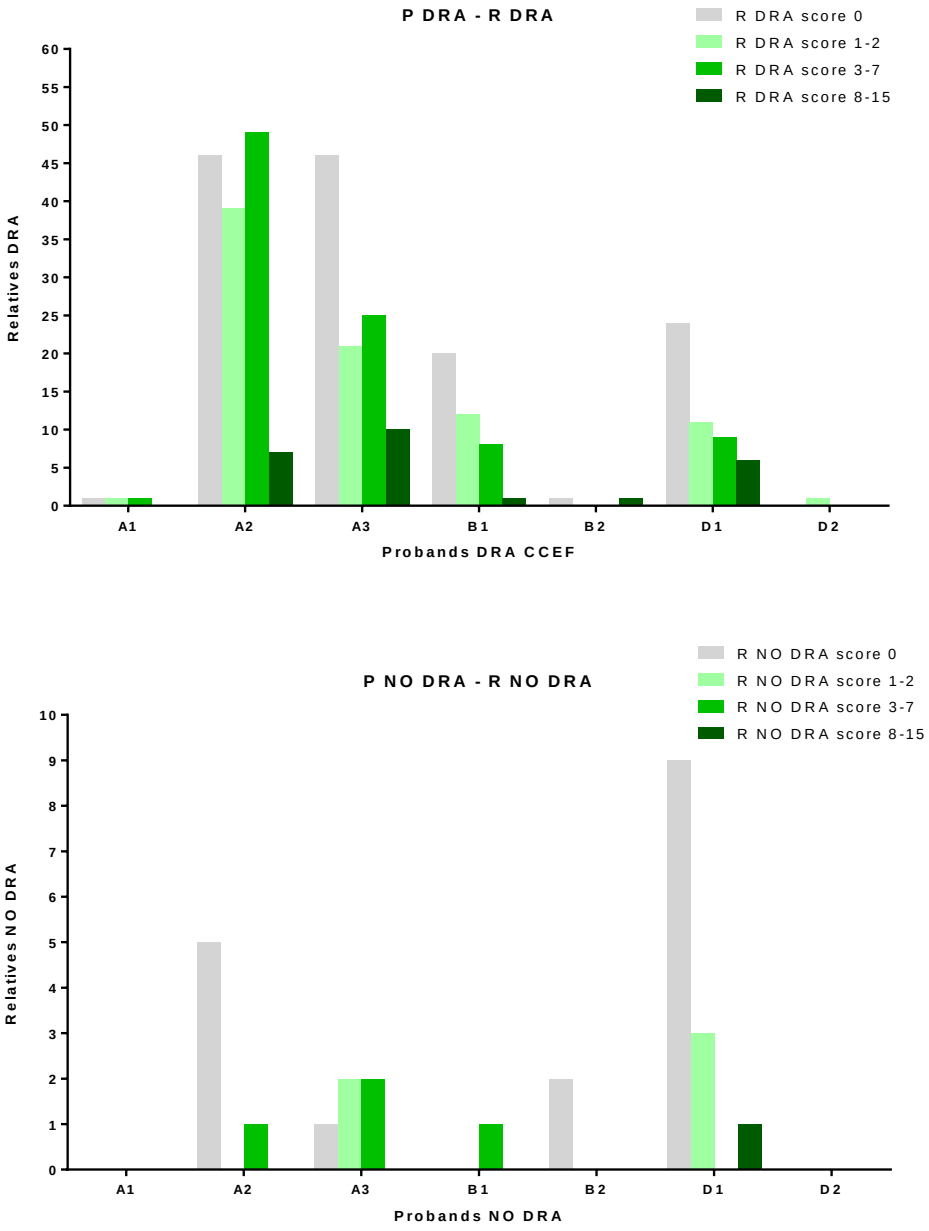
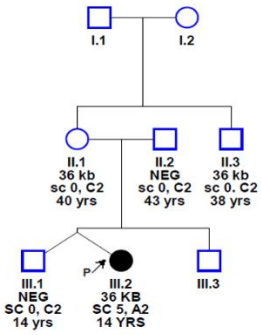


Figure 26. Relationship between probands categories and FSHD score of respective relatives in DRA and NO DRA cohort

3.4 TRIO WGS analysis

We investigated family trios composed by the affected subject and its healthy parents with a pattern of inheritance compatible to an autosomal recessive trait, in order to identify alternative genetic mechanisms by exploiting the WGS/WES technology. The analyzes have been carried out are: identification of de novo mutations in probands, profiling of mutations with autosomal recessive inheritance patterns, relatively rare in the general population (AF <= 0.01), but homozygous in the proband and heterozygous in parents. In both cases, the mutations were noted to predict possible functional effects. A very rich annotation system was used (with annotations of regulatory elements and the like), but we have focused only on the "main" annotations, i.e., effects on genes per the proteins. The coverage level (> = 20x) in all subjects of the trio allowed the following analysis of **4** families. To confirm the reliability of the data and for the type of analysis, we wanted to carry out, it was necessary / fundamental that the different polymorphic sites seen in an individual are covered by at least 10 reads (and are therefore genotyped) also in all the other individuals. In the families analyzed this happened for a corrected proportion between 93.6 and 98.7 of the observable polymorphic sites. Next step was the number and proportion of genotypes that do not agree with a Mendelian inheritance. Given the error rate of this type of experiment, it is safe to expect that c.a. 1% of all named genotypes may fall into this category. For all families, the proportion of sites not compatible with Mendelian inheritance was in the expected range, and was compatible with the error rate. For each child/probands has been reported the number of de novo mutations and rare autosomic homozygotic mutation that are heterozygotic in parents. A limited number of potentially harmful mutations have been found for all families which appear to be denovo and / or are compatible with autosomal recessive inheritance.

Fam 6:



In families 6, the proband III.2 is a carrier of 9 RU DRA (36 kb) inherited by her mother, which is completely asymptomatic. The same healthy phenotype is present in the mother' brother, II.3, which is carrier of 9 RU DRA too. The proband has a dizygotic twin brother who is not a carrier. The

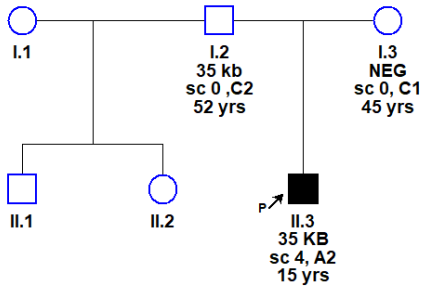
proband arrives in counseling at the age of 14; already at the age of 5 was noticed that she sleeps with eyes open, and she appears winged shoulder blade at the age of 12. She has a partial ability to protrude her lips and is unable to puff out her cheeks. He has incomplete abduction of the arms but greater than 45°, hyperlordosis and atrophy of the pectoral muscles. We found 5 candidate mutations, 4 de novo non synonymous SNV and 1 autosomic recessive frameshift deletion, all in exonic regions (see table).

FAMILY 6			
VARIANTS	GENE REF.	EXONIC FUNC. REF	AA CHANGE
DE NOVO	RAPGEF2	nonsynonymous SNV	R958H, R1003H, R1119H
	HSPA14	nonsynonymous SNV	C394G
	APOBR	nonsynonymous SNV	T347I
	CDPF1	nonsynonymous SNV	P32H
AUTOSOMAL RECESSIVE	IVD	frameshift deletion	F106Sfs*66

For the novo mutations, the genes involved with interesting functions in this contex are: 1) RAPGEF2, is a GTPase activator and functions as a guanine nucleotide exchange factor (GEF), which activates Rap and Ras family by exchanging GDP bound for GTP free in a cAMP-dependent manner. Serves as a link between cell surface receptors and Rap/Ras GTPases in intracellular signaling cascades. Its binding to ligand-activated beta-1 adrenergic receptor ADRB1 leads to the Ras activation through the G(s)-alpha signaling pathway. Involved in the cAMP-induced Ras and Erk1/2 signaling pathway that leads to sustained inhibition of long term melanogenesis by reducing dendrite extension and melanin synthesis. Provides also inhibitory signals for cell proliferation of melanoma cells and promotes their apoptosis in a cAMP-independent nanner. Plays a role in the regulation of embryonic blood vessel formation and in the establishment of basal junction integrity and endothelial barrier function. In two families, after nanopore sequencing of genomic DNA, are found in introns of RAPGEF2 abnormal expansions of TTTCA and TTTTA repeats similar to one associated at benign adult familial myoclonic epilepsy. It is associated to susceptibility to childhood (< 16) pneumonia.

2) HSPA14, heat shock protein family A (Hsp70) member 14, is component of the ribosome-associated complex (RAC) involved in folding or maintaining nascent polypeptides in a folding-competent state. In the RAC complex, HSPA14 binds to the nascent polypeptide chain, while its ATPase activity and binding is stimulated.

Fam 8:



In family 8, the proband II.3 is a carrier of 8 RU DRA (35 kb) inherited by his father, which is completely asymptomatic. From the history, no other family members are affected or show dystrophy signs. The proband arrives in counseling at the age of 15; shoulder girdle muscles are first involved by weakness, but parents have already noticed he sleeps with eyes open and has winged scapula years before. Moreover, he has partial ability to protrude her lips and to puff out cheeks with asymmetric involvement (left). He has incomplete abduction of the arms but greater than 45°, horizontal clavicles, hyperlordosis and atrophy of the pectoral muscles. He practiced water polo and football until the age of 12 without any problems. We found 6 candidate mutations, 2 de novo (1 non synonymous SNV, 1 stopgain) and 4 autosomic recessive (1 non synonymous SNV, 1 frameshift substitution, 1 frameshift insertion, 1 frameshift deletion) all in exonic regions (see table).

FAMILY 8			
VARIANTS	GENE REF.	EXONIC FUNC. REF	AA CHANGE
DE NOVO	DEPDC1*	stopgain	p.S240X*
	HOXA3	nonsynonymous SNV	G252S
AUTOSOMAL RECESSIVE	ANKRD55	nonsynonymous SNV	E279K
	ZAN	frameshift substitution	H1922Cfs*19
	SIGLEC12	frameshift insertion	A66Gfs*50
	HLA-A	frameshift deletion	D101Tfs*50, N101Tfs*21

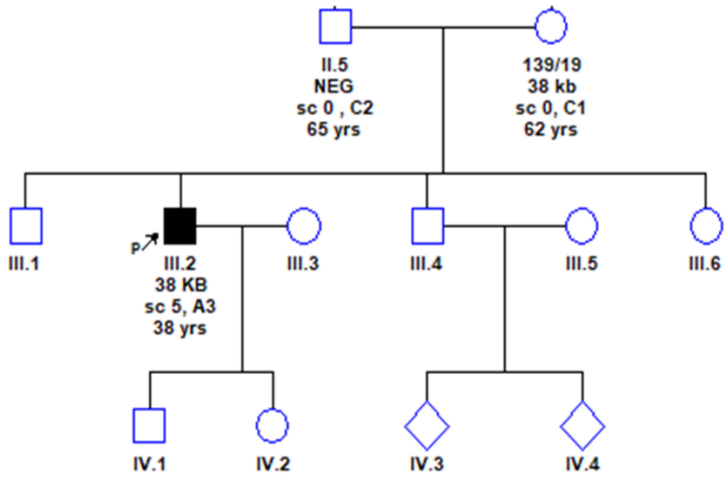
The novo stopgain mutations involved DEPDC1 gene, a GTPase activator, which recent studied and published to be a key regulatore of myoblast proliferation in mouse and man. DEPDC1B is expressed in proliferating murine and human myoblasts, with expression then decreasing markedly during myogenic differentiation. SiRNA-mediated knockdown of DEPDC1B reduce myoblast proliferation and induced entry into myogenic differentiation, with deregulation of key cell cycle regulators (cyclins, CDK, CDKi). DEPDC1B and β-catenin co-knockdown is unable to rescue proliferation in myoblasts, suggesting that DEPDC1B functions independently of canonical WNT signalling during

myogenesis, considering that it is also a positive regulator of Wnt signaling pathway. Genome-wide associated loci prioritisate DEPDC1B as likely causal gene for adolescent idiopathic scoliosis.

2) It is found in this proband a autosomic recessive mutation in ANKRD55, Ankyrin repeat domain-containing protein 55, whose function is very little known. In genomic panels has been associated Oligoarticular Juvenile Idiopathic Arthritis and Rheumatoid Factor-Negative Polyarticular Juvenile Idiopathic Arthritis.

3) A very interesting de novo mutation in HOXA3 is found. Homeobox protein Hox-A3 is a sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis. It is an RNA polymerase II cis-regulator and is involved in process like embryonic skeletal system morphogenesis, angiogenesis, animal organ formation, anterior/posterior pattern specification. It is associated to different musculoskeletal disease but no specific genetic association has been studied or validated.

Fam 9:



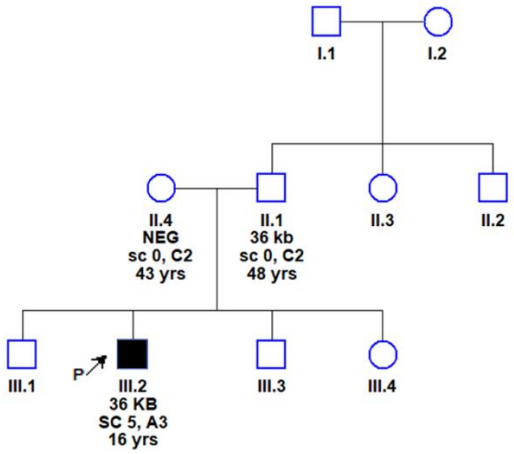
In family 9, the proband III.2 is a carrier of 9 RU DRA (38 kb) inherited by his mother, which is healthy. No other family members were reported as affected or showed dystrophy signs. The proband admitted for clinical examination at the age of 38; he referred appearance of winged shoulder blade at the age of 5. At clinical examination he presented partial ability to close eyes but incapable to close eyes with asymmetric involvement; the lower part of the face was not involved. He has complete but anormal abduction of both arms. Asymmetric scapular winging was observed. The proband presented horizontal clavicles, atrophy of the pectoral muscles and hyperlordosis. He was able to climb stairs but abnormally and presented hyperlordotic and steppage gait. He needed support to stand up from a

chair and rise from the floor. He was able to walk on tiptoes only and no on heels. The Beevor's sign was positive. Analysis of WGS results revealed 7 detrimental variants, 3 de novo (3 non synonymous SNV, 1 stopgain) and 4 autosomic recessive (3 non synonymous SNV, 1 frameshift deletion) all in exonic regions (see table).

FAMILY 9			
VARIANTS	GENE REF.	EXONIC FUNC. REF	AA CHANGE
DE NOVO	FAM90A1	nonsynonymous SNV	P368A
	MUC16	nonsynonymous SNV	L11142P
	TUBGCP6	nonsynonymous SNV	Q1074R
AUTOSOMAL RECESSIVE	PCDHA2	nonsynonymous SNV	A567V
	PCDHA12	nonsynonymous SNV	V758L
	DIAPH1	nonsynonymous SNV	P694A, P703A
	C12orf60	frameshift deletion	M184Dfs*14

For the autosomic recessive mutations, the genes involved with interesting functions in this contex is: 1) DIAPH1, for its homology of the Drosophila gene, it has been speculated that may have a role in the regulation of actin polymerization in hair cells of the inner ear. It is involved in actin cytoskeleton organization, indeed is an actin nucleation and elongation factor in the nucleus and binds to the barbed end of the actin filament, slowing down actin polymerization and depolymerization. In addition to the role during signalling and the regulation of actin dynamics, it functions as a scaffold protein for to stabilize microtubules and promote cell migration, plays a role in the regulation of cell morphology and cytoskeletal organization and it is required in the control of cell shape. Mutations in this gene have been associated hearing loss with or without macrothrombocytopenia, also seizures, cortical blindness, and microcephaly syndrome (SCBMS) considering its role in in brain development.

Fam 10:



In familiy 10, the proband III.2 is a carrier of 9 RU DRA (36 kb) inherited by his father, which is completely asymptomatic. From the history, no other family members are affected or show signs of dystrophy. The proband arrives in counseling at the age of 16; however, he appears the winged shoulder blade at the age of 12. He has a partial ability to protrude lips and puff out cheek without asymmetric involvement, but the upper part of the face is not involved. He has incomplete abduction of the arms but greater than 45°, horizontal clavicles, hyperlordosis and atrophy of the pectoral muscles, in addiction to asymmetric scapular winging. He not has pelvic girdle involvment. The Beevor's sign is positive. He presents infraspinatus and triceps atrophy (MRC score 3). Creatine phosphokinase are < 4x normal value. He practiced cycling since childhood without any problems. We found 5 candidate mutations, 3 de novo (3 non synonymous SNV) and 2 autosomic recessive (2 frameshift deletion) all in exonic regions (see table).

FAMILY 10			
VARIANTS	GENE REF.	EXONIC FUNC. REF	AA CHANGE
DE NOVO	ZCCHC7	nonsynonymous SNV	K508R
	GRIN2C	nonsynonymous SNV	p.F1046S
	GRIN2C	nonsynonymous SNV	p.F1046L
AUTOSOMAL RECESSIVE	OR2B11	frameshift deletion	F8Sfs*2
	APOL4*	frameshift deletion	E111Vfs*21, E108Vfs*21

The interesting mutation de novo is in GRIN2C gene: Glutamate receptor ionotropic-NMDA 2C, is a component of NMDA receptor complexes that function as heterotetrameric ligand-gated ion channels with high calcium permeability and voltage-dependent magnesium sensitivity. It plays a role in regulating the balance between excitatory and inhibitory activity of pyramidal neurons in the prefrontal cortex. It contributes to the slow phase of excitatory postsynaptic current, long-term synaptic potentiation, and learning. Mainly it is expressed in brain but detected in the heart, skeletal muscle and pancreas.

3.5 Preimplantation diagnosis test

FSHD Preimplantation genetic diagnosis (PGD) is a frequent request from families with several affected relatives and is based on the identification of genetic markers. In FSHD, it's necessary that these polymorphic markers segregate with the D4Z4 reduce allele. Considering that the traditional single cell method cannot be performed in FSHD by Southern Blotting technique, we started from the PCR-based multiplex approach of Haouri et al, which select the four most proximal polymorphic markers at 4q35.2 between several available (D4S2390, D4S1652, D4S2930 and D4S1523, <1.23 Mb), showing the highest heterozygote frequencies (67–91%). We selected 3 markers more near D4Z4, located between 550 kb and 210 kb: the microsatellite STS (Sequence-Tagged Site) D4S1523, the VNTR D4S163 and D4S139 (see figure 27). The VNTR reported heterozygosity of 90% for D4S163 and 89% for D4S139. Proximity to D4Z4 is fundamental to lower the probability of recombination events occurring between D4Z4 and the hypothetical markers. Analysis has been completed on 75 individuals from 10 families, 19 individuals in the first generations, 30 individuals in the second generations and 26 individuals in the third generations. Control subjects (n. 96) were also selected to determine allelic frequencies for each marker.

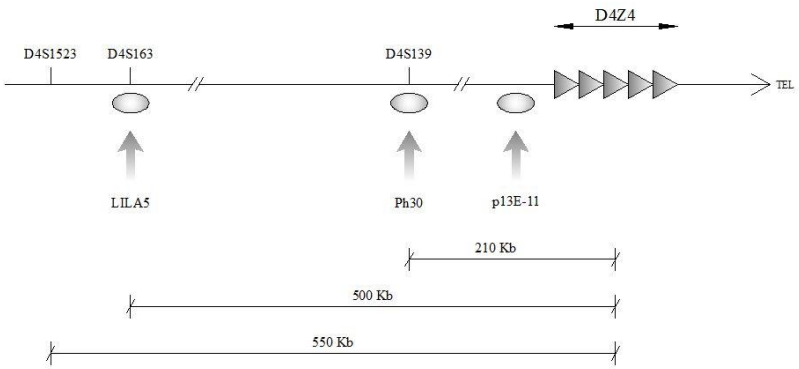


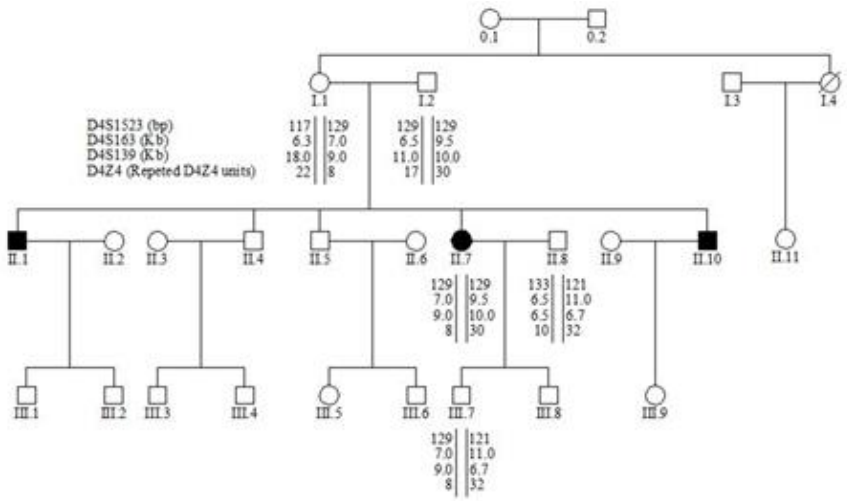
Figure 27. Graphical representation of the analyzed markers. D4S163 and D4S139 are both digested with PstI enzyme and seen with LILA probe and PH30 probe respectively.

3.5.1 Analysis of microsatellite D4S1523

The segregation of alleles was annotated to identify recombination events between D4S1523 and D4Z4. Initially, to evaluate the microsatellite D4S1523 it has been followed the PCR-based multiplex protocol of Barat-Haouri and colleagues and amplicons were subsequently run in 5% acrylamide gel. Despite the size of fragments being as expected, PCR products contained many non-specific

amplicons. To improve the specificity of PCR amplification we set up a nested PCR and a touchdown thermal protocol. In details, first we obtained an amplicon of about 1 Kb that subsequently was used as template for the nested PCR. to lower the size of amplicon to about 125 bp. The nested forward primer was labeiled with a FAM green fluorophore to make the amplification product detectable in capillary electrophoresis. The Applied Biosystems® 3130xi Genetic Analysers created a calibration curve using a fluorescent ladder labelled with ROX fluorophore, therefore, alleles dimension was determined by interpolation. In conclusion, the use of capillary electrophoresis allowed to correctly discriminate and was used as the default technique for further family studies. We observed 100 heterozygous subjects and 25 homozygous subjects: figure 28 depicts the tree (A) and output file (B) of family 1012, heterozygous subjects present two peaks of comparable intensity while homozygous subjects present only one peak of double dimension. This is consistent with the presence of a double quantity of labelled DNA on the same position. Within the families we found 8 different alleles in totally corresponding to D4S1523 for a total of 250 alleles. The low variability of alleles for D4S1523 strongly impairs the direct identification of recombinational events in the analyzed pedigrees. Considering the required information for the use of statistical tools to asses combinations value, we made an evaluation of the allelic frequencies considering only the 33 unrelated individuals from families and 96 unrelated controls. The table 5 summarizes the type and the number of alleles found, the frequencies within selected population and the percentage of homozygous observed. We counted 25 homozygous subjects and thus 100 heterozigous subjects. We calculated the value of expected frequency for each allele (p^2, not showed) and using this data in the equation to calculated heterozygsity (see introduction). We obtained a level of heterozygosity equal to **76,27%**, which corroborate our initial observation of low allelic variability and explains the struggle in individuating recombinant or non-recombinant subjects in the families third generations.

A



B

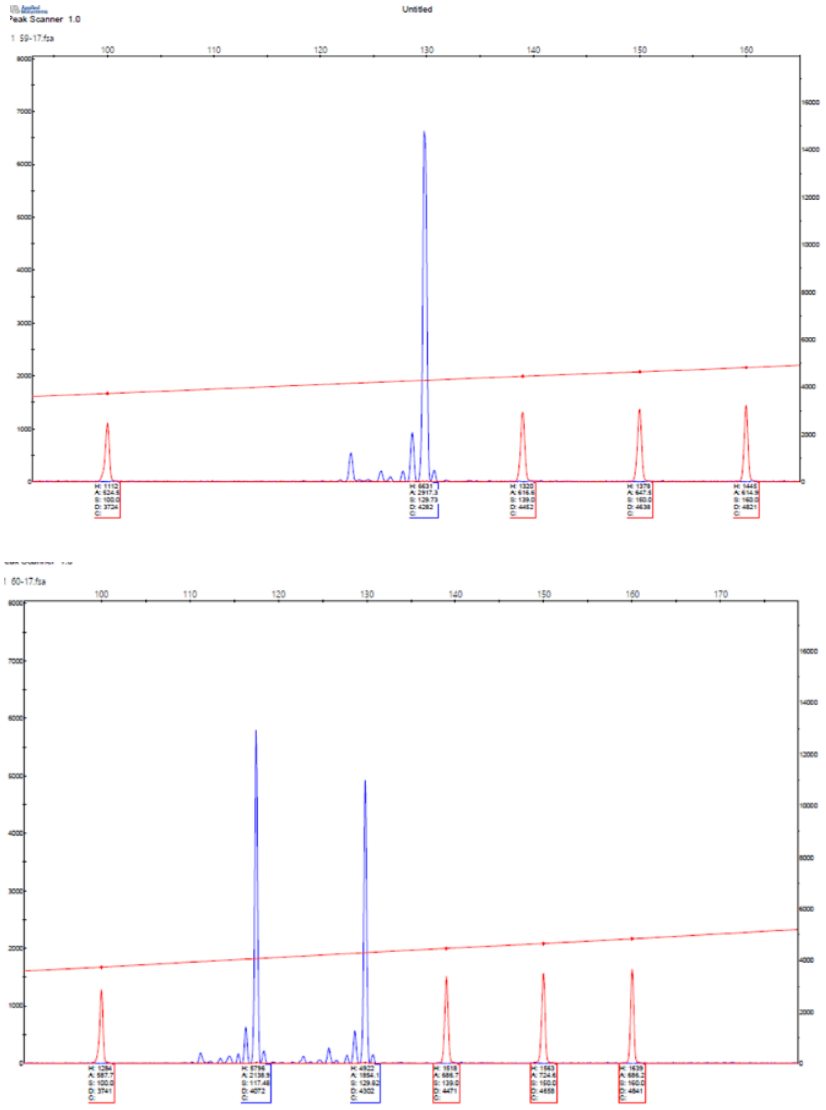


Figure 28. (A) The pedigree of family 1012. For each analyzed subjects the size of D4Z4, D4S1523, D4S163 and D4S139 alleles are reported. (B) Example of homozygous subjects (up) and a heterozygous subject (down) for D4S1523. Horizontal asses report the allele dimension in bp while the vertical asses indicate the peak intensity

ALLELE SIZE D4S1523	REPEAT (TAGA) n SEQUENCES	N. ALLELES	ALLELE FREQUENCIES (p)	N ° HOMOZYGOUS	p hom
109 bp	5	1	0,4 %	0	0 %
113 bp	6	1	0,4 %	0	0 %
117 bp	7	12	4,8 %	0	0 %
121 bp	8	59	23,6 %	4	3,2 %
125 bp	9	42	16,8 %	3	2,4 %
129 bp	10	85	34 %	14	11,2 %
133 bp	11	47	18,8 %	4	3,2 %
137 bp	12	3	1,2 %	0	0 %
TOT		**250**		**25**	

Table 5-Schematic description of allele frequencies of D4S1523 marker. From left, n of alleles found and the number of repeated TAGA sequences associated. Allele frequencies are calculated on n total allele found (250); n. homozygous of total subjects analyzed (125). p hom is the percentage observed of n homozygous on n total subjects analyzed.

3.5.2 Analysis of VNTR D4S163 and D4S139

D4S163 and D4S139 polymorphic markers are detected by Southern blot hybridization of PstI digested DNA, using the probe LILA (920 bp) and the probe PH30 (932 bp) respectively at the same standardized hybridation conditions. The southern blot analysis allows to determine the allele size and trace marker segregation in the family: we performed analysis in all three generation families initially collected. For both probe we obtained a valid result with clearly identifiable alleles. The detected fragments were in the range of expected dimension, from 4.2 kb to 18 kb for D4S163 and from 3.5 kb to 21 kb for D4S139. This enabled us to follow the segregation of the marker in the analyzed families. From literature, we know that for D4S163 the digestion allows to obtained alleles with 90% heterozygosity as calculated by Neuweiler et al. on 72 unrelated Caucasians. D4S139 was firstly reported by Milner et al. and 89% of heterozygosity observed analyzing 85 unrelated Caucasian subjects.

Both markers were analyzed on the 75 individuals from the FSHD collected families. Allelic frequencies of D4S163 were calculated on 34 unrelated individuals from families and 50 unrelated control subjects: 47 different alleles were overall identified and allelic frequencies calculated varied from 0.0059 to 0.0828. A heterozygosity of 95.3% was observed, confirming data previously reported. D4S139 was calculated on 47 control subjects and 35 unrelated individuals from families: 55 different alleles were overall identified of which 39 were found in family's members. Allelic frequencies varied from 0.0061 to 0.0610 and a heterozygosity of **91.5%** was observed. Furthermore, we calculated the allele frequency of D4Z4 on the same 75 individuals from FSHD families selected

in addiction to controls subjects: 45 different alleles were overall identified ranging from 1 to 88 D4Z4 units. Allelic frequencies varied from 0.0018 to 0.0600 and a heterozygosity of **98.5%** was observed.

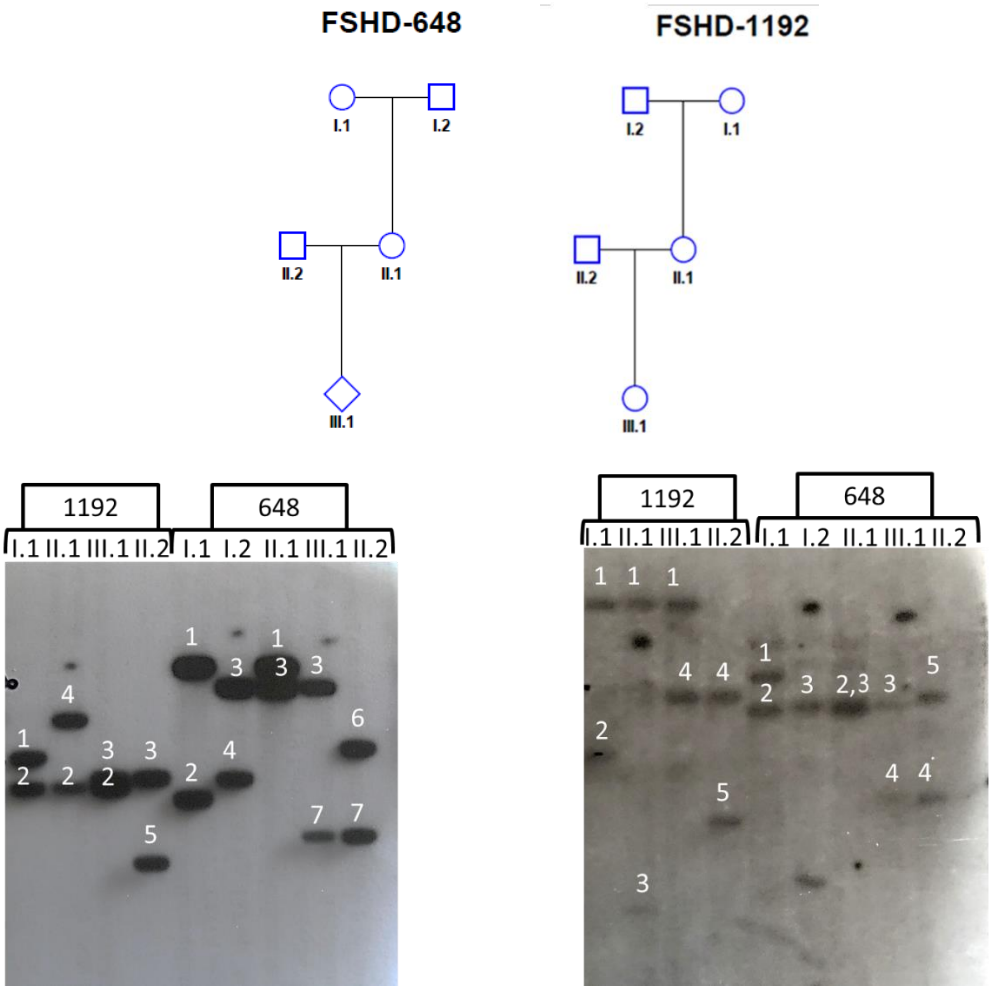


Figura 29. Analysis of D4S163 (sx) and D4S139 using LILA and PH30 probes in two different families. For each allele an arbitrary number was assigned and the marker at the center define allele size.

3.5.3 Computation of the recombination fraction and LOD SCORE

The basic program (therefore the set of programs) is Linkage. The starting file has .txt extension and present several columns: first columns are filled with all family members, alleles at all analyzed loci. A number will be assigned for each information arbitrarily so the same information will have the same code if it is repeated. In this way the program generates other information about the families, like particular conditions of kinship, heredity, the state of disease, description of the loci to adding or excluding in analysis. About recombination, M-link is a program that calculates the LOD SCORE for a certain number of recombination frequencies, that we can set: for example, from 0 to 0.5 separated by 0.1, so the calculation will be done on 0, 0.1, 0.2, 0.3, 0.4,0.5. This allows us to see which is the

most probable recombination fraction which is the one corresponding to the highest LOD SCORE. Subsequently I-LINK evaluates a range of recombination fraction around the one we have assigned and will only return the recombination fraction relative to the higher LOD SCORE. So, if we give different inputs with different recombination fraction values, we can have more probable recombination fraction values in output. The results of M-LINK analysis are the following:

D4S139 MLINK		
θ	Z	Log10(Likelyhood)
0	-∞	
0.1	**6.75294**	-260.257499
0.2	5.673958	-261.336482
0.3	3.918535	-263.091904
0.4	1.956288	-265.054152
0.5	0	-267.010439

D4S163 MLINK		
θ	Z	Log10(Likelyhood)
0	-∞	
0.1	**6.754145**	-252.612749
0.2	5.636527	-253.730367
0.3	3.861039	-255.505855
0.4	1.907696	-257.459199
0.5	0	-259.366894

D4S1523 MLINK		
θ	Z	Log10(Likelyhood)
0	-∞	
0.1	**2.874386**	-186.019321
0.2	2.817447	-186.07626
0.3	2.004418	-186.889289
0.4	0.983814	-187.909893
0.5	0	-188.893707

LOD SCORE equal to or greater than 3 confirm that the marker is in linkage with a given locus that recombination fraction associated. For the two VNTR markers, LOD SCORE around 6 confirms that the most likely recombination fraction is 10% (0.1). For D4S1523, the M-LINK program calculated a maximum LOD SCORE of 2.87 at the recombination fraction of 0.1. To confirm these results, data has obtained with I-LINK, which gives which calculates the data more accurately.

D4Z4-D4S139 ILINK	
θ	Z
0,086	6,780377

D4Z4-D4S163 MLINK	
θ	Z
0,085	6,792649

D4Z4-D4S1523 ILINK	
θ	Z
0,14	3,012391

This data shows that the most probable recombination fraction for D4S1523 microsatellite is at 14% (0.14) and the linkage with D4Z4 is confirmed by LOD SCORE 3. However, these observations are in contrast with the data reported in Barat-Haouri 's work which reported a LOD SCORE of 2.98 to the recombination fraction 0. Together to heterozygosity 76.75% data, these observations confirm that D4S1523 is not reliable enough for an accurate prediction of D4Z4 dimension to be used for PGD in FSHD. Indeed, we would have 86% certainty that specific allele is linked and same observations could be done for the two VNTR. So, to have absolute confirmation We should still do an antenatal subsequently [data unpublished].

4. DISCUSSION

Facioscapulohumeral muscular dystrophy (FSHD, OMIM #158900) is third most common muscular dystrophy with a considerable clinical heterogeneity also in genetically homogeneous cohorts [Ruggero et al 2020; Butz et al 2003]. It is characterized by insidious onset and selective weakness, affecting mainly facial, shoulder girdle muscle, upper limbs, distal legs, pelvic girdle and abdominal [Vercelli et al, 2021]. The published guidelines for the diagnosis and management of FSHD proposed that molecular testing, including the measurement of the size of D4Z4 alleles, the presence of 4qA polymorphism, and the D4Z4 methylation status, become a determinant aspect for diagnosis [Tawill et al, 2015]. While there is a natural tendency to look for a relationship between number of repeat units and clinical severity, this approach has presented several flaws through years and not to be accurate [Ricci et al, 2013, Nikolic et al, 2014, Nikolic et al 2020]. Diagnostic difficulties arise also from an overlap in D4Z4 numbers between controls and FSHD individuals and different phenotype with same size of D4Z4. To date, the considerable variability in the progression has not been defined by predictors of decline of muscle strength and there is no established treatment for FSHD.

A NOVEL PROTOTYPE FOR STUDYING RARE HEREDITARY NEUROMUSCULAR DISEASE

Through time, DNA molecular analysis has become the gold standard for diagnostics of hereditary neuromuscular diseases (HNMDs). However its extensive use revealed an increasingly varied phenotypic spectrum with different disease presentations in patients carrying mutations in the same gene that makes difficult to obtain a precise correlation between the genotype and clinical observations. In a time in which it is possible to obtain genomic data at low cost it is necessary to be aware that detection of a mutation in a specific gene may not precisely describe the clinical phenotype and predict the evolution of disease. Nowadays affected patients presenting similar symptoms, also including imaging of muscles or study of muscle biopsy, may carry mutations in different genes, or people presenting different clinical conditions may carry mutation in the same single gene. These observations reveal that HNMD, as well as many other genetic disorders, are the results of complex biological interactions that are mostly unknown. All these observations open a crucial question: is it possible to develop novel approach to fight diseases that even though they are not fatal, severely affect the social life of a person limiting its activities and independency? At present, there are no adequate social and medical interventions for HNMD patients. The Italian National Registry for facioscapulohumeral muscular dystrophy (FSHD) is the prototype of interventions applied to a rare HNMD. FSHD is a hereditary neuromuscular disease with prevalence of 5 in 100,000. A data

management framework, called MOMIS (Mediator Environment for Multiple Information Sources) FSHD Web Platform, has been developed to provide charts, maps and search tools customized for specific needs. It consists of two main web modules, based on the open-source EDC platform OpenClinica and the MOMIS Data Integration System for data search, analysis, and extraction. The integration of multiple distributed data sources with genomic information, family origin and instrumental investigations will result in a greater depth and fast growth of information that will be managed. The MOMIS FSHD Web Platform is an open framework to enable collaboration and knowledge exchange for biomedical and clinical research. All information from different sources is integrated to constitute a solid support for clinical practice and research on the disease. The tools used to collect, integrate, and visualize clinical, molecular and natural history information about patients affected by FSHD and their relatives are described. All information provides a basis to address complex system challenges in health and health care. The MOMIS FSHD Web Platform has collected the highest number of clinical and molecular data described so far. This result was achieved thanks to the Italian Clinical Network for FSHD (ICNF). Thanks union of the clinical centres and accessibility of National Health Service, we have obtained data representative of the Italian population for FSHD, in fact the number of registered families is 2951 and 532 subjects showed classic phenotype (category A). Our analysis shows that the establishment of the INFR generated a steady relationship between clinicians and parents with an average of 387 cases to be registered each year from 2007, split between index cases and relatives, and this shows that an official data collection makes patients more compliant. The INRF created a valid case series for epidemiological studies and above all it allowed us the follow up of patients and families through the years. For example, in 2020 subjects were registered who belong to families examined 20 years before. So, the Registry creates a relationship with the patient and allows the dynamic study of penetrance, heredity and variability within shared genetic background. Thanks to the Registry, we observed among the most severe phenotypes that females have up to 10 shorter disease duration than males. This is most evident in the DRA cohort, but also in the NO DRA cohort. This strongly suggests an implication of sex-related factors in the progression of the disease.

CARRIERS OF D4Z4 ALLELES VERSUS CARRIERS OF D4Z4 ALLELES OF NORMAL SIZE

Since the description of FSHD2 cases, clinicians were encouraged to study subjects with common FSHD signs who carry D4Z4 allele of normal size (NO DRA) and their families. However, the present study shows that the number of NO DRA single cases is greater three times than of NO DRA family probands in all 4 courts considered. This suggest that the finding of a DRA carrier prompts the

clinician to extended to family study. with DRA and that the study of NO DRA subjects is not extended to the family. However, the fact that of relatives DRA, only 57,8% have phenotypic manifestations, indicates that particular attention should be paid to clinical aspects and clinical conclusions cannot be based on molecular data exclusively. Thanks to the use of the CCEF we can characterize the FSHD population with precision. The CCEF tool has taken data research to a higher level than simply collecting molecular data. It has created new questions and new hypotheses also for molecular research. A 5-year follow-up study found that the clinical phenotype as described by the CCEF categories could be a predictor of disability progression, with more rapid disease progression in individuals with a classic FSHD (category A) phenotype than patients with an incomplete facial phenotype (category B1) [Vercelli et al, 2021]. On the basis of this hypothesis, this study highlighted that in DRA probands of category A (classical phenotype), B1 (incomplete phenotype) and D1 (mixed phenotype) the percentage of having a relative of the same category is the higher in comparison with NO DRA subjects. In the latter group relatives have a homogeneous phenotypic distribution, even though it should be considered the small number of relatives of NO DRA probands that were collected (n.288). Our study also identified 135 D2 subjects (45 DRA and 90 NO DRA) and we observed 16 probands with DRA (12%). This result indicates that molecular analysis is requested patients with signs that are in common with other dystrophies to rule out FSHD diagnosis in the context of differential diagnosis. The age of onset (40) of D2 subjects is higher compared to other phenotypes and this suggest that the clinical signs evaluated could be due to aging and presence of short allele is not linked with develpment other disease. However, D2 DRA probands have relatives with classical phenotype, so it could be interesting studying a correlation with some specific disease. In any case, the data collected in the Registry allow us to deeply the phenotypic heterogeneity, partly resolved by the molecular data but is not defined. This leads to an integrated method for the study of families, creating centralized diagnostics.

COMPARISON BETWEEN PROBANDS AND RELATIVES

Equally important is to observe some characteristics in the relative's study. I found 73 families with DRA probands and NO DRA relatives with phenotipical signs and viceversa. This can be explained by the request of analysis by relatives of affected subjects not yet analysed or positive results in other centres. This request arrives also from clinicians, in fact 28% of registered DRA relatives have a score of 0, and 34% have no defined clinical evaluation. These subjects are often younger than proband and this explains also why the average age of the NO DRA relatives reported is lower than it expected or than NO DRA probands. While clinicians tend not to take responsibility for not making a diagnosis, on the other hand there are biases when family members are examined and certain muscle signs are

not accurately interpreted. Therefore, register's data help us to respect appropriateness of request: if the suspicion does not derive from the presence of muscle deficits, subject should not necessarily do the analysis. Moreover, if the patient has only a typical sign without weakness (such as winged shoulder blades), it would be more correct and prudent to wait for clinical evaluation, and do it in case there is a worsening or development of the dystrophic phenotype. These secondary observations to the general aspects of the clinical and the molecular analysis, suggests that the study of probands and their related needs of two different models. If we remove the number of relatives with score 0 (from the score distribution graph) we obtain a Gaussian, which is typical of complex or multifactorial diseases, but different respect to probands. These factors must be considered in clinical practice, for this INRF can be considered a prototype for the study of complex diseases. It has an educational effect and offer a method of work, as well as a guide for studying families.

INTEGRATED APPROACH WITH GENOMIC ANALYSIS

Patients' genomic and clinical information can be used as big data to extrapolate and relate the enormous amount of heterogeneous, structured and unstructured data uncovers hidden patterns, unknown correlations, and other insights through examining large-scale various data sets. The data reported here was obtained using excel function expansions: it was possible to cross-reference the general data within families, but the program could not handle the volume, speed and variety of all the data collected. In fact, specific information technologies and sophisticated statistical methods are required. Although complicated, big data offers a feasible opportunity to develop an efficient and effective approach to identify genetic and clinical factors that can be used for personalized diagnosis and therapies. To date, over 6000 Mendelian disorders have been studied at the genetic level [McKusick et al, 2007; Brunham et al 2013] and over 1500 clinically-relevant complex traits have been studied with genome-wide association study (GWAS) approaches [Welter et al, 2014]. Next-generation sequencing (NGS) technologies are progressively more applied to biomedical study and medical practice to identify disease- and/or drug-associated genetic variants to advance precision medicine. Additionally, a number of studies have been designed to combine genomic and EHR data to improve clinical research and/or healthcare outcome. [Collins et al 2015; Carter et al, 2016]. Precision medicine allows scientists and clinicians to predict more accurately which therapeutic and preventive approaches to a specific illness can work effectively in subgroups of patients based on their genetic make-up, lifestyle, and environmental factors. Genomic data generated by NGS technologies are a vital component in supporting genomic medicine, but the volume and complexity of the data raise challenges for its use in clinical practice. Therefore, the development of novel

bioinformatics infrastructures is required to implement NGS in clinical practice [Gullapalli et al, 2012].

GENETIC COUNSELING IN A COMPLEX GENETIC DISEASE

In general, a couple with FSHD or from a family in which the disease segregates ask for genetic counseling to make an informed procreative decision. However the wide clinical variability in the FSHD is the greatest challenge in FSHD genetic counselling, Despite that, an increasing number of couples belonging to families with FSHD request preconceptional and prenatal genetic counseling and diagnosis (PND). In the most recent years our laboratory received several questions about preimplantation genetic testing (PGT). The identification of suitable markers for PGD applied to FSHD could be a valuable tool to improve the quality of genetic counseling for this disease. Indeed, this would help geneticists ensure delivery of healthy babies and provide early diagnosis by preventing the need to terminate pregnancy. In the case of FSHD, PGT must rely on an indirect test, because the D4Z4 region cannot be investigated. To investigate the reliability of the test presently available we analyzed the segregation of the loci D4S163 and D4S139, which are located between the D4Z4 locus and D4S1523, used in PGT. The extension of this segregation analysis on a significant number of meiosis allowed performing a linkage analysis to determine the exact fraction of recombination between these loci. Our analysis demonstrates that preimplantation diagnosis for FSHD still requires confirmation of prenatal testing. PGT should be followed by the classic procedure of FSHD molecular diagnosis on CVS. PND and PGT are one of the most sensitive aspects of genetic counseling. In these conditions, multiple aspects of informed decision-making come out, which may involve risks for the individual to procreative autonomy, in addition to the potential post-consequences of selective termination of pregnancy, and the stigmatization and discrimination of disabled people. In the last 15-20 years, awareness of the issue has led to the creation of national and European commissions, such as the Public Health Genomics European Network" (PHGEN) and the Ministerial Commission for Genetics in the National Health Service in Italy. These have carried out an intense activity of dissemination of the "Guidelines for the Activities of Medical Genetics, which define the multidisciplinary of genetic counseling, therefore from medical-genetic information, to the help of the person and the family to understand the available options and more appropriate. It is expected that big data technologies will revolutionize also new-born early genetic screening, intervention and treatment as a pathway to accelerating new therapies. In these years, there has been a massive investment in the progression of genetic testing and precision medicine and this will radically change the course of diseases, including neuromuscular ones. Opportunities that could offset these challenges and promote optimal patient outcomes also include increasing the use of patient

registries, data hubs and other ways to centralize data, and the adoption of technologies for remote appointments and real-time monitoring.

CONCLUSION AND PERSPECTIVES

This work strongly wants to demonstrate that the facioscapulohumeral muscular dystrophy, like other rare or neuromuscular diseases, should be studied by focusing the forces mainly on data collection. The complexity, that characterizes them, could have a daunting impact for clinicians and researchers. Over time, we understood that the data on allele size is insufficient and that phenotypic variability can be assessed with the help of precision medicine. The Registry does not simplify the analysis, but it has the power to diversify and categorize patients and families in an unambiguous way. In this sense, this study has improved our understanding of different clinical conditions, especially within the family context. Based on the differences found between males and females, we would like to understand if and which specific sex-related or environmental factors could be linked to the progression of the disease. We believe that this study contributes to the evolution of genetic counseling, not only with the request for greater involvement of the clinician, but with a more precise and in-depth understanding of factors which condition patient's life.

BIBLIOGRAFIA

- Agresti, A., Meneveri, R., Siccardi, A.G., Marozzi, A., Corneo, G., Gaudi, S., Ginelli, E., 1989. Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family). J. Mol. Biol. 205, 625–631. https://doi.org/10.1016/0022-2836(89)90308-2
- Agresti, A., Rainaldi, G., Lobbiani, A., Magnani, I., Di Lernia, R., Meneveri, R., Siccardi, A.G., Ginelli, E., 1987. Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. Hum. Genet. 75, 326–332. https://doi.org/10.1007/BF00284102
- Bakker, E., Wijmenga, C., Vossen, R.H., Padberg, G.W., Hewitt, J., van der Wielen, M., Rasmussen, K., Frants, R.R., 1995. The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve. Suppl. 2, S39-44.
- Barat-Houari, M., Nguyen, K., Bernard, R., Fernandez, C., Vovan, C., Bareil, C., Khau Van Kien, P., Thorel, D., Tuffery-Giraud, S., Vasseur, F., Attarian, S., Pouget, J., Girardet, A., Lévy, N., Claustres, M., 2010. New multiplex PCR-based protocol allowing indirect diagnosis of FSHD on single cells: can PGD be offered despite high risk of recombination? Eur. J. Hum. Genet. EJHG 18, 533–538. https://doi.org/10.1038/ejhg.2009.207
- Barber, T.D., Barber, M.C., Tomescu, O., Barr, F.G., Ruben, S., Friedman, T.B., 2002. Identification of target genes regulated by PAX3 and PAX3-FKHR in embryogenesis and alveolar rhabdomyosarcoma. Genomics 79, 278–284. https://doi.org/10.1006/geno.2002.6703
- Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R., Kyba, M., 2008. DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Exp. Neurol. 214, 87–96. https://doi.org/10.1016/j.expneurol.2008.07.022
- Brezina, P.R., Brezina, D.S., Kearns, W.G., 2012. Preimplantation genetic testing. BMJ 345, e5908. https://doi.org/10.1136/bmj.e5908
- Brouwer, O.F., Padberg, G.W., Wijmenga, C., Frants, R.R., 1994. Facioscapulohumeral muscular dystrophy in early childhood. Arch. Neurol. 51, 387–394. https://doi.org/10.1001/archneur.1994.00540160085011
- Buckingham, M., Relaix, F., 2007. The Role of Pax Genes in the Development of Tissues and Organs: Pax3 and Pax7 Regulate Muscle Progenitor Cell Functions. Annu. Rev. Cell Dev. Biol. 23, 645–673. https://doi.org/10.1146/annurev.cellbio.23.090506.123438
- Butz, M., Koch, M.C., Müller-Felber, W., Lemmers, R.J.L.F., van der Maarel, S.M., Schreiber, H., 2003. Facioscapulohumeral muscular dystrophy. Phenotype-genotype correlation in patients with borderline D4Z4 repeat numbers. J. Neurol. 250, 932–937. https://doi.org/10.1007/s00415-003-1116-y
- Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., Gabellini, D., 2012. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149, 819–831. https://doi.org/10.1016/j.cell.2012.03.035
- 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., 2013. Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. PLoS Genet. 9, e1003550. https://doi.org/10.1371/journal.pgen.1003550

- Chen, H.-F., Chen, S.-U., Ma, G.-C., Hsieh, S.-T., Tsai, H.-D., Yang, Y.-S., Chen, M., 2018. Preimplantation genetic diagnosis and screening: Current status and future challenges. J. Formos. Med. Assoc. Taiwan Yi Zhi 117, 94–100. https://doi.org/10.1016/j.jfma.2017.08.006
- Chiara, M., Gioiosa, S., Chillemi, G., D'Antonio, M., Flati, T., Picardi, E., Zambelli, F., Horner, D.S., Pesole, G., Castrignanò, T., 2018. CoVaCS: a consensus variant calling system. BMC Genomics 19, 120. https://doi.org/10.1186/s12864-018-4508-1
- Chuenkongkaew, W.L., Lertrit, P., Limwongse, C., Nilanont, Y., Boonyapisit, K., Sangruchi, T., Chirapapaisan, N., Suphavilai, R., 2005. An unusual family with Leber's hereditary optic neuropathy and facioscapulohumeral muscular dystrophy. Eur. J. Neurol. 12, 388–391. https://doi.org/10.1111/j.1468-1331.2004.01060.x
- Clapp, J., Mitchell, L.M., Bolland, D.J., Fantes, J., Corcoran, A.E., Scotting, P.J., Armour, J.A.L., Hewitt, J.E., 2007. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. 81, 264–279. https://doi.org/10.1086/519311
- Dahdouh, E.M., Balayla, J., Audibert, F., Genetics Committee, Wilson, R.D., Audibert, F., Brock, J.-A., Campagnolo, C., Carroll, J., Chong, K., Gagnon, A., Johnson, J.-A., MacDonald, W., Okun, N., Pastuck, M., Vallée-Pouliot, K., 2015. Technical Update: Preimplantation Genetic Diagnosis and Screening. J. Obstet. Gynaecol. Can. JOGC J. Obstet. Gynecol. Can. JOGC 37, 451–463. https://doi.org/10.1016/s1701-2163(15)30261-9
- D'Antona, G., Brocca, L., Pansarasa, O., Rinaldi, C., Tupler, R., Bottinelli, R., 2007. Structural and functional alterations of muscle fibres in the novel mouse model of facioscapulohumeral muscular dystrophy. J. Physiol. 584, 997–1009. https://doi.org/10.1113/jphysiol.2007.141481
- Daxinger, L., Tapscott, S.J., van der Maarel, S.M., 2015. Genetic and epigenetic contributors to FSHD. Curr. Opin. Genet. Dev. 33, 56–61. https://doi.org/10.1016/j.gde.2015.08.007
- de Greef, J.C., Lemmers, R.J.L.F., van Engelen, B.G.M., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R., van der Maarel, S.M., 2009. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum. Mutat. 30, 1449–1459. https://doi.org/10.1002/humu.21091
- Deidda, G., Cacurri, S., Piazzo, N., Felicetti, L., 1996. Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). J. Med. Genet. 33, 361–365. https://doi.org/10.1136/jmg.33.5.361
- Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Mattéotti, C., van Acker, A.M., Leo, O., Figlewicz, D., Barro, M., Laoudj-Chenivesse, D., Belayew, A., Coppée, F., Chen, Y.-W., 2007. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc. Natl. Acad. Sci. U. S. A. 104, 18157–18162. https://doi.org/10.1073/pnas.0708659104
- Dolled-Filhart, M.P., Lee, M., Ou-Yang, C., Haraksingh, R.R., Lin, J.C.-H., 2013. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. ScientificWorldJournal 2013, 730210. https://doi.org/10.1155/2013/730210
- Dorobek, M., Szmidt-Sałkowska, E., Rowińska-Marcińska, K., Gaweł, M., Hausmanowa-Petrusewicz, I., 2013. Relationships between clinical data and quantitative EMG findings in facioscapulohumeral muscular dystrophy. Neurol. Neurochir. Pol. 47, 8–17. https://doi.org/10.5114/ninp.2013.32936
- Duchenne Guillaume- Benjamin, 1869. " Revue photographique des hopitaux de Paris - Bulletin medical ", n.d.
- DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis [WWW Document], n.d. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003947 (accessed 9.3.21).

- Facioscapulohumeral Muscular Dystrophy: More Complex than it Appears - PubMed [WWW Document], n.d. URL https://pubmed.ncbi.nlm.nih.gov/25323867/ (accessed 9.3.21).
- Filosto, M., Tonin, P., Scarpelli, M., Savio, C., Greco, F., Mancuso, M., Vattemi, G., Govoni, V., Rizzuto, N., Tupler, R., Tomelleri, G., 2008. Novel mitochondrial tRNA Leu(CUN) transition and D4Z4 partial deletion in a patient with a facioscapulohumeral phenotype. Neuromuscul. Disord. NMD 18, 204–209. https://doi.org/10.1016/j.nmd.2007.12.005
- Gabellini, D., Green, M.R., Tupler, R., 2002. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell 110, 339–348. https://doi.org/10.1016/s0092-8674(02)00826-7
- Gabriëls, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., Belayew, A., 1999. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene 236, 25–32. https://doi.org/10.1016/s0378-1119(99)00267-x
- Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., Tapscott, S.J., 2012. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev. Cell 22, 38–51. https://doi.org/10.1016/j.devcel.2011.11.013
- Hamel, J., Tawil, R., 2018. Facioscapulohumeral Muscular Dystrophy: Update on Pathogenesis and Future Treatments. Neurother. J. Am. Soc. Exp. Neurother. 15, 863–871. https://doi.org/10.1007/s13311-018-00675-3
- Hanel, M.L., Wuebbles, R.D., Jones, P.L., 2009. Muscular dystrophy candidate gene FRG1 is critical for muscle development. Dev. Dyn. Off. Publ. Am. Assoc. Anat. 238, 1502–1512. https://doi.org/10.1002/dvdy.21830
- Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M., 1994. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum. Mol. Genet. 3, 1287–1295. https://doi.org/10.1093/hmg/3.8.1287
- Jacobsen, S.J., Diala, E.S., Dorsey, B.V., Rising, M.B., Graveline, R., Falls, K., Schultz, P., Hogan, C., Rediker, K., D'Amico, C., 1990. A clinically homogeneous group of families with facioscapulohumeral (Landouzy-Déjérine) muscular dystrophy: linkage analysis of six autosomes. Am. J. Hum. Genet. 47, 376–388.
- Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Mattéotti, C., Arias, C., Corona, E.D., Nuñez, N.G., Leo, O., Wattiez, R., Figlewicz, D., Laoudj-Chenivesse, D., Belayew, A., Coppée, F., Rosa, A.L., 2007. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul. Disord. NMD 17, 611–623. https://doi.org/10.1016/j.nmd.2007.04.002
- Krom, Y.D., Thijssen, P.E., Young, J.M., den Hamer, B., Balog, J., Yao, Z., Maves, L., Snider, L., Knopp, P., Zammit, P.S., Rijkers, T., van Engelen, B.G.M., Padberg, G.W., Frants, R.R., Tawil, R., Tapscott, S.J., van der Maarel, S.M., 2013. Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. PLoS Genet. 9, e1003415. https://doi.org/10.1371/journal.pgen.1003415
- L, D., Sj, T., Sm, van der M., 2015. Genetic and epigenetic contributors to FSHD. Curr. Opin. Genet. Dev. 33. https://doi.org/10.1016/j.gde.2015.08.007
- Lamperti, C., Fabbri, G., Vercelli, L., D'Amico, R., Frusciante, R., Bonifazi, E., Fiorillo, C., Borsato, C., Cao, M., Servida, M., Greco, F., Di Leo, R., Volpi, L., Manzoli, C., Cudia, P., Pastorello, E., Ricciardi, L., Siciliano, G., Galluzzi, G., Rodolico, C., Santoro, L., Tomelleri, G., Angelini, C., Ricci, E., Palmucci, L., Moggio, M., Tupler, R., 2010. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: The FSHD clinical score. Muscle Nerve 42, 213–217. https://doi.org/10.1002/mus.21671

- Landouzy and Dejerine, 1885. De la myopathie atrophique progressive: myopathie héréditaire, sans neuropathie, débutant d, n.d.
- Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., Müller, C.R., 2015. Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. Eur. J. Hum. Genet. EJHG 23, 808–816. https://doi.org/10.1038/ejhg.2014.191
- Lecky, B.R., MacKenzie, J.M., Read, A.P., Wilcox, D.E., 1991. X-linked and FSH dystrophies in one family. Neuromuscul. Disord. NMD 1, 275–278. https://doi.org/10.1016/0960-8966(91)90101-w
- Lemmers, R.J.L.F., de Kievit, P., Sandkuijl, L., Padberg, G.W., van Ommen, G.-J.B., Frants, R.R., van der Maarel, S.M., 2002. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat. Genet. 32, 235–236. https://doi.org/10.1038/ng999
- Lemmers, R.J.L.F., Osborn, M., Haaf, T., Rogers, M., Frants, R.R., Padberg, G.W., Cooper, D.N., van der Maarel, S.M., Upadhyaya, M., 2003. D4F104S1 deletion in facioscapulohumeral muscular dystrophy: phenotype, size, and detection. Neurology 61, 178–183. https://doi.org/10.1212/01.wnl.0000078889.51444.81
- Lemmers, R.J.L.F., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W.E., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., Krom, Y.D., Klooster, R., Sun, Y., den Dunnen, J.T., Helmer, Q., Donlin-Smith, C.M., Padberg, G.W., van Engelen, B.G.M., de Greef, J.C., Aartsma-Rus, A.M., Frants, R.R., de Visser, M., Desnuelle, C., Sacconi, S., Filippova, G.N., Bakker, B., Bamshad, M.J., Tapscott, S.J., Miller, D.G., van der Maarel, S.M., 2012. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat. Genet. 44, 1370–1374. https://doi.org/10.1038/ng.2454
- Lemmers, R.J.L.F., van den Boogaard, M.L., van der Vliet, P.J., Donlin-Smith, C.M., Nations, S.P., Ruivenkamp, C.A.L., Heard, P., Bakker, B., Tapscott, S., Cody, J.D., Tawil, R., van der Maarel, S.M., 2015. Hemizygosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. Hum. Mutat. 36, 679–683. https://doi.org/10.1002/humu.22792
- Lemmers, R.J.L.F., van der Maarel, S.M., van Deutekom, J.C.T., van der Wielen, M.J.R., Deidda, G., Dauwerse, H.G., Hewitt, J., Hofker, M., Bakker, E., Padberg, G.W., Frants, R.R., 1998. Inter-and Intrachromosomal Sub-Telomeric Rearrangements on 4q35: Implications for Facioscapulohumeral Muscular Dystrophy (FSHD) Aetiology and Diagnosis. Hum. Mol. Genet. 7, 1207–1214. https://doi.org/10.1093/hmg/7.8.1207
- Lemmers, R.J.L.F., van der Vliet, P.J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., Miller, D.G., Tapscott, S.J., Tawil, R., Frants, R.R., van der Maarel, S.M., 2010. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650–1653. https://doi.org/10.1126/science.1189044
- Lemmers, R.J.L.F., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M.M., de Knijff, P., Padberg, G.W., Frants, R.R., van der Maarel, S.M., 2007. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. 81, 884–894. https://doi.org/10.1086/521986
- Lin, M.Y., Nonaka, I., 1991. Facioscapulohumeral muscular dystrophy: muscle fiber type analysis with particular reference to small angular fibers. Brain Dev. 13, 331–338. https://doi.org/10.1016/s0387-7604(12)80128-8
- Liu, Q., Jones, T.I., Tang, V.W., Brieher, W.M., Jones, P.L., 2010. Facioscapulohumeral muscular dystrophy region gene-1 (FRG-1) is an actin-bundling protein associated with muscle-attachment sites. J. Cell Sci. 123, 1116–1123. https://doi.org/10.1242/jcs.058958

- Lunt, P.W., Compston, D.A., Harper, P.S., 1989. Estimation of age dependent penetrance in facioscapulohumeral muscular dystrophy by minimising ascertainment bias. J. Med. Genet. 26, 755–760. https://doi.org/10.1136/jmg.26.12.755
- Lunt, P.W., Harper, P.S., 1991. Genetic counselling in facioscapulohumeral muscular dystrophy. J. Med. Genet. 28, 655–664. https://doi.org/10.1136/jmg.28.10.655
- Lunt, P.W., Jardine, P.E., Koch, M.C., Maynard, J., Osborn, M., Williams, M., Harper, P.S., Upadhyaya, M., 1995. Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). Hum. Mol. Genet. 4, 951–958. https://doi.org/10.1093/hmg/4.5.951
- Lyle, R., Wright, T.J., Clark, L.N., Hewitt, J.E., 1995. The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics 28, 389–397. https://doi.org/10.1006/geno.1995.1166
- Masciullo, M., Iannaccone, E., Bianchi, M.L.E., Santoro, M., Conte, G., Modoni, A., Monforte, M., Tasca, G., Laschena, F., Ricci, E., Silvestri, G., 2013. Myotonic dystrophy type 1 and de novo FSHD mutation double trouble: a clinical and muscle MRI study. Neuromuscul. Disord. NMD 23, 427–431. https://doi.org/10.1016/j.nmd.2013.02.002
- Matsumura, T., Goto, K., Yamanaka, G., Lee, J.H., Zhang, C., Hayashi, Y.K., Arahata, K., 2002. Chromosome 4q;10q translocations; comparison with different ethnic populations and FSHD patients. BMC Neurol. 2, 7. https://doi.org/10.1186/1471-2377-2-7
- Meneveri, R., Agresti, A., Marozzi, A., Saccone, S., Rocchi, M., Archidiacono, N., Corneo, G., Della Valle, G., Ginelli, E., 1993. Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. Gene 123, 227–234. https://doi.org/10.1016/0378-1119(93)90128-p
- Minkoff, H., Berkowitz, R., 2014. The case for universal prenatal genetic counseling. Obstet. Gynecol. 123, 1335–1338. https://doi.org/10.1097/AOG.0000000000000267
- Mostacciuolo, M.L., Pastorello, E., Vazza, G., Miorin, M., Angelini, C., Tomelleri, G., Galluzzi, G., Trevisan, C.P., 2009. Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. Clin. Genet. 75, 550–555. https://doi.org/10.1111/j.1399-0004.2009.01158.x
- Nigro, V., Savarese, M., 2016. Next-generation sequencing approaches for the diagnosis of skeletal muscle disorders. Curr. Opin. Neurol. 29, 621–627. https://doi.org/10.1097/WCO.0000000000000371
- Nigro, V., Savarese, M., 2014. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. Acta Myol. Myopathies Cardiomyopathies Off. J. Mediterr. Soc. Myol. 33, 1–12.
- Nikolic, A., Jones, T.I., Govi, M., Mele, F., Maranda, L., Sera, F., Ricci, G., Ruggiero, L., Vercelli, L., Portaro, S., Villa, L., Fiorillo, C., Maggi, L., Santoro, L., Antonini, G., Filosto, M., Moggio, M., Angelini, C., Pegoraro, E., Berardinelli, A., Maioli, M.A., D'Angelo, G., Di Muzio, A., Siciliano, G., Tomelleri, G., D'Esposito, M., Della Ragione, F., Brancaccio, A., Piras, R., Rodolico, C., Mongini, T., Magdinier, F., Salsi, V., Jones, P.L., Tupler, R., 2020. Interpretation of the Epigenetic Signature of Facioscapulohumeral Muscular Dystrophy in Light of Genotype-Phenotype Studies. Int. J. Mol. Sci. 21, E2635. https://doi.org/10.3390/ijms21072635
- Nikolic, A., Ricci, G., Sera, F., Bucci, E., Govi, M., Mele, F., Rossi, M., Ruggiero, L., Vercelli, L., Ravaglia, S., Brisca, G., Fiorillo, C., Villa, L., Maggi, L., Cao, M., D'Amico, M.C., Siciliano, G., Antonini, G., Santoro, L., Mongini, T., Moggio, M., Morandi, L., Pegoraro, E., Angelini, C., Di Muzio, A., Rodolico, C., Tomelleri, G., Grazia D'Angelo, M., Bruno, C., Berardinelli, A., Tupler, R., 2016. Clinical expression of facioscapulohumeral

muscular dystrophy in carriers of 1-3 D4Z4 reduced alleles: experience of the FSHD Italian National Registry. BMJ Open 6, e007798. https://doi.org/10.1136/bmjopen-2015-007798
- Padberg, 1982. Doctoral thesis: Facioscapulohumeral disease, n.d.
- Padberg, G.W., Lunt, P.W., Koch, M., Fardeau, M., 1991. Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul. Disord. NMD 1, 231–234. https://doi.org/10.1016/0960-8966(91)90094-9
- Pearson K, 1933. Two new pedigrees of muscular dystrophy. Ann Eugen 5: 179-191, n.d.
- Pirooznia, M., Goes, F.S., Zandi, P.P., 2015. Whole-genome CNV analysis: advances in computational approaches. Front. Genet. 6, 138. https://doi.org/10.3389/fgene.2015.00138
- Pistoni, M., Shiue, L., Cline, M.S., Bortolanza, S., Neguembor, M.V., Xynos, A., Ares, M., Gabellini, D., 2013. Rbfox1 downregulation and altered calpain 3 splicing by FRG1 in a mouse model of Facioscapulohumeral muscular dystrophy (FSHD). PLoS Genet. 9, e1003186. https://doi.org/10.1371/journal.pgen.1003186
- Rappsilber, J., Ryder, U., Lamond, A.I., Mann, M., 2002. Large-scale proteomic analysis of the human spliceosome. Genome Res. 12, 1231–1245. https://doi.org/10.1101/gr.473902
- Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V., Silvestri, G., Mirabella, M., Mangiola, F., Tonali, P., Felicetti, L., 1999. Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype. Ann. Neurol. 45, 751–757. https://doi.org/10.1002/1531-8249(199906)45:6<751::aid-ana9>3.0.co;2-m
- Ricci, G., Ruggiero, L., Vercelli, L., Sera, F., Nikolic, A., Govi, M., Mele, F., Daolio, J., Angelini, C., Antonini, G., Berardinelli, A., Bucci, E., Cao, M., D'Amico, M.C., D'Angelo, G., Di Muzio, A., Filosto, M., Maggi, L., Moggio, M., Mongini, T., Morandi, L., Pegoraro, E., Rodolico, C., Santoro, L., Siciliano, G., Tomelleri, G., Villa, L., Tupler, R., 2016. A novel clinical tool to classify facioscapulohumeral muscular dystrophy phenotypes. J. Neurol. 263, 1204–1214. https://doi.org/10.1007/s00415-016-8123-2
- Ricci, G., Scionti, I., Alì, G., Volpi, L., Zampa, V., Fanin, M., Angelini, C., Politano, L., Tupler, R., Siciliano, G., 2012. Rippling muscle disease and facioscapulohumeral dystrophy-like phenotype in a patient carrying a heterozygous CAV3 T78M mutation and a D4Z4 partial deletion: Further evidence for "double trouble" overlapping syndromes. Neuromuscul. Disord. NMD 22, 534–540. https://doi.org/10.1016/j.nmd.2011.12.001
- Ricci, G., Scionti, I., Sera, F., Govi, M., D'Amico, R., Frambolli, I., Mele, F., Filosto, M., Vercelli, L., Ruggiero, L., Berardinelli, A., Angelini, C., Antonini, G., Bucci, E., Cao, M., Daolio, J., Di Muzio, A., Di Leo, R., Galluzzi, G., Iannaccone, E., Maggi, L., Maruotti, V., Moggio, M., Mongini, T., Morandi, L., Nikolic, A., Pastorello, E., Ricci, E., Rodolico, C., Santoro, L., Servida, M., Siciliano, G., Tomelleri, G., Tupler, R., 2013a. Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. Brain J. Neurol. 136, 3408–3417. https://doi.org/10.1093/brain/awt226
- Ricci, G., Scionti, I., Sera, F., Govi, M., D'Amico, R., Frambolli, I., Mele, F., Filosto, M., Vercelli, L., Ruggiero, L., Berardinelli, A., Angelini, C., Antonini, G., Bucci, E., Cao, M., Daolio, J., Di Muzio, A., Di Leo, R., Galluzzi, G., Iannaccone, E., Maggi, L., Maruotti, V., Moggio, M., Mongini, T., Morandi, L., Nikolic, A., Pastorello, E., Ricci, E., Rodolico, C., Santoro, L., Servida, M., Siciliano, G., Tomelleri, G., Tupler, R., 2013b. Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. Brain J. Neurol. 136, 3408–3417. https://doi.org/10.1093/brain/awt226
- Rijkers, T., Deidda, G., van Koningsbruggen, S., van Geel, M., Lemmers, R.J.L.F., van Deutekom, J.C.T., Figlewicz, D., Hewitt, J.E., Padberg, G.W., Frants, R.R., van der Maarel,

S.M., 2004. FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. J. Med. Genet. 41, 826–836. https://doi.org/10.1136/jmg.2004.019364

- Rossi, M., Ricci, E., Colantoni, L., Galluzzi, G., Frusciante, R., Tonali, P.A., Felicetti, L., 2007. The Facioscapulohumeral muscular dystrophy region on 4qter and the homologous locus on 10qter evolved independently under different evolutionary pressure. BMC Med. Genet. 8, 8. https://doi.org/10.1186/1471-2350-8-8
- Rudnik-Schöneborn, S., Weis, J., Kress, W., Häusler, M., Zerres, K., 2008. Becker's muscular dystrophy aggravating facioscapulohumeral muscular dystrophy--double trouble as an explanation for an atypical phenotype. Neuromuscul. Disord. NMD 18, 881–885. https://doi.org/10.1016/j.nmd.2008.06.387
- Ruggiero, L., Mele, F., Manganelli, F., Bruzzese, D., Ricci, G., Vercelli, L., Govi, M., Vallarola, A., Tripodi, S., Villa, L., Di Muzio, A., Scarlato, M., Bucci, E., Antonini, G., Maggi, L., Rodolico, C., Tomelleri, G., Filosto, M., Previtali, S., Angelini, C., Berardinelli, A., Pegoraro, E., Moggio, M., Mongini, T., Siciliano, G., Santoro, L., Tupler, R., 2020. Phenotypic Variability Among Patients with D4Z4 Reduced Allele Facioscapulohumeral Muscular Dystrophy. JAMA Netw. Open 3, e204040. https://doi.org/10.1001/jamanetworkopen.2020.4040
- Sacconi, S., Lemmers, R.J.L.F., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., Casarin, A., Pegoraro, E., Tawil, R., Bakker, E., Tapscott, S.J., Desnuelle, C., van der Maarel, S.M., 2013. The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. Am. J. Hum. Genet. 93, 744–751. https://doi.org/10.1016/j.ajhg.2013.08.004
- Sakellariou, P., Kekou, K., Fryssira, H., Sofocleous, C., Manta, P., Panousopoulou, A., Gounaris, K., Kanavakis, E., 2012. Mutation spectrum and phenotypic manifestation in FSHD Greek patients. Neuromuscul. Disord. NMD 22, 339–349. https://doi.org/10.1016/j.nmd.2011.11.001
- Salsi, V., Magdinier, F., Tupler, R., 2020. Does DNA Methylation Matter in FSHD? Genes 11, E258. https://doi.org/10.3390/genes11030258
- Sancisi, V., Germinario, E., Esposito, A., Morini, E., Peron, S., Moggio, M., Tomelleri, G., Danieli-Betto, D., Tupler, R., 2014. Altered Tnnt3 characterizes selective weakness of fast fibers in mice overexpressing FSHD region gene 1 (FRG1). Am. J. Physiol. Regul. Integr. Comp. Physiol. 306, R124-137. https://doi.org/10.1152/ajpregu.00379.2013
- Sarfarazi, M., Wijmenga, C., Upadhyaya, M., Weiffenbach, B., Hyser, C., Mathews, K., Murray, J., Gilbert, J., Pericak-Vance, M., Lunt, P., Frants, R.R., Jacobsen, S., Harper, P.S., Padberg, G.W., 1992. Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: Combined analysis of an international consortium. Am. J. Hum. Genet. 51, 396–403.
- Savarese, M., Sarparanta, J., Vihola, A., Jonson, P.H., Johari, M., Rusanen, S., Hackman, P., Udd, B., 2020. Panorama of the distal myopathies. Acta Myol. Myopathies Cardiomyopathies Off. J. Mediterr. Soc. Myol. 39, 245–265. https://doi.org/10.36185/2532-1900-028
- 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., 2012. Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. J. Med. Genet. 49, 171–178. https://doi.org/10.1136/jmedgenet-2011-100454
- Sevy, A., Cerino, M., Gorokhova, S., Dionnet, E., Mathieu, Y., Verschueren, A., Franques, J., Maues de Paula, A., Figarella-Branger, D., Lagarde, A., Desvignes, J.P., Béroud, C., Attarian, S., Levy, N., Bartoli, M., Krahn, M., Campana-Salort, E., Pouget, J., 2016. Improving molecular diagnosis of distal myopathies by targeted next-generation sequencing. J. Neurol. Neurosurg. Psychiatry 87, 340–342. https://doi.org/10.1136/jnnp-2014-309663

- Shaw, N.D., Brand, H., Kupchinsky, Z.A., Bengani, H., Plummer, L., Jones, T.I., Erdin, S., Williamson, K.A., Rainger, J., Stortchevoi, A., Samocha, K., Currall, B.B., Dunican, D.S., Collins, R.L., Willer, J.R., Lek, A., Lek, M., Nassan, M., Pereira, S., Kammin, T., Lucente, D., Silva, A., Seabra, C.M., Chiang, C., An, Y., Ansari, M., Rainger, J.K., Joss, S., Smith, J.C., Lippincott, M.F., Singh, S.S., Patel, N., Jing, J.W., Law, J.R., Ferraro, N., Verloes, A., Rauch, A., Steindl, K., Zweier, M., Scheer, I., Sato, D., Okamoto, N., Jacobsen, C., Tryggestad, J., Chernausek, S., Schimmenti, L.A., Brasseur, B., Cesaretti, C., García-Ortiz, J.E., Buitrago, T.P., Silva, O.P., Hoffman, J.D., Mühlbauer, W., Ruprecht, K.W., Loeys, B.L., Shino, M., Kaindl, A.M., Cho, C.-H., Morton, C.C., Meehan, R.R., van Heyningen, V., Liao, E.C., Balasubramanian, R., Hall, J.E., Seminara, S.B., Macarthur, D., Moore, S.A., Yoshiura, K., Gusella, J.F., Marsh, J.A., Graham, J.M., Lin, A.E., Katsanis, N., Jones, P.L., Crowley, W.F., Davis, E.E., FitzPatrick, D.R., Talkowski, M.E., 2017. SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. Nat. Genet. 49, 238–248. https://doi.org/10.1038/ng.3743
- Smith, B.K., Holloway, G.P., Reza-Lopez, S., Jeram, S.M., Kang, J.X., Ma, D.W.L., 2010. A decreased n-6/n-3 ratio in the fat-1 mouse is associated with improved glucose tolerance. Appl. Physiol. Nutr. Metab. 35, 699–706. https://doi.org/10.1139/H10-066
- Stadler, G., Rahimov, F., King, O.D., Chen, J.C.J., Robin, J.D., Wagner, K.R., Shay, J.W., Emerson, C.P., Wright, W.E., 2013. Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. Nat. Struct. Mol. Biol. 20, 671–678. https://doi.org/10.1038/nsmb.2571
- Statland, J.M., Sacconi, S., Farmakidis, C., Donlin-Smith, C.M., Chung, M., Tawil, R., 2013. Coats syndrome in facioscapulohumeral dystrophy type 1: frequency and D4Z4 contraction size. Neurology 80, 1247–1250. https://doi.org/10.1212/WNL.0b013e3182897116
- Tasca, G., Pescatori, M., Monforte, M., Mirabella, M., Iannaccone, E., Frusciante, R., Cubeddu, T., Laschena, F., Ottaviani, P., Ricci, E., 2012. Different molecular signatures in magnetic resonance imaging-staged facioscapulohumeral muscular dystrophy muscles. PloS One 7, e38779. https://doi.org/10.1371/journal.pone.0038779
- Tawil, R., van der Maarel, S., Padberg, G.W., van Engelen, B.G.M., 2010. 171st ENMC international workshop: Standards of care and management of facioscapulohumeral muscular dystrophy. Neuromuscul. Disord. NMD 20, 471–475. https://doi.org/10.1016/j.nmd.2010.04.007
- Tonini, M.M.O., Passos-Bueno, M.R., Cerqueira, A., Matioli, S.R., Pavanello, R., Zatz, M., 2004. Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromuscul. Disord. NMD 14, 33–38. https://doi.org/10.1016/j.nmd.2003.07.001
- Traeger-Synodinos, J., 2017. Pre-implantation genetic diagnosis. Best Pract. Res. Clin. Obstet. Gynaecol. 39, 74–88. https://doi.org/10.1016/j.bpobgyn.2016.10.010
- Tsien, F., Sun, B., Hopkins, N.E., Vedanarayanan, V., Figlewicz, D., Winokur, S., Ehrlich, M., 2001. Methylation of the FSHD syndrome-linked subtelomeric repeat in normal and FSHD cell cultures and tissues. Mol. Genet. Metab. 74, 322–331. https://doi.org/10.1006/mgme.2001.3219
- TYLER, F.H., STEPHENS, F.E., 1950. Studies in disorders of muscle. II Clinical manifestations and inheritance of facioscapulohumeral dystrophy in a large family. Ann. Intern. Med. 32, 640–660. https://doi.org/10.7326/0003-4819-32-4-640
- Upadhyaya, M., Jardine, P., Maynard, J., Farnham, J., Sarfarazi, M., Wijmenga, C., Hewitt, J.E., Frants, R., Harper, P.S., Lunt, P.W., 1993. Molecular analysis of British facioscapulohumeral dystrophy families for 4q DNA rearrangements. Hum. Mol. Genet. 2, 981–987. https://doi.org/10.1093/hmg/2.7.981

- Upadhyaya, M., Lunt, P.W., Sarfarazi, M., Broadhead, W., Daniels, J., Owen, M., Harper, P.S., 1991. A closely linked DNA marker for facioscapulohumeral disease on chromosome 4q. J. Med. Genet. 28, 665–671. https://doi.org/10.1136/jmg.28.10.665
- Upadhyaya, M., Lunt, P.W., Sarfarazi, M., Broadhead, W., Daniels, J., Owen, M., Harper, P.S., 1990. DNA marker applicable to presymptomatic and prenatal diagnosis of facioscapulohumeral disease. Lancet Lond. Engl. 336, 1320–1321. https://doi.org/10.1016/0140-6736(90)93005-a
- Upadhyaya, M., MacDonald, M., Ravine, D., 1999. Prenatal diagnosis for facioscapulohumeral muscular dystrophy (FSHD). Prenat. Diagn. 19, 959–965.
- Upadhyaya, M., Maynard, J., Rogers, M.T., Lunt, P.W., Jardine, P., Ravine, D., Harper, P.S., 1997. Improved molecular diagnosis of facioscapulohumeral muscular dystrophy (FSHD): validation of the differential double digestion for FSHD. J. Med. Genet. 34, 476–479. https://doi.org/10.1136/jmg.34.6.476
- van der Maarel, S.M., Tawil, R., Tapscott, S.J., 2011. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol. Med. 17, 252–258. https://doi.org/10.1016/j.molmed.2011.01.001
- van Deutekom, J.C., Bakker, E., Lemmers, R.J., van der Wielen, M.J., Bik, E., Hofker, M.H., Padberg, G.W., Frants, R.R., 1996. Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. Hum. Mol. Genet. 5, 1997–2003. https://doi.org/10.1093/hmg/5.12.1997
- van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., Frants, R.R., 1993. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum. Mol. Genet. 2, 2037–2042. https://doi.org/10.1093/hmg/2.12.2037
- van Geel, M., Dickson, M.C., Beck, A.F., Bolland, D.J., Frants, R.R., van der Maarel, S.M., de Jong, P.J., Hewitt, J.E., 2002. Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics 79, 210–217. https://doi.org/10.1006/geno.2002.6690
- van Koningsbruggen, S., Dirks, R.W., Mommaas, A.M., Onderwater, J.J., Deidda, G., Padberg, G.W., Frants, R.R., van der Maarel, S.M., 2004. FRG1P is localised in the nucleolus, Cajal bodies, and speckles. J. Med. Genet. 41, e46. https://doi.org/10.1136/jmg2003.012781
- van Overveld, P.G., Lemmers, R.J., Deidda, G., Sandkuijl, L., Padberg, G.W., Frants, R.R., van der Maarel, S.M., 2000. Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum. Mol. Genet. 9, 2879–2884. https://doi.org/10.1093/hmg/9.19.2879
- van Overveld, P.G.M., Lemmers, R.J.F.L., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.-J.B., Frants, R.R., van der Maarel, S.M., 2003. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat. Genet. 35, 315–317. https://doi.org/10.1038/ng1262
- Vasli, N., Laporte, J., 2013. Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders. Acta Neuropathol. (Berl.) 125, 173–185. https://doi.org/10.1007/s00401-012-1072-7
- Vasyutina, E., Stebler, J., Brand-Saberi, B., Schulz, S., Raz, E., Birchmeier, C., 2005. CXCR4 and Gab1 cooperate to control the development of migrating muscle progenitor cells. Genes Dev. 19, 2187–2198. https://doi.org/10.1101/gad.346205
- Vercelli, L., Mele, F., Ruggiero, L., Sera, F., Tripodi, S., Ricci, G., Vallarola, A., Villa, L., Govi, M., Maranda, L., Di Muzio, A., Scarlato, M., Bucci, E., Maggi, L., Rodolico, C., Moggio, M., Filosto, M., Antonini, G., Previtali, S., Angelini, C., Berardinelli, A., Pegoraro, E., Siciliano, G., Tomelleri, G., Santoro, L., Mongini, T., Tupler, R., 2021. A 5-year clinical

follow-up study from the Italian National Registry for FSHD. J. Neurol. 268, 356–366. https://doi.org/10.1007/s00415-020-10144-7
- Walton, J.N., Nattrass, F.J., 1954. On the classification, natural history and treatment of the myopathies. Brain 77, 169–231.
- Wijmenga, C., Frants, R.R., Brouwer, O.F., Moerer, P., Weber, J.L., Padberg, G.W., 1990. Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet Lond. Engl. 336, 651–653. https://doi.org/10.1016/0140-6736(90)92148-b
- Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.-M., Hofker, M.H., Moerer, P., Williamson, R., van Ommen, G.-J.B., Padberg, G.W., Frants, R.R., 1992. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat. Genet. 2, 26–30. https://doi.org/10.1038/ng0992-26
- Wijmenga, C., Padberg, G.W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O.F., Milner, E.C., Weber, J.L., van Ommen, G.B., Sandkuyl, L.A., 1991. Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics 9, 570–575. https://doi.org/10.1016/0888-7543(91)90348-i
- Wijmenga, C., van Deutekom, J.C., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., Frants, R.R., 1994. Pulsed-field gel electrophoresis of the D4F104S1 locus reveals the size and the parental origin of the facioscapulohumeral muscular dystrophy (FSHD)-associated deletions. Genomics 19, 21–26. https://doi.org/10.1006/geno.1994.1006
- Wright, T.J., Wijmenga, C., Clark, L.N., Frants, R.R., Williamson, R., Hewitt, J.E., 1993. Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. Hum. Mol. Genet. 2, 1673–1678. https://doi.org/10.1093/hmg/2.10.1673
- Yao, Y.L., Yang, W.M., Seto, E., 2001. Regulation of transcription factor YY1 by acetylation and deacetylation. Mol. Cell. Biol. 21, 5979–5991. https://doi.org/10.1128/MCB.21.17.5979-5991.2001
- Zatz, M., Marie, S.K., Passos-Bueno, M.R., Vainzof, M., Campiotto, S., Cerqueira, A., Wijmenga, C., Padberg, G., Frants, R., 1995. High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. Am. J. Hum. Genet. 56, 99–105.
- Zeng, W., de Greef, J.C., Chen, Y.-Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., Kimonis, V.E., Balog, J., Frants, R.R., Ball, A.R., Lock, L.F., Donovan, P.J., van der Maarel, S.M., Yokomori, K., 2009. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet. 5, e1000559. https://doi.org/10.1371/journal.pgen.1000559
- Zhang, X.Y., Loflin, P.T., Gehrke, C.W., Andrews, P.A., Ehrlich, M., 1987. Hypermethylation of human DNA sequences in embryonal carcinoma cells and somatic tissues but not in sperm. Nucleic Acids Res. 15, 9429–9449. https://doi.org/10.1093/nar/15.22.9429