

The Italian National Registry for FSHD: An Enhanced Data Integration and an Analytics Framework Towards Smart Health Care and Precision Medicine for a Rare Disease

Cinzia Bettio

University of Modena and Reggio Emilia Department of Biomedical Sciences: Università degli Studi di Modena e Reggio Emilia Dipartimento di Scienze Biomediche Metaboliche e Neuroscienze

Valentina Salsi

University of Modena and Reggio Emilia Department of Life Sciences: Università degli Studi di Modena e Reggio Emilia Dipartimento di Scienze della Vita

Mirko Orsini

DataRiver SRL

Enrico Calanchi

DataRiver SRL

Luca Magnotta

DataRiver SRL

Luca Gagliardelli

University of Modena and Reggio Emilia Faculty of Engineering - Modena Campus: Università degli Studi di Modena e Reggio Emilia Dipartimento di Ingegneria Enzo Ferrari

Sonia Bergamaschi

University of Modena and Reggio Emilia Department of Material and Environmental Engineering: Università degli Studi di Modena e Reggio Emilia Dipartimento di Ingegneria Enzo Ferrari

Rossella Tupler (✉ rossella.tupler@unimore.it)

Università degli Studi di Modena e Reggio Emilia <https://orcid.org/0000-0002-3797-1789>

Research Article

Keywords: Rare disease registry, Data collection, Data integration, FSHD

Posted Date: May 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-447106/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

BACKGROUND

The *Italian Clinical network for FSHD* (ICNF) has established the Italian National Registry for FSHD (INRF), collecting data from patients affected by Facioscapulohumeral dystrophy (FSHD) and their relatives. The INRF has gathered data from molecular analysis, clinical evaluation, anamnestic information, and family history from more than 3500 participants.

METHODS

A data management framework, called MOMIS FSHD Web Platform, has been developed to provide charts, maps and search tools customized for specific needs. Patients' samples and their clinical information derives from the *Italian Clinical network for FSHD* (ICNF), a consortium consisting of fourteen neuromuscular clinics distributed across Italy. The tools used to collect, integrate, and visualize clinical, molecular and natural history information about patients affected by FSHD and their relatives are described.

RESULTS

The INRF has collected the molecular data of FSHD conducted on 7197 subjects, and identified 3362 individuals carrying a DRA: 1634 are unrelated individuals, 602 isolated cases. In 1032 cases the molecular testing has been extended to 3747 relatives, 1728 carrying a DRA. Since 2009 molecular analysis has been accompanied by clinical evaluation based standardized evaluation protocols. In total 3577 clinical forms have been collected, 2059 follow the Comprehensive Clinical Evaluation form (CCEF). The integration of standardized clinical information and molecular data has made possible to demonstrate the wide phenotypic variability of FSHD. The MOMIS (Mediator Environment for Multiple Information Sources) data integration framework allowed performing genotype-phenotype correlation studies, and generated information of medical importance either for clinical practice or genetic counseling.

CONCLUSION

The platform implemented for the FSHD Registry data collection based on OpenClinica meets the requirement to integrate patient/disease information, as well as the need to adapt dynamically to security and privacy concerns. Our results indicate that the quality of data collection in a multi-integrated approach is fundamental for clinical and epidemiological research in a rare disease and allows to redefine diagnostic criteria and disease markers for FSHD.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) (MIM#158900), the third most common hereditary myopathy with prevalence of 1 in 20,000 [1], is characterized by progressive and variable atrophy and weakness of the facial, shoulder, and upper-arm muscles[2]. Wide variability of the clinical spectrum

ranging from asymptomatic subjects to patients who are wheelchair dependent has been described among and within FSHD families [3]. Before the advent of molecular diagnosis, penetrance of FSHD was evaluated on clinical examination of patients at risk and estimated between 83% and 95% by the age of 20 years[4]. FSHD is considered an autosomal dominant disorder, associated with rearrangements occurring in a 3.3 kilobase (kb) tandemly repeated sequence (D4Z4) located at the 4q subtelomere[5]. The analysis of FSHD penetrance [3,6–13] reinforced the idea that additional elements take part in disease onset and progression. Remarkably the clinical variability and non-penetrance observed in presence of the same molecular defect affect diagnosis, prognosis, genetic counseling. Besides having a great impact in clinical practice, it is also affecting research and readiness to clinical trials. To address these issues, the Italian Clinical Network for FSHD (ICNF) developed clinical tools, the FSHD scale (that computes the FSHD score [14], and the Comprehensive Clinical Evaluation Form (CCEF) [15], for the standardized evaluation of subjects. Based on the observed phenotypic features, one individual can be assigned to one of four different phenotypic categories (A, B, C, D). To integrate molecular data with clinical data the Italian National Consortium for FSHD (INCF) established the Italian National Registry for FSHD (INRF). The INRF has accrued 3362 subjects carrying *D4Z4 Reduced Alleles* (DRA), including 1032 families of unrelated subjects. This large collection of data, the largest world-wide, has enormous potential for providing clinical and epidemiological information and requires suitable tools for proper collection, storage and analysis.

In 2016, a collaboration between the INCF and DataRiver S.r.l. realized the *MOMIS FSHD Web Platform*. 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.

Methods

MOMIS Framework Architecture

The MOMIS FSHD Web Platform integrates data coming from different, heterogeneous and fragmented data sources. The platform provides a unified vision of integrated clinical data to facilitate data exploration and analytics through the MOMIS data integration system.

Figure 1 shows the architecture of the FSHD Registry. Integrated data came from three different sources:

- *Family Pedigree Charts* reporting the family pedigree of each index case;
- *Molecular DB web platform* contains molecular analysis data of patients and their relatives;
- *Clinical DB web platform* contains clinical data of patients and their relatives.

For the integration of these three data sources, we used the MOMIS (Mediator Environment for Multiple Information Sources) data integration framework designed by the DBGROUP at the University of Modena & Reggio Emilia and developed by DataRiver. MOMIS builds a virtual unified schema, which preserves the independence and security of original data sources allowing users to formulate queries on it. The integration process consists of four main steps: (i) connection sources and automatic extraction of schema of the selected data sources; (ii) automatic annotation of the schemas attributes, i.e., one or more meanings is associated to each of them; (iii) attributes with the same meanings are automatically connected by semantic mappings; (iv) finally, a global schema is generated and can be manually refined by the data integrator designer.

MOMIS has already been used in several integration projects [16], including clinical trials and public authorities data management [17]. The MOMIS FSHD Web Platform relies on OpenClinica™ Community Edition (Fig. 1) to manage Clinical Data information, strengthening data quality and efficiency. The data collected through OpenClinica are displayed through the MOMIS Dashboard [16,17]: a MOMIS framework's web tool for data analysis and monitoring tool realized by DataRiver (Fig. 1).

OpenClinica database

The INRF Database collects clinical and molecular data from both index cases and their relatives.

To manage an auditable collection of clinical data we used the open-source web application OpenClinica™ Community Edition. The OpenClinica Electronic Data Capture (EDC) system is a widely used and well-trusted clinical research web platform used for all types of clinical studies, population and rare disease registries. It is currently used by hundreds of research centers and institutions in over 100 countries and can rely on a community of about 18,000 users. The platform is compliant with NIH, HIPAA, 21 CFR Part 11 requirements and regulatory guidelines for clinical data management [18]. OpenClinica is designed as a standards-based, extensible, and modular platform, to guarantee the absolute and strict compliance with the confidentiality and security constraints of access to highly sensitive data.

The OpenClinica platform has been extended with the EDC module to provide the user with specific features to manage sample providers, families, probands and pedigrees. Blood relationships between subjects are crucial to study in detail the families and sporadic cases in which there is an identified molecular defect typical of FSHD. The customized software modules allow users to manage family relationships, identifying the proband (the first family subject affected by disease) and streamlining searches on family branches that could be far from their geographical origin. Through a secure connection, the EDC module allows users to easily switch between the MOMIS Dashboard search and monitoring tool; the custom software modules to manage families, probands and subjects; the data capture forms implemented on OpenClinica.

MOMIS Data processing workflow

When the sample of a new patient arrives at the Miogen Lab, the MOMIS FSHD Web Platform provides a new and unique identifier in the format ID/YEAR. The system then creates two new *Events* in OpenClinica: "Registration" and "Visit". The *Registration* form is designed to collect the participant's personal data and

other information concerning the arrival day of the sample, the National Health System request, and the informed consent. The visit event includes the *Molecular* and the *Comprehensive Clinical Evaluation Form* (CCEF) modules. Case Report Forms (CRFs) are designed to accurately collect all this information from each subject investigated by the ICNF. This allows performing a solid epidemiologic data analysis based on the study of families and sporadic cases.

The modularity, the flexibility, and the ability to create custom forms and fields made the platform a very suitable resource to accomplish our purposes.

The MOMIS Dashboard in the clinical practice of FSHD:

1-The Molecular module

The Molecular module is used for tracking the sample in each step of molecular analyses (e.g., the identifiers of the autoradiographs showing the results of molecular investigation, or the sample storage location). As the final step of the molecular analysis, the allelic profile of the patient is reported. The molecular report contains the size of each D4Z4 allele at specific loci (e.g., Chr-4 q35 or Chr-10 q26), alleles of the qA/qB polymorphism. Additional molecular information such as the presence of a *polyadenylation signal* (PAS) and/or *simple sequence length polymorphism* (SSLP) or D4Z4 methylation status can be uploaded. It is also possible to register additional molecular information such as Next Generation Sequencing data.

2-The Comprehensive Clinical Evaluation Form

Medical experience and systematic clinical evaluation of hundreds of participants guided the evolution of the clinical form from the original FSHD Clinical Form, created in 2009 to the *Comprehensive Clinical Evaluation Form* (CCEF) [15], which collects several additional information such as comorbidities, common and uncommon signs of the disease, lifestyle habits, and more importantly the clinical categories. The CCEF has been validated and frozen in 2016 [15]. The CCEF is organized in four sections as summarized in Table 1.

The *FSHD Evaluation Scale* is used to calculate the FSHD clinical score based on the regional distribution of muscle weakness and the functionality of muscle groups. The FSHD score ranges from 0 to 15 and indicates the severity of the motor impairment.

The *Clinical Diagnostic section* is used to report the combination of typical or uncommon clinical features of FSHD.

The *Sect. 4, Clinical Categories*, assigns the individual to one out of four different phenotypic categories: typical FSHD (category A, subcategories A1, A2, A3), incomplete phenotypes (category B, subcategories B1, B2), asymptomatic/healthy subjects (category C, subcategories C1, C2) and atypical/complex phenotypes (category D, subcategories D1, D2)

The development of the Molecular report and the CCEF in OpenClinica has generated unprecedented possibilities towards the understanding of the FSHD disease. The collection and management of numeric and objective data regarding patients and families makes the search for patterns and correlations not only feasible but also easy.

3-The Filtering module

In the clinical practice, the MOMIS FSHD Web Platform can be used to obtain an integrated view of data coming from several commercial DBMSs. The graphical interface of MOMIS Dashboard allows a quick and effective data extraction from the database. Since 2007, the Italian Registry for FSHD has collected data coming from 7194 subjects (3362 DRA carriers). For a first cursory analysis subjects can be grouped, classified, and filtered for one or more features. For this reason, the MOMIS FSHD Web Platform's homepage provides different features to filter the overall database population and to extract valuable information in the Microsoft Excel Spreadsheet format.

The filter options include:

1. **Subject:** this option filters by subject ID, name, surname, restricting the search to only 'proband' (index cases of unrelated families). It is also possible to select only subjects who received a FSHD Score (an index of degree muscle impairment), within a configurable range.
2. **Category:** it is possible to filter by the phenotypic clinical category. On the basis of the observed phenotypic features, one individual can be assigned to one out of four different phenotypic clinical categories: typical FSHD (category A, subcategories A1, A2, A3), incomplete phenotypes (category B, subcategories B1, B2), asymptomatic/healthy subjects (category C, subcategories C1, C2) and atypical/complex phenotypes (category D, subcategories D1, D2) [15].
3. **Family:** filters by family name.
4. **Registration Year:** the year of patient registration in the database can be selected.
5. **Alleles:** filters by the characteristics of the monitored alleles, e.g., encountered peculiarities, type of translocation and allele size.

The MOMIS FSHD Web Platform shows other different tabs, clustering patients for other features, including:

1. The number of subjects registered every year as well as the enrolment trend is a metric that is useful for planning resources and procurement.
2. Subjects enrolled by location, showing on a map the geographic distribution of the patients.
3. The distribution of alleles within subjects accrued by the INRF, on the basis of their size, for a specific phenotypic category.

All these filtering options can be combined for a finer level of analysis and The MOMIS FSHD Web Platform could then generate graphs allowing an overview of data.

Results

Data collection and management model for a rare disease

The MOMIS FSHD Web Platform is the Italian national data repository for FSHD and represents a fundamental tool for collecting and managing anamnestic, clinical, and molecular data of all Italian subjects accrued through the INRF. Biological samples of participants and their clinical information were collected by the ICNF, which consists of 14 centers specialized in neuromuscular diseases located across Italy and a centralized laboratory, the Miogen lab, dedicated to molecular analyses for FSHD (Fig. 2). The Miogen lab has been a reference center for the molecular diagnosis of FSHD which has been conducted on 7197 subjects. 3362 individuals carry a DRA: 1634 are unrelated individuals, 602 isolated cases and 1032 familiar cases. These families, 3747 relatives were subject to molecular testing. Of 3747 subjects molecularly evaluated, 1728 resulted to be carriers of a DRA.

Thanks to the analysis of 3577 clinical forms collected since 2009, it has been possible establishing that 2165 participants (60.5%) had muscle impairment with different degree of severity, from mild (FSHD score 1–2) to moderate (FSHD score 3–7) and severe (FSHD score 8–15). In the period 2015–2019, 2059 subjects were evaluated following the CCEF, which lead to a more precise phenotypic description. Through the CCEF it was possible to demonstrate that the wide phenotypic variability of participants: (55.4%) present a myopathic phenotype compatible with FSHD (Clinical Categories A, B and D1) and 134 (6.5%) presented muscle impairment that was not compatible with FSHD diagnosis. From the time the CCEF has been applied for the clinical evaluation in 2015, the enrollment of subjects with a myopathic FSHD-like phenotype has been steady (Fig. 3A). Overall, since the establishment of the INRF in 2006, the number of enrolled subjects has been steadily increasing, until a plateau of about 300 subjects per year was reached in 2015 (Fig. 3B). Taken together these data, Since FSHD prevalence is 1 in 20,000 [1] with an estimate of 3,030 people affected by FSHD in Italy, indicate that the collection of subjects accrued by the INRF might be largely representative of the Italian population. As a result, this centralized data collection offers the possibility of conducting clinical and epidemiological studies with high statistical power.

The MOMIS FSHD Web Platform application in clinical research

The most relevant feature of the MOMIS FSHD Web Platform is the population-based systematic collection of clinical and anamnestic data. This has fostered clinical and epidemiological research in a rare disease such as FSHD. It is now clear that clinical variability is a relevant part of hereditary diseases and that the disease expression and evolution is determined by several factors. In the case of FSHD this aspect is of great importance especially because the molecular marker used for FSHD diagnosis is a common polymorphism[19]. Assumed that the presence of DRA constitutes a susceptibility factor there might be several elements contributing to disease expression. In particular, the genotype-phenotype correlation on the basis of the MOMIS FSHD Web Platform's database has generated information of medical importance either for clinical practice or for genetic counseling (Table 2). Based on our earlier consideration, it clearly appears that one of the most important usage of the MOMIS FSHD Web Platform is represented by its

filtering options that allows clustering subjects based on specific characteristics. The selection of cohorts of subjects for epidemiological studies aimed at addressing specific questions is based on the MOMIS interrogation, which strongly shortens the time for subject's collection and minimizes errors. Patients are selected based on the information reported in the database, including anamnestic information. Selection can be very specific: for example, it is possible to select people from a certain region, carrying DRA with 4–8 repeat units, who practiced sports; or we may filter for females, category A, with at least one pregnancy belonging to a family with other relatives affected. Another valuable aspect of the MOMIS FSHD Web Platform is the ability to carry out family studies, defining in each family the characteristics of the proband and relatives. As shown in Table 1, the platform has been successfully applied to many different aspects of FSHD research, starting from the CCEF validation[15] to genotype/phenotype characterization of people carrying specific molecular features: 1–3 DRA[13] or 7–8 DRA[20] or 9–10[21] and to follow up studies[22]. The MOMIS FSHD Web Platform has also been used for molecular research on D4Z4 epigenetic regulation [23] or to identify genetic interactions with other rare diseases[24].

Table 1
Description of the main features in Comprehensive Clinical Evaluation Form (CCEF)

SECTION 1: THE EVALUATION FORM	SECTION 2: FSHD EVALUATION SCALE AND FSHD CLINICAL SCORE	SECTION 3: CLINICAL DIAGNOSTIC FORM	SECTION 4: CLINICAL CATEGORY ASSESSMENT
<p>Evaluation form collects several information and is subdivided in three part:</p> <p><i>Part A</i></p> <p>Investigates the subject's clinical history</p> <p><i>Part B</i></p> <p>Evaluates the patient's disability (neurological examination)</p> <p><i>Part C</i></p> <p>Assesses muscle segmental involvement using the <i>Medical Research Council</i> (MRC) scale</p>	<p>Evaluation scale allows the computation of the FSHD SCORE which is based on the regional distribution of muscle weakness and the functionality of muscle groups.</p> <p>Five groups of muscles are investigated and evaluated by a specific score:</p> <p>1 <i>facial muscles</i> (scored from 0 to 2)</p> <p>2 <i>scapular girdle muscles</i> (scored from 0 to 3)</p> <p>3 <i>upper limb muscles</i> (scored from 0 to 2)</p> <p><i>leg muscles</i> (scored from 0 to 2)</p> <p>4 <i>pelvic girdle muscles</i> (scored from 0 to 5)</p> <p>5 <i>abdominal muscles</i> (scored from 0 to 1).</p> <p>The FSHD score ranges from 0 to 15 and indicates the severity of the motor impairment.</p>	<p>This section established the assessment of typical and atypical features of FSHD.</p> <p>An accurate evaluation of these features allows to assign clinical categories.</p>	<p>Assignment to each individual one out of four different phenotypic categories:</p> <p>typical FSHD:</p> <p>category A, subcategories A1, A2, A3.</p> <p>incomplete phenotypes:</p> <p>category B, subcategories B1, B2.</p> <p>asymptomatic/healthy subjects:</p> <p>category C, subcategories C1, C2</p> <p>atypical/complex phenotypes:</p> <p>category D, subcategories D1, D2</p>

Table 2

List of the major Literature Reports Based on the use of MOMIS and OPEN CLINICA tools for the Italian FSHD Registry

REFs	MOMIS FSHD WEB PLATFORM USAGE	PARTICIPANTS (n)	STUDY DESIGN	CONCLUSIONS
<p>Nikolic et al., 2016 "Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1–3 D4Z4 reduced alleles: experience of the FSHD Italian National Registry" BMJ.^[13]</p>	<ul style="list-style-type: none"> • Subjects filtered by year: collected between 2008 and 2013. • Subjects filtered by allele: 1 to 3 D4Z4 units • Subjects filtered by term: proband and relatives • Subjects classified using CCEF annotations: category and FSHD score 	<p>66 probands 33 relatives</p>	<p>Subjects were analyzed searching for signs of perinatal onset and evaluated for disease outcome through the systematic collection of clinical and anamnestic records of de novo and familial index cases and their relatives, carrying 1–3 D4Z4 units. To investigate the earliest signs of disease, the Infantile Anamnestic Questionnaire (IAQ) was used.</p>	<p>The size of the D4Z4 allele is not always predictive of severe clinical outcome. The high degree of clinical variability suggests that additional factors contribute to the phenotype complexity.</p>

REFs	MOMIS FSHD WEB PLATFORM USAGE	PARTICIPANTS (n)	STUDY DESIGN	CONCLUSIONS
<p>Ricci et al., 2016 "A novel clinical tool to classify facioscapulohumeral muscular dystrophy phenotypes" J Neurol.[15]</p>	<ul style="list-style-type: none"> • Subjects filtered by year: collected between 2008 and 2009 • Subjects filtered by allele: 1 to 9 D4Z4 units • Subjects filtered by geographic location: Modena, Turin, and Naples • Subjects classified using CCEF annotations: category and FSHD score 	<p>56 cases</p>	<p>Based on the 7-year experience of the Italian Clinical Network for FSHD, the inter-rater reproducibility of the CCEF was assessed between two examiners using kappa statistics by evaluating 56 subjects carrying the molecular marker used for FSHD diagnosis.</p>	<p>CCEF is an easy clinical tool useful to capture various phenotypes from classic FSHD to individuals with incomplete phenotype, or asymptomatic carriers as well as subjects with atypical signs for which alternative diagnoses may be supposed.</p>

REFs	MOMIS FSHD WEB PLATFORM USAGE	PARTICIPANTS (n)	STUDY DESIGN	CONCLUSIONS
<p>Ruggiero et al., 2020 "Phenotypic Variability Among Patients With D4Z4 Reduced Allele Facioscapulohumeral Muscular Dystrophy Epigenetics" JAMA Net Open.[20]</p>	<ul style="list-style-type: none"> • Subjects filtered by year: collected between 2008 and 2016. • Subjects filtered by allele: 7 to 8 D4Z4 units • Subjects filtered by term proband and relatives • Subjects classified using CCEF annotations: category and FSHD score 	<p>187 probands 235 relatives</p>	<p>Subjects carrying 7 to 8 D4Z4 units and their relatives were phenotypically stratified based on CCEF categories.</p>	<p>A large phenotypic variability associated with individuals carrying a DRA with 7 to 8 RUs, was found in contrast to the indication that a positive molecular test is the only determining aspect for FSHD diagnosis. Carriers of a DRA with 7 to 8 RUs constitute a genetic subgroup different from classic FSHD.</p> <p>The use of CCEF and the family study are both required for clinical management and genetic counseling.</p>

REFs	MOMIS FSHD WEB PLATFORM USAGE	PARTICIPANTS (n)	STUDY DESIGN	CONCLUSIONS
<p>Ricci et al., 2020 “Large genotype–phenotype study in carriers of D4Z4 borderline alleles provides guidance for facioscapulohumeral muscular dystrophy diagnosis. Sci Rep . [21]</p>	<p>Subjects filtered by year: collected between 2008 and 2016.</p> <p>Subjects filtered by allele: 9 to 10 D4Z4 units</p> <p>Subjects filtered by term proband and relatives</p> <p>Subjects classified using CCEF annotations: category and FSHD score</p>	<p>134 probands</p> <p>110 relatives</p>	<p>Subjects carrying 9 to 10 D4Z4 units and their relatives were phenotypically stratified based on CCEF categories</p>	<p>A large phenotypic variability associated with individuals carrying a DRA with 9 to 10 RUs. D4Z4 alleles with 9–10 repeat has been found in healthy individuals, in subjects with FSHD or affected by other myopathies.</p> <p>The use of CCEF and the family study are both required for clinical management and genetic counseling. Stratification of patients using the number of D4Z4 repeat units is not accurate.</p>
<p>Vercelli et al., 2020 “A 5-year clinical follow-up study from the Italian National Registry for FSHD” J Neurol. [22]</p>	<ul style="list-style-type: none"> • Subjects filtered by term proband and relatives • Subjects classified using CCEF annotations: Category and FSHD score • Disease progression measured as ΔFSHD score 	<p>141 probands</p> <p>105 relatives</p>	<p>Subjects were analyzed to estimate the disease worsening calculated as the FSHD score performed at baseline and at the end of 5-year follow-up (ΔFSHD score)</p>	<p>The progression of disease is different between index cases and carrier relatives and the assessment of the CCEF categories has strong prognostic effect in FSHD1 patients.</p>

REFs	MOMIS FSHD WEB PLATFORM USAGE	PARTICIPANTS (n)	STUDY DESIGN	CONCLUSIONS
<p>Nikolic et al., 2020 "Interpretation of the Epigenetic Signature of Facioscapulohumeral Muscular Dystrophy in Light of Genotype-Phenotype Studies" Int J Mol Sci.[23]</p>	<ul style="list-style-type: none"> • Subject filtered by clinical category: A1-A3, FSHD score > 1 • Subjects filtered by term proband and relatives • Subjects (relatives) classified using CCEF annotations: category and FSHD score 	<p>122 probands 110 relatives</p>	<p>The D4Z4 methylation level at 4q35 was assessed in 122 FSHD1 index cases with FSHD clinical score ≥ 1 and in relatives carrying the same molecular defect and presenting classical FSHD (category A), or incomplete /complex phenotype (categories B and D)) or no muscle impairment (category C).</p>	<p>The D4Z4 methylation levels among index cases and in family study show a high variability revealing no association with clinical manifestation or disease severity and no predictive value for disease progression.</p> <p>The results of D4Z4 methylation analysis must be cautiously interpreted in respect to disease prognosis which requires family studies.</p>
<p>Rodolico et al., 2020 "Deletion of the Williams Beuren syndrome critical region unmasks facioscapulohumeral muscular dystrophy" Eur J Paediatr Neurol. [24]</p>	<ul style="list-style-type: none"> • Subjects filtered by term proband • Subjects filtered for non-familiar cases • Subjects filtered by comorbidities/complex phenotypes 	<p>1339 FSHD unrelated cases</p>	<p>Among 1339 unrelated FSHD cases three unrelated cases were found who presented signs of Williams-Beuren Syndrome (WBS) in early childhood and later developed FSHD. All three cases carry the molecular defects associated with the two disorders.</p>	<p>The rarity of WBS and FSHD, 1 in 7500 and 1 in 20,000 respectively, argued for a nonrandom association of the two diseases. These cases open novel and unexpected interpretation of genetic findings providing hints for the identification of genes and functional pathways involved.</p>

Table 2: Systematic analysis and comparison of the major literature reports based on the MOMIS FSHD Web Platform application.

Discussion

The expanding use of DNA molecular analysis applied to diseases with a genetic basis has revealed increasingly varied phenotypic spectra in patients carrying mutations in the same gene, as one can see by the number of reviews[3,25,26].

It is now clear that the well-phenotyped and genetically characterized patient cohorts can promote a more accurate treatment of patients. These steps are necessary to develop appropriate outcome measures and biomarkers and to proceed towards trial readiness. This requires both the integration of data available in different locations and the management of continuously increasing information related to a single sample. Now, the greatest effort is directed towards combining genomic, anamnestic and clinical investigation; a process that requires appropriate Big Data techniques. In the case of FSHD the MOMIS FSHD Web Platform serve these purposes. The collection of participants evaluated with a standardized methodology with demonstrated interrater reliability is the major strength of the INRF and the platform implemented for INRF data management based on MOMIS and OpenClinica is a solid foundation both for the secure users access to individual's data and for the clinical management of patients. This structure fosters the definition of epidemiology of FSHD as well as definition of its natural history, two elements that are crucial for the design of therapeutic interventions.

A mid-term goal of the INRF is the identification of factors, including comorbidities or lifestyle habits that might influence the disease onset and progression. These might be identified through epidemiology studies based on the analysis of anamnestic records of a vast number of subjects and pedigrees. Moreover, the collection of precisely phenotyped participants might permit their stratification on the basis of standardized criteria, providing the framework for further studies that might serve clinical practice and basic research in a rare, understudied disease.

The major weakness is that FSHD registry is an evolving tool that generates data requiring registry constantly updates and users' training in order to fully exploit its function, otherwise it is in danger for becoming a statical survey of a fixed cohort, which is not its main purpose.

To this purpose, it would be of essential help in the future to look for an implement in our database which could facilitate the fill in of subject's clinical records both for clinicians and patients themselves.

Conclusion

The future of medicine lies on precision medicine, in which genetic counseling and treatments are tailored to each individual integrating genomics to clinical, environmental, and lifestyle data. The systematic collection of clinical data for rare genetic diseases on a National scale, providing informative observational data that can support clinicians, researchers, and the health system.

We propose our experience with the integrated management of data in a rare disease such as FSHD as a white label that could be applied to other complex diseases in which the genotype/phenotype correlation is far to be resolved.

List Of Abbreviation

FSHD, Facioscapulohumeral Dystrophy; DRA, D4Z4 Reduced Allele; PAS, Polyadenilation Signal; CCEF, Comprehensive Clinical Evaluation Form; INRF Italian National Registry for FSHD; ICNF, Italian Clinical Network for FSHD; MOMIS: Mediator Environment for Multiple Information Sources; DB, Database; EDC, Electronic Data Capture.

Declarations

ACKNOWLEDGEMENTS: We would like to thank all of the patients, their families and staff that participated in collecting samples at clinical specialized centres, particularly we thank all **the members of Italian Clinical Network for FSHD (ICNF) Group:** Giulia Ricci, Gabriele Siciliano (Department of Clinical and Experimental Medicine, University of Pisa); Lucia Ruggiero, Lucio Santoro (Department of Neurosciences, Reproductive, and Odontostomatological Sciences, University Federico II, Naples); Silvia Bonanno, Lorenzo Maggi (Neuroimmunology and Neuromuscular Disease Unit, Foundation IRCCS Carlo Besta Neurological Institute, Milano); Liliana Vercelli, Tiziana Mongini (Neurological Clinic, Center for Neuromuscular Diseases, Department of Neurosciences, University of Turin); Stefano Previtali, Marina Scarlato (Neuromuscular Repair Unit, Inspe and Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan); Elena Pegoraro (Department of Neurosciences, University of Padua); Grazia D'Angelo (Scientific Institute IRCCS Institute Medea, Neuromuscular unit, Neurohabilitation Dept, Bosisio Parini, Turin); Antonio Di Muzio, Angela Berardinelli (Child Neurology and Psychiatry Unit, IRCCS, Casimiro Mondino Foundation, Pavia); Massimiliano Filosto (Neurology Clinic, Spedali Civili Hospital, Brescia); Carmelo Rodolico (Department of Clinical and Experimental Medicine, University of Messina); Maurizio Moggio, (Neuromuscular Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Dino Ferrari Center, University of Milan); Giovanni Antonini, Elisabetta Bucci (Department of Neuroscience, Mental Health, and Sensory Organs, S. Andrea Hospital, University of Rome Sapienza); Maria Antonietta Maioli (Department of Public Health, Molecular and Cincial Medicine, University of Cagliari); Corrado Angelini (IRCCS San Camillo, Venezia); Maria Lucia Valentino (Unit of Neurology, Department of Biomedical and NeuroMotor Sciences (DIBINEM), University of Bologna); Roberto Massa (Unit Neuromuscular diseases, System medicine Dept, University of Roma Tor Vergata, Rome).

ETHICS APPROVAL AND CONSENT TO PARTICIPATE: Not applicable

CONSENT FOR PUBLICATION: Not applicable

AVAILABILITY OF DATA AND MATERIAL: Please contact author for data requests. To date, the INRF is not an open access database. The stored data have been available for the Italian Clinical Network for FSHD (ICNF) since its establishment. The Miogen Lab welcomes any researcher focused on FSHD or different

genetic diseases to make contact for collaborative project proposals. Applications for Access to the INRF data may be directed (providing the corresponding documentation required) to www.FSHD.it (managed by Miogen Lab at University of Modena and Reggio Emilia) and should contain a brief description of the related research project.

FUNDING: Regione Emilia Romagna Progetti - Alte Competenze per la Ricerca e il Trasferimento Tecnologico. Regione Emilia Romagna Piano triennale alte competenze alla ricerca POR/FSE 2014/2020

AUTHOR CONTRIBUTIONS: CB, VS and RT contributed to study design, molecular analysis, data collection, data analysis and interpretation, literature search, preparing of figures/tables and manuscript writing. EC, SB, MO, LG, LM contributed in data collection and integration, manuscript editing.

COMPETING INTERESTS: The authors declare that they have no competing interests

References

1. Mostacciuolo ML, Pastorello E, Vazza G, Miorin M, Angelini C, Tomelleri G, et al. Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. *Clin Genet* [Internet]. 2009 [cited 2019 Sep 6];75:550–5. Available from: <http://doi.wiley.com/10.1111/j.1399-0004.2009.01158.x>
2. Padberg GWAM, LUMC. Facioscapulohumeral disease. Faculty of Medicine, Leiden University Medical Center (LUMC), Leiden University; [cited 2019 Oct 7]; Available from: <https://openaccess.leidenuniv.nl/handle/1887/25818>
3. Ricci G, Scionti I, Sera F, Govi M, D'Amico R, Frambolli I, et al. Large scale genotype–phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. *Brain* [Internet]. 2013 [cited 2019 Sep 18];136:3408–17. Available from: <https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awt226>
4. Padberg GW, Lunt PW, Koch M, Fardeau M. Diagnostic criteria for facioscapulohumeral muscular dystrophy. *Neuromuscul Disord* [Internet]. Elsevier; 1991 [cited 2019 Sep 18];1:231–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1822799>
5. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* [Internet]. 1992 [cited 2019 Sep 26];2:26–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1363881>
6. NAKAGAWA M, MATSUZAKI T, HIGUCHI I, FUKUNAGA H, INUI T, NAGAMITSU S, et al. Facioscapulohumeral Muscular Dystrophy: Clinical Diversity and Genetic Abnormalities in Japanese Patients. *Intern Med* [Internet]. 1997 [cited 2019 Sep 18];36:333–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9213170>
7. Goto K, Nishino I, Hayashi YK. Very low penetrance in 85 Japanese families with facioscapulohumeral muscular dystrophy 1A. *J Med Genet* [Internet]. 2004 [cited 2019 Sep 18];41:12e – 12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14729852>

8. Tonini MMO, Passos-Bueno MR, Cerqueira A, Matioli SR, Pavanello R, Zatz M. Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). *Neuromuscul Disord* [Internet]. 2004 [cited 2019 Sep 18];14:33–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14659410>
9. de Greef JC, Lemmers RJLF, van Engelen BGM, Sacconi S, Venance SL, Frants RR, et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum Mutat* [Internet]. John Wiley & Sons, Ltd; 2009 [cited 2019 Oct 7];30:1449–59. Available from: <http://doi.wiley.com/10.1002/humu.21091>
10. Scionti I, Fabbri G, Fiorillo C, Ricci G, Greco F, D'Amico R, et al. Facioscapulohumeral muscular dystrophy: New insights from compound heterozygotes and implication for prenatal genetic counselling. *J Med Genet* [Internet]. 2012 [cited 2020 Feb 14];49:171–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22217918>
11. Sakellariou P, Kekou K, Fryssira H, Sofocleous C, Manta P, Panousopoulou A, et al. Mutation spectrum and phenotypic manifestation in FSHD Greek patients. *Neuromuscul Disord* [Internet]. 2012 [cited 2019 Sep 18];22:339–49. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22357364>
12. Ricci G, Zatz M, Tupler R. Facioscapulohumeral Muscular Dystrophy: More Complex than it Appears. *Curr Mol Med* [Internet]. 2014 [cited 2019 Sep 6];14:1052–68. Available from: <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1566-5240&volume=14&issue=8&spage=1052>
13. Nikolic A, Ricci G, Sera F, Bucci E, Govi M, Mele F, et al. Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1-3 D4Z4 reduced alleles: Experience of the FSHD Italian National Registry. *BMJ Open*. 2016;6:1–10.
14. Lamperti C, Fabbri G, Vercelli L, D'Amico R, Frusciante R, Bonifazi E, et al. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: The FSHD clinical score. *Muscle Nerve* [Internet]. 2010 [cited 2019 Sep 18];42:213–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20544930>
15. Ricci G, Ruggiero L, Vercelli L, Sera F, Nikolic A, Govi M, et al. A novel clinical tool to classify facioscapulohumeral muscular dystrophy phenotypes. *J Neurol* [Internet]. 2016 [cited 2019 Sep 18];263:1204–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27126453>
16. Bergamaschi S, Bouquet P, Giazomuzzi D, Guerra F, Po L, Vincini M. An Incremental Method for the Lexical Annotation of Domain Ontologies. *Int J Semant Web Inf Syst* [Internet]. IGI Global; 2007 [cited 2021 Feb 10];3:57–80. Available from: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jswis.2007070103>
17. Magnotta L, Gagliardelli L, Simonini G, Orsini M, Bergamaschi S. MOMIS Dashboard: A Powerful Data Analytics Tool for Industry 4.0. 2018 [cited 2021 Feb 10]; Available from: www.datariver.it/data-integration/momis
18. Beneventano D, Bergamaschi S, Gagliardelli L, Po L. Open Data for Improving Youth Policies [Internet]. *pdfs.semanticscholar.org*. [cited 2021 Feb 10]. Available from: www.opendatabarometer.org

19. Scionti I, Greco F, Ricci G, Govi M, Arashiro P, Vercelli L, et al. Large-scale population analysis challenges the current criteria for the molecular diagnosis of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* [Internet]. The American Society of Human Genetics; 2012;90:628–35. Available from: <http://dx.doi.org/10.1016/j.ajhg.2012.02.019>
20. Ruggiero L, Mele F, Manganelli F, Bruzzese D, Ricci G, Vercelli L, et al. Phenotypic Variability Among Patients With D4Z4 Reduced Allele Facioscapulohumeral Muscular Dystrophy. *JAMA Netw open* [Internet]. NLM (Medline); 2020 [cited 2021 Feb 10];3:e204040. Available from: <https://jamanetwork.com/>
21. Ricci G, Mele F, Govi M, Ruggiero L, Sera F, Vercelli L, et al. Large genotype–phenotype study in carriers of D4Z4 borderline alleles provides guidance for facioscapulohumeral muscular dystrophy diagnosis. *Sci Rep* [Internet]. Nature Research; 2020 [cited 2021 Feb 22];10:1–12. Available from: <https://doi.org/10.1038/s41598-020-78578-7>
22. Vercelli L, Mele F, Ruggiero L, Sera F, Tripodi S, Ricci G, et al. A 5-year clinical follow-up study from the Italian National Registry for FSHD. *J Neurol* [Internet]. Springer Science and Business Media Deutschland GmbH; 2021 [cited 2021 Feb 10];268:356–66. Available from: <https://doi.org/10.1007/s00415-020-10144-7>
23. Nikolic A, Jones TI, Govi M, Mele F, Maranda L, Sera F, et al. Interpretation of the Epigenetic Signature of Facioscapulohumeral Muscular Dystrophy in Light of Genotype-Phenotype Studies. *Int J Mol Sci* [Internet]. *Int J Mol Sci*; 2020 [cited 2020 Apr 16];21:2635. Available from: <https://www.mdpi.com/1422-0067/21/7/2635>
24. Rodolico C, Politano L, Portaro S, Murru S, Boccone L, Sera F, et al. Deletion of the Williams Beuren syndrome critical region unmasks facioscapulohumeral muscular dystrophy. *Eur J Paediatr Neurol* [Internet]. W.B. Saunders Ltd; 2020 [cited 2021 Feb 10];27:25–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/32553920/>
25. Krasnianski M, Eger K, Neudecker S, Jakubiczka S, Zierz S. Atypical Phenotypes in Patients With Facioscapulohumeral Muscular Dystrophy 4q35 Deletion. *Arch Neurol* [Internet]. 2003 [cited 2019 Sep 18];60:1421. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14568813>
26. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease [Internet]. *Hum. Genet. Hum Genet*; 2013 [cited 2021 Feb 22]. p. 1077–130. Available from: <https://pubmed.ncbi.nlm.nih.gov/23820649/>

Figures

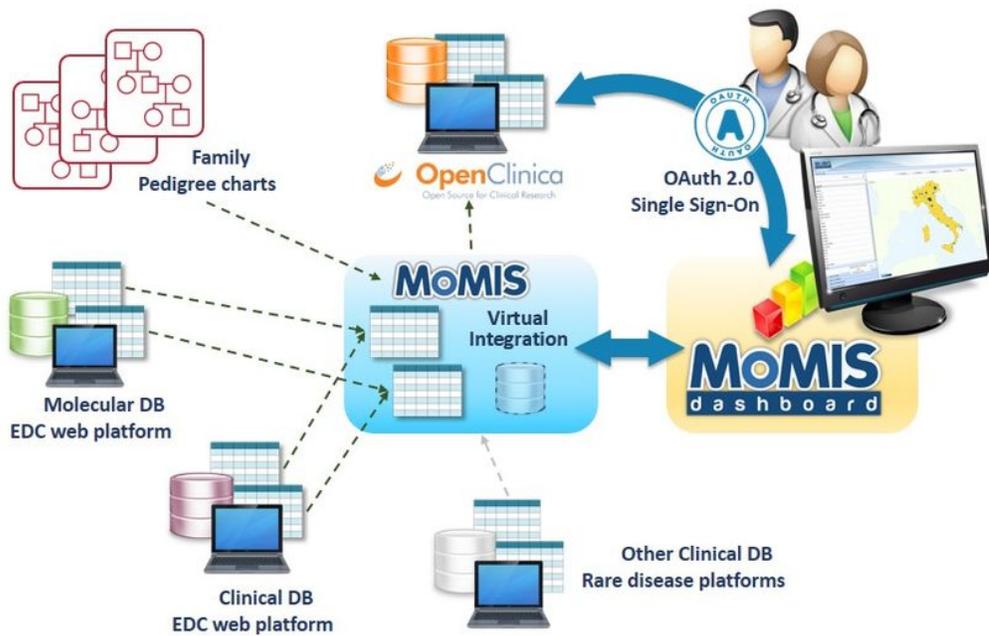


Figure 1

Figure 1

Architecture of the INRF based on the MOMIS FSHD Web Platform.

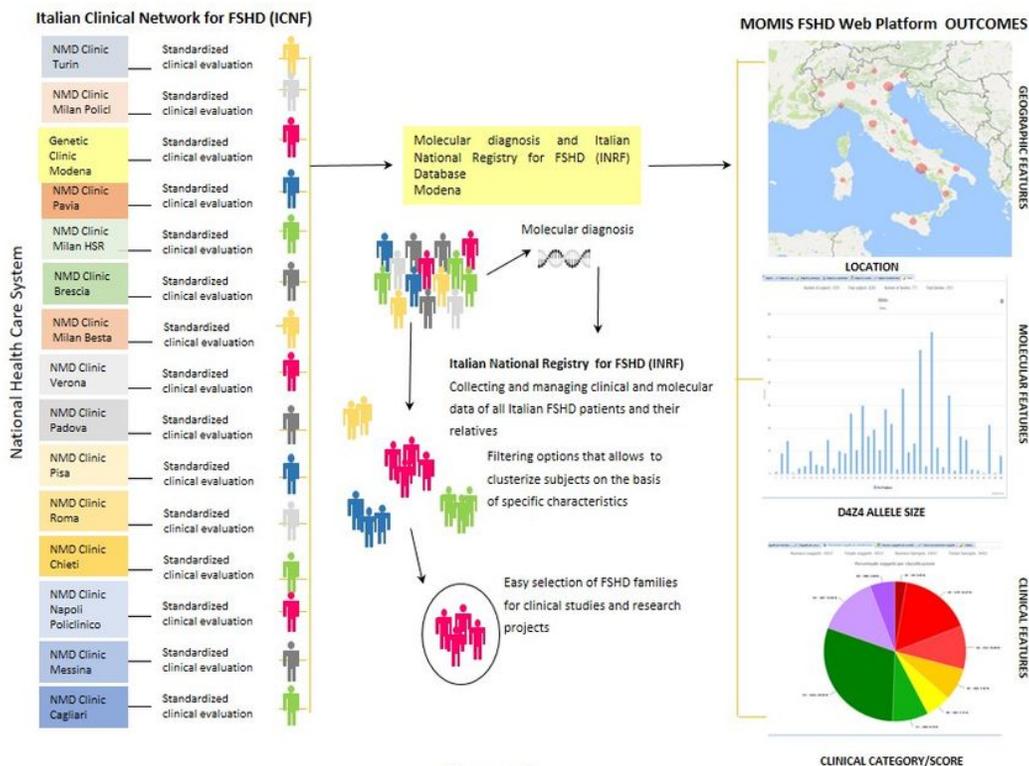


Figure 2

Figure 2

The CNF-INRF-MOMIS FSHD Web Platform interaction network. Subjects are recruited by the ICNF, a consortium consisting of 14 centers specialized in neuromuscular diseases located across Italy. Biological samples of participants and their clinical information are collected in Modena for molecular analysis and database integration in INRF. MOMIS FSHD Web Platform filtering options allow patients selection for clinical and research purposes. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

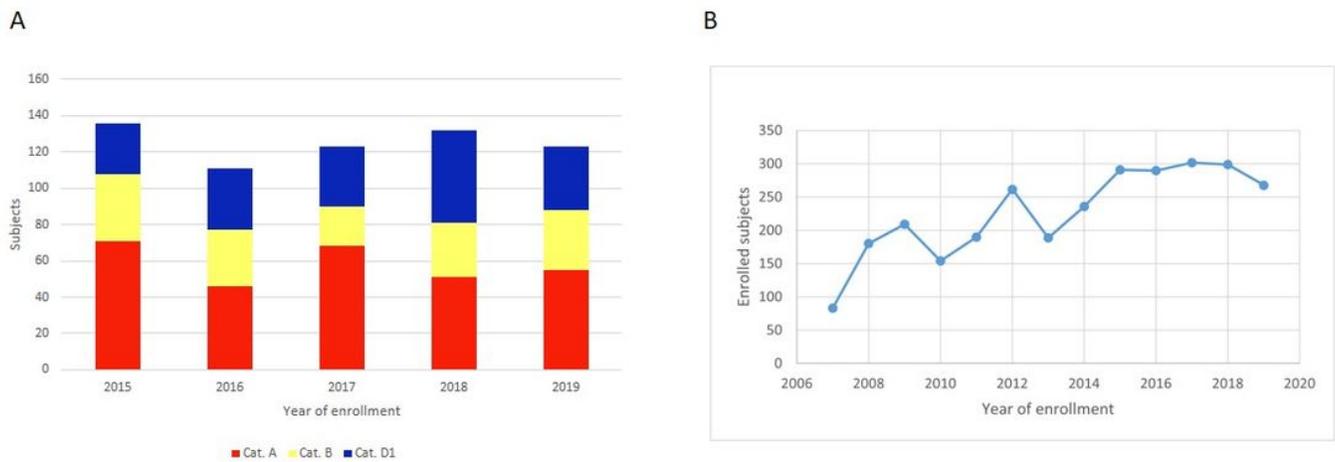


Figure 3

Figure 3

Trends in enrolment of patients collected by INRF. (A) Distribution of clinical categories as described by the CCEF within subjects with a myopathic FSHD-like phenotype since the start of the CCEF utilization (2015-2019) (B) Graphic representation of the INRF enrolment in the period 2007 to 2019. The annual number of subjects is reported.