

RESEARCH

Open Access



# 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<sup>1</sup>, Valentina Salsi<sup>2</sup>, Mirko Orsini<sup>3</sup>, Enrico Calanchi<sup>3</sup>, Luca Magnotta<sup>3</sup>, Luca Gagliardelli<sup>4</sup>, June Kinoshita<sup>5</sup>, Sonia Bergamaschi<sup>4</sup> and Rossella Tupler<sup>1,6\*</sup> 

## 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 *Mediator Environment for Multiple Information Sources* (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 collected the molecular data regarding FSHD diagnosis conducted on 7197 subjects and identified 3362 individuals carrying a *D4Z4 Reduced Allele* (DRA): 1634 were unrelated index 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 the period 2009–2020, 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 may have great value in allowing us to redefine diagnostic criteria and disease markers for FSHD. By extending the use of the MOMIS data integration framework to other

\*Correspondence: rossella.tupler@unimore.it

<sup>1</sup> Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy  
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

countries and the longitudinal systematic collection of standardized clinical data will facilitate the understanding of disease natural history and offer valuable inputs towards trial readiness. This approach is of high significance to FSHD medical community and also to rare disease research in general.

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

## 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, 6–14] reinforced the idea that additional elements take part in disease onset and progression. Indeed, 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, the observed phenotypic heterogeneity 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 [15]), and the Comprehensive Clinical Evaluation Form (CCEF) [16], 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 unrelated families with 1728 relatives heterozygotes for a DRA. This large collection of data, the largest worldwide, 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.

Here we describe the organization and summarize the results obtained from the implementation of the *MOMIS FSHD Web Platform*.

## Methods

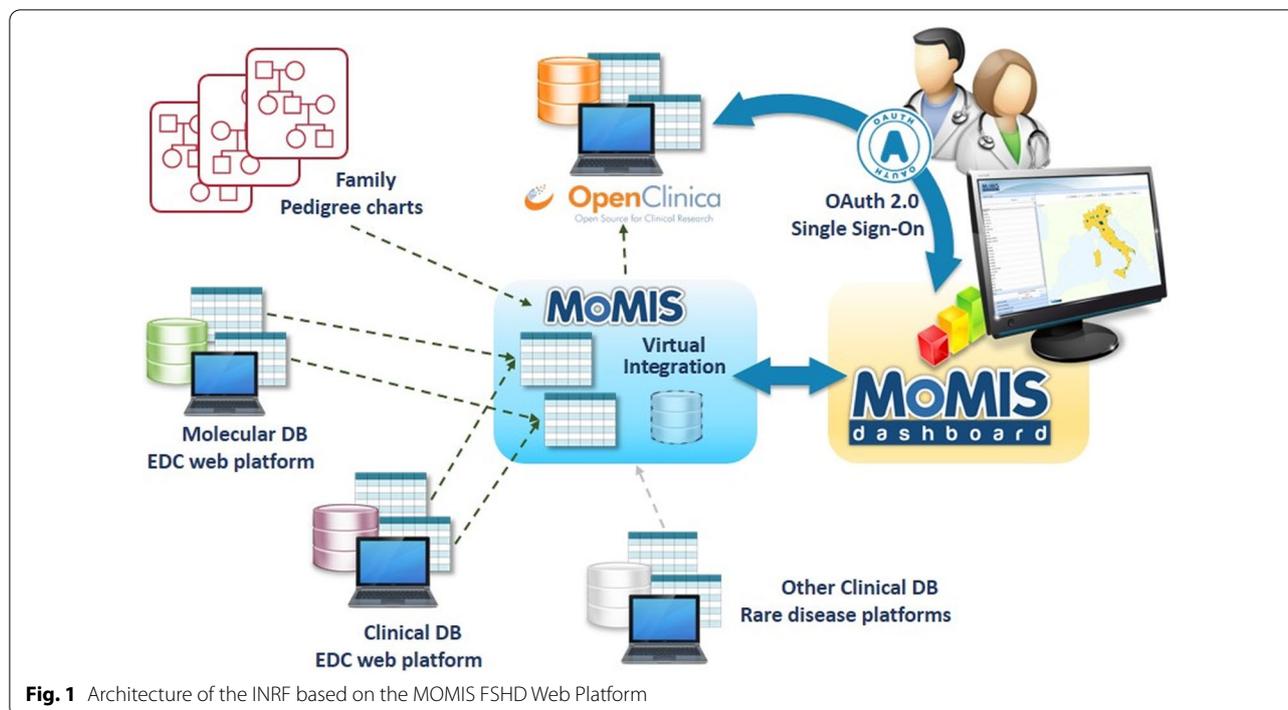
### MOMIS framework architecture

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 [17], including clinical trials and public authorities data management [18]. 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 [17, 18]: 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 [19]. 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

**Table 1** Description of the main features in the 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> Investigates the subject's clinical history</p> <p><i>Part B</i> Evaluates the patient's disability (neurological examination)</p> <p><i>Part C</i> Assesses muscle segmental involvement using the <i>Medical Research Council (MRC)</i> 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> <ul style="list-style-type: none"> <li>1 <i>facial muscles</i> (scored from 0 to 2)</li> <li>2 <i>scapular girdle muscles</i> (scored from 0 to 3)</li> <li>3 <i>upper limb muscles</i> (scored from 0 to 2)</li> <li><i>leg muscles</i> (scored from 0 to 2)</li> <li>4 <i>pelvic girdle muscles</i> (scored from 0 to 5)</li> <li>5 <i>abdominal muscles</i> (scored from 0 to 1)</li> </ul> <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><i>typical FSHD</i>:</p> <ul style="list-style-type: none"> <li>category A, subcategories A1, A2, A3</li> </ul> <p><i>incomplete phenotypes</i>:</p> <ul style="list-style-type: none"> <li>category B, subcategories B1, B2</li> </ul> <p><i>asymptomatic/healthy subjects</i>:</p> <ul style="list-style-type: none"> <li>category C, subcategories C1, C2</li> </ul> <p><i>atypical/complex phenotypes</i>:</p> <ul style="list-style-type: none"> <li>category D, subcategories D1, D2</li> </ul>

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 (i.e., 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 [15], to the *Comprehensive Clinical Evaluation Form* (CCEF) [16], 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 [16]. 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 section 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 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 clinical data coming from 3577 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:

- 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 of muscle impairment), within a configurable range.
- 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].
- Family:** filters by family name.
- Registration Year:** the year of patient registration in the database can be selected.
- 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:

- The number of subjects registered every year as well as the enrollment trend is a metric that is useful for planning resources and procurement.
- Subjects enrolled by location, showing on a map the geographic distribution of the patients.
- The distribution of alleles within subjects accrued by the INRE, 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.

#### 4. Principle of data governance

The INRF complies with practices indicated for the good governance of a registry. The team has different and complementary expertise, including a named registry lead, a statistician, data stewardship and database management, development and hosting and a legal specialist. Each clinical center, headed by a designated candidate, has the responsibility to train their medical staff about the clinical procedures adopted by the registry. Once per year, a round table discussion is planned in order to chew over the concerns of the group and define key objectives.

With regard to privacy policies, INRF adheres to “EU General Data Protection Regulation (GDPR)”.

#### 5. Data access

Data management is performed by DataRiver srl, which exercises the utmost care to ensure that all the confidential information stays protected from prying access.

Data is collected according to the principles of lawfulness, fairness and transparency, carefully applying specific Standard Operation Procedures (SOP) in order to prevent and detect any exposure of personal and sensitive data and adopt proper risk mitigation measures.

DataRiver follows all the best practices and the laws to ensure the protection of the personal data of the patients, all the system is compliant with EU General Data Protection Regulation (GDPR).

Data is stored on a dedicated server hosted by a UNI EN ISO 27001:2005 certified server farm to ensure the physical, logical and organizational security of information. The server is protected by firewall configurations and remote access is allowed only to the authorized system administrator through SSH connection with authentication via asymmetric key pair.

Sensitive data are stored in a database implementing encryption at the filesystem level, in compliance with the requirements of the GDPR (art.32) and with the published guidelines by the European Union Agency for Network and Information Security (ENISA—Manual on the Security of Personal Data Processing).

Data that necessary allow the identification of the subjects for research and patient’s care purposes are stored in a separate database and made available to authorized users on a role-based permission system implemented by the registry platform in order to guarantee the minimization of the access to personal data.

The access to patients’ personal information is authorized exclusively for staff working at Miogen lab, with the consensus of the curator of the Registry. Every member of the staff has personal credentials for accessing Momis platform and in agreement with the curator of the

Registry have the permission to manage the data (insertion, corrections, extraction, quality control). The agreement between the curator of the Registry and the staff is to maintain the confidentiality of all the information collected. When the employment relationship is over, the access for the member of the staff is immediately interrupted. If requested, anonymized data are provided to collaborators for cohort studies.

#### 6. Data ownership

The principal investigator of the INFR is the curator of the data and acts as the “data controller” of the registries and assumes all the responsibility for the protection of the data, its storage, use and access. When processed, the data become research data and are then the intellectual property of the investigator who is the “third party”.

#### 7. Data management

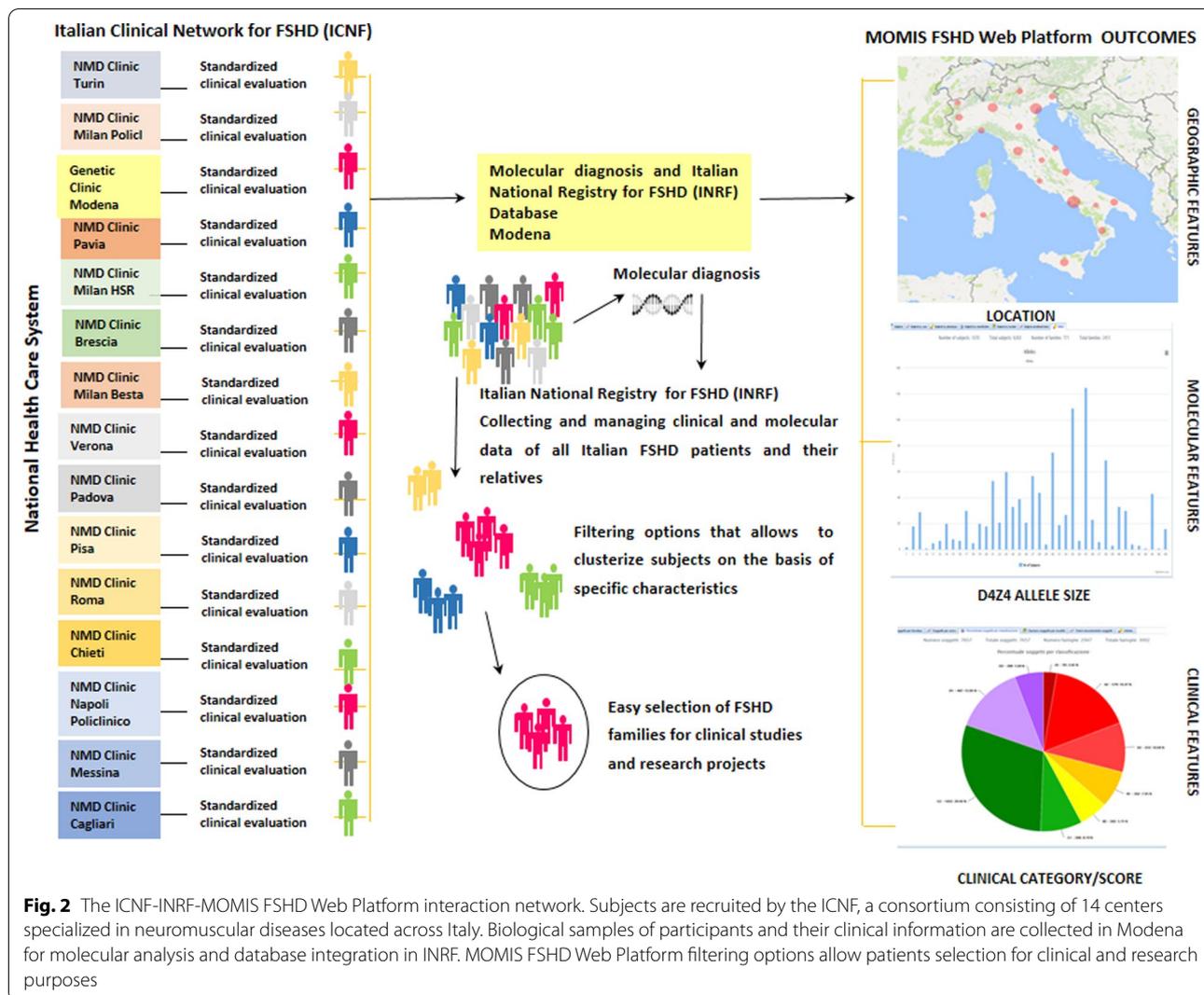
DataRiver provides remote help-desk service and support to the users of the platform on weekdays during working hours, as established by contract. The helpdesk and support service includes the creation and management of user accounts, the solution of problems related to the ordinary platform usage and maintenance, the analysis, application and documentation of any extraordinary interventions to address issues related to the process of entry, edit and management of the registry’s clinical data. The INRF is operating to reach high-quality data level, and interoperability based on FAIR principles (i.e., Findable, Accessible, Interoperable and Reusable) in order to make data available for wider use.

## 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 ICNE, 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 probands unrelated to each other, of these 602 are isolated cases (with no other family members investigated). In 1032 cases, molecular investigation has been extended to other family members. Of 3747 molecularly tested relatives, 1728 were heterozygotes for a DRA.

Thanks to the analysis of 3577 clinical forms collected since 2009, it has been possible to establish 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 led 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. Since the establishment of the INRF in 2007, the number of enrolled subjects has been steadily increasing until 2015, the time the CCEF has been applied for the clinical evaluation (Additional file 1A). From 2015, the enrollment of subjects with a myopathic FSHD-like phenotype (categories A, B, and D) has been steady (Additional File 1B). 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 [20]. Assuming 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 usages 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 the 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: 1) the CCEF validation [16] 2) the genotype/phenotype characterization of people carrying specific molecular features: 1–3 DRA [14] or 7–8 DRA [21] or 9–10 [22] 3) follow up studies [23]. The MOMIS FSHD Web Platform has also been used 4) for molecular research on D4Z4 epigenetic regulation [24] or 5) to identify genetic interactions with other rare diseases [25].

#### INRF in the scheme of the world

Several entities for patients’ data collection have been established in the last years including national, European and global Rare Diseases (RDs) registries responding to the rising requests from patient organizations, regulatory agencies and healthcare professionals for curated data serving epidemiological and clinical research on RDs.

At the European level, several platforms have been created: the European Platform for rare disease registries (EPIRARE: [www.epirare.eu](http://www.epirare.eu)), the International Rare Disease Research Consortium, (IRDIRC: [www.irdirc.org](http://www.irdirc.org)) and the Integrated Platform connecting databases, registries, biobanks and clinical bioinformatics for Rare Disease research (RD-Connect: [www.rd-connect.eu](http://www.rd-connect.eu)). In February 2017, 24 European Reference Networks (ERNs) (<https://www.ern-rnd.eu>) were established in a European legal framework, of which 23 are dedicated to rare or low prevalence complex diseases or conditions. An Italian national institutional registry of rare disease patients,

the *Registro Nazionale Malattie Rare* (RNMR), has been established in Italy in 2001 by the National Centre for Rare Diseases (*Centro Nazionale Malattie Rare*, CNMR) of the *Istituto Superiore di Sanità* (National Institute of Health) [26]. The registry is not exhaustive of all RDs and FSHD is underrepresented as well as other Neuromuscular disorders (NMD) as this database comprises RDs characterized by congenital syndromes, severe clinical conditions, high level of invalidity, and costly treatments. Here, diseases of the musculoskeletal system and connective tissue are globally the 5.5% of total records.

Neuromuscular diseases specific registries are present worldwide and globally described within the TREAT-NMD neuromuscular network website (<https://treat-nmd.org/patient-registries>). The TREAT-NMD network has developed a specific tool, the TREAT-NMD Global Registry Network, where registries are listed regarding the principal NMD disorders worldwide, including FSHD. Recently, a collaboration between TREAT-NMD and ERNs was launched, leading to the establishment of a “FSHD European Trial Network”. The INRF is part of this network.

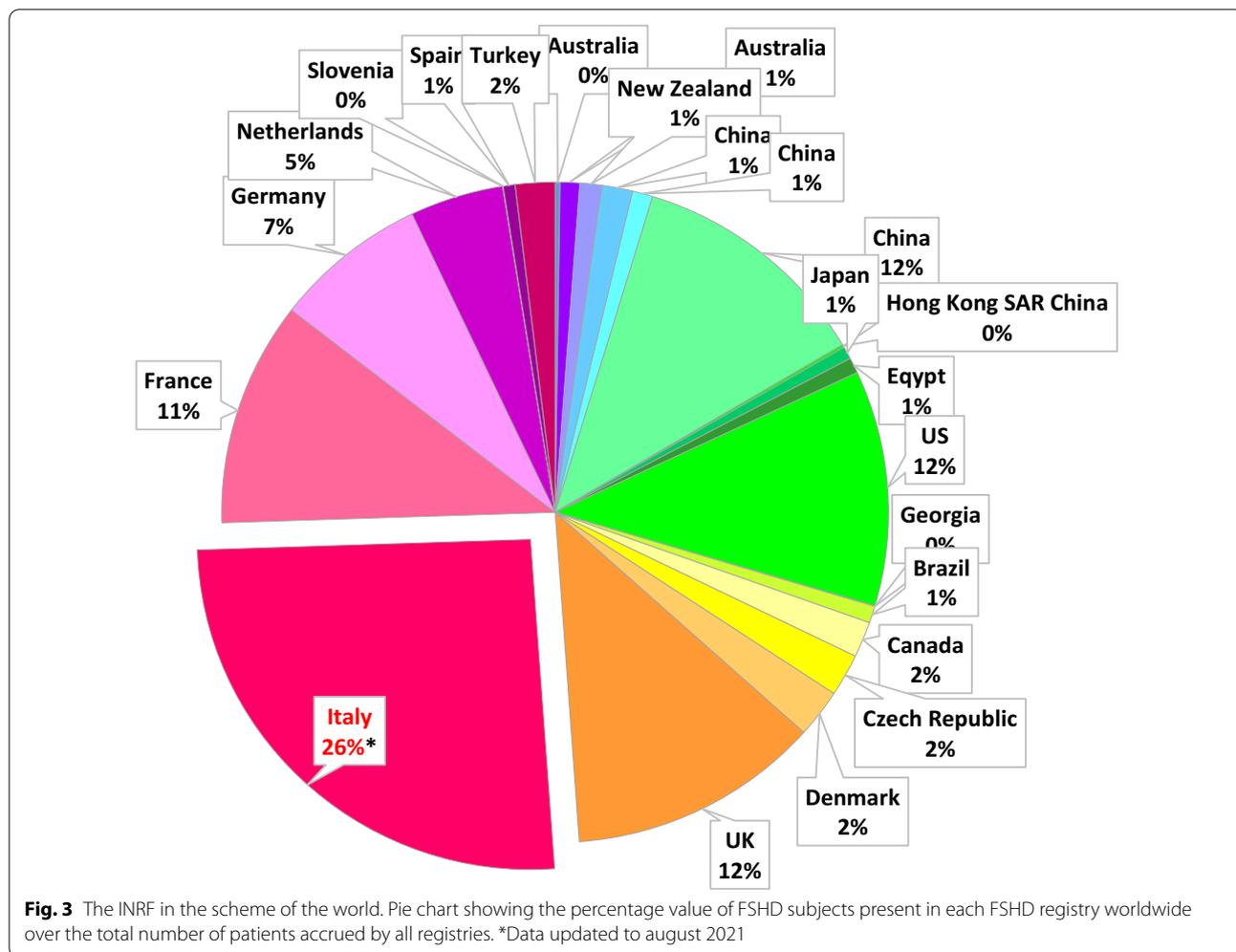
According to data provided by the FSH Society, we were able to compare a total of 23 FSHD registries located in 5 continents (4 in Asia, 1 in Africa, 4 in America, 11 in Europe and 3 in Oceania) [27–32] which were collectively analyzed based on patient’s number within and on specific features as reported in Fig. 3 and in Additional file 2. In particular, Fig. 3 shows the number of FSHD subjects present in each platform as the percentage value over the total number of patients accrued by all registries. Notably, the pie chart highlights that the INFR alone accounts for 26% of total FSHD patients worldwide. Given the reported values, we selected registries including a significant number of patients, at least 5% of total number, for further consideration. Additional file 2 shows demographic, clinical/molecular and logistics parameters. We also evaluated the ratio between the number of accrued FSHD subjects over the total country population. This parameter varies among countries and may be useful to estimate the ‘recruitment power’ of single registries. Additional file 2 also shows that the INFR platform is unique among the various FSHD registries, as clinical and anamnestic information are totally reported by the ICNF-clinicians through the CCEF template and all accrued subjects are molecularly investigated.

#### 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, 33, 34].

**Table 2** Literature reports based on the use of MOMIS and OPEN CLINICA tools for the Italian FSHD Registry

Momis FSHD platform usage	Participants (n)	Study design	Conclusions
Subjects filtered by year: 2008–2013 Subjects filtered by allele: 1–3 RUs Subjects filtered by term: proband and relatives Subjects classified using CCEF annotations: category and FSHD score	66 probands 33 relatives	Subjects carrying 1–3 DRA were searched for disease early onset. Disease outcome was evaluated in de novo and familial probands and their relatives. To investigate the earliest signs of disease, the Infantile Anamnestic Questionnaire (IAQ) was used. Nikolic et al., 2016	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
Subjects filtered by year: 2008–2009 Subjects filtered by allele: 1–9 RUs Subjects filtered by geographic location: Modena, Turin, and Naples Subjects classified using CCEF annotations: category and FSHD score	56 cases	Based on the 7-year experience of the ICNF, the inter-rater reproducibility of the CCEF was assessed between two examiners using kappa statistics in 56 subjects carrying the molecular marker used for FSHD diagnosis. Ricci et al., 2016	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
Subjects filtered by year: 2008–2016 Subjects filtered by allele: 7–8 RUs Subjects filtered by term: proband and relatives Subjects classified using CCEF annotations: category and FSHD score	187 probands 235 relatives	Probands and their relatives carrying 7–8 DRA were phenotypically stratified based on the CCEF categories. Ruggiero et al., 2020	Large phenotypic variability was found in individuals carrying a 7–8 DRA, in contrast to the indication that a positive molecular test is the only determining aspect for FSHD diagnosis. Carriers of a 7–8 DRA constitute a genetic subgroup different from classic FSHD. The use of CCEF and the family study are both required for clinical management and genetic counseling
Subjects filtered by year: 2008–2016 Subjects filtered by allele: 9–10 RUs Subjects filtered by term: proband and relatives Subjects classified using CCEF annotations: category and FSHD score	134 probands 110 relatives	Probands and their relatives carrying 9 to 10 DRA were phenotypically stratified based on the CCEF categories. Ricci et al., 2020	Large phenotypic variability in 9–10 DRA carriers, including 70% of healthy relatives. Penetrance in families ranged between 20–100%. 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 RUs is not accurate
Subjects filtered by term: proband and relatives Subjects classified using CCEF annotations: Category and FSHD score Disease progression measured as $\Delta$ FSHD score	141 probands 105 relatives	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 ( $\Delta$ FSHD score) Vercelli et al., 2020	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
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	122 probands 110 relatives	The D4Z4 methylation level at 4q35 was assessed in 122 FSHD1 probands and in relatives carrying the same molecular defect presenting classical FSHD (category A), or incomplete/complex phenotype (categories B or D) or no muscle impairment (category C). Nikolic et al., 2020	The D4Z4 methylation levels among index cases and in families show a high variability with no association with clinical manifestation or disease progression The results of D4Z4 methylation analysis must be cautiously interpreted in respect to disease prognosis, which requires family studies
Subjects filtered by term: proband Subjects filtered for non-familial cases Subjects filtered by comorbidities/complex phenotypes	1339 FSHD unrelated cases	Among 1339 unrelated FSHD cases 3 unrelated cases presented signs of Williams-Beuren Syndrome (WBS) in early childhood and later developed FSHD. All 3 cases carry both molecular defects associated with the two disorders. Rodolico et al., 2020	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



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. The limited knowledge about the natural history of RDs, their chronicity, need for long-term follow-up and high demand for assistance represent a special challenge for the healthcare systems. The management of all these needs and questions requires a large amount of information to be collected. In this respect RD or disease-specific data collection hubs with the possibility of integrating data available in different locations and the management of continuously increasing information related to a single sample are key instruments for the surveillance of these diseases, with the aim of improving patients and caregivers quality of life and for therapeutic planning. Thus, the term "Registry" has now gone beyond the original significance of a collection of mere anagraphic/demographic data. The great effort is directed towards

combining genomic, anamnestic and clinical investigation. This important process requires a solid description of clinical data and appropriate Big Data techniques to exploit the potential of the wealth of information that can be acquired. In the case of FSHD, the MOMIS FSHD Web Platform serves 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 FSHD registry is an evolving tool: this is one of its most important features in order to allow longitudinal studies and increasing amount of data elaboration, but it represents a double edged sword as the registry requires to be constantly updated with follow-up information of patients to fully exploit its function, otherwise it is in danger for becoming a static survey of a fixed cohort, which is not our purpose. It would be of essential help to look for an implementation in our database which could facilitate the fill in of individual clinical records both for clinicians and patients themselves.

In the future, Patient Engagement and Mobile data capture solutions can be integrated into the existing database to empower patients through self-assessment of health indicators. Wearable and clinical devices can be used to automatically collect physiological data. Innovative communication strategies will also improve the patient's engagement and Patient Reported Outcomes data (ePRO) can be collected in real time employing the patient's smartphones and mobile devices.

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

Our experience with the INRF shows the possibility of the systematic collection of clinical data for rare genetic diseases on a national scale and its value is providing informative observational data that can support clinicians, researchers and the health system in acquiring precise information on rare disorders. We are confident that our experience with the integrated management of FSHD patients would be taken as an example and applied to other complex diseases in which the genotype/phenotype correlation is far to be resolved.

## Conclusion

The future of medicine lies in personalized treatments. In the case of rare genetic diseases, the integration of data from genomic analyses, standardized clinical evaluation, collection of environmental and lifestyle data, is a precondition to implement person-centered care. The systematic collection of clinical data for rare genetic diseases on a National scale is the tool providing informative observational data that can support this process.

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.

## Abbreviations

FSHD: Facioscapulohumeral dystrophy; DRA: D4Z4 reduced allele; PAS: Polyadenylation 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.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13023-021-02100-z>.

**Additional file 1.** 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.

**Additional file 2.** Summary of main FSHD-registries features.

## 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, Fondazione IRCCS Carlo Besta Neurological Institute, Milan); 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 Clinical 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).

## Authors' 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. JK, EC, SB, MO, LG, LM contributed in data collection and integration, manuscript editing. All authors read and approved the final manuscript.

## Funding

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

**Availability of data and materials**

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](http://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.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy. <sup>2</sup>Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy. <sup>3</sup>DataRiver Srl, Modena, Italy. <sup>4</sup>Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Modena, Italy. <sup>5</sup>FSHD Society, Reading, MA, USA. <sup>6</sup>Department of Molecular Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, USA.

Received: 26 April 2021 Accepted: 24 October 2021

Published online: 04 November 2021

**References**

- Mostacciolo ML, Pastorello E, Vazza G, Miorin M, Angelini C, Tomelleri G, et al. Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. *Clin Genet*. 2009;75:550–5. <https://doi.org/10.1111/j.1399-0004.2009.01158.x>.
- Padberg GW. Facioscapulohumeral disease. Faculty of Medicine, Leiden University Medical Center (LUMC), Leiden University; 1982;247. <http://hdl.handle.net/1887/25818>.
- 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*. 2013;136:3408–17. <https://doi.org/10.1093/brain/awt226>.
- Padberg GW, Lunt PW, Koch M, Fardeau M. Diagnostic criteria for facioscapulohumeral muscular dystrophy. *Neuromuscul Disord Elsevier*. 1991;1:231–4.
- Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet*. 1992;2:26–30.
- 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*. 1997;36:333–9.
- Goto K, Nishino I, Hayashi YK. Very low penetrance in 85 Japanese families with facioscapulohumeral muscular dystrophy 1A. *J Med Genet*. 2004;41:12e–12.
- 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*. 2004;14:33–8.
- 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*. 2009;30:1449–59. <https://doi.org/10.1002/humu.21091>.
- Salsi V, Magdinier F, Tupler R. Does DNA methylation matter in FSHD? *Genes (Basel)*. 2020;11:258.
- 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*. 2012;49:171–8.
- Sakellariou P, Kekou K, Fryssira H, Sofocleous C, Manta P, Panousopoulou A, et al. Mutation spectrum and phenotypic manifestation in FSHD Greek patients. *Neuromuscul Disord*. 2012;22:339–49.
- Ricci G, Zatz M, Tupler R. Facioscapulohumeral muscular dystrophy: more complex than it appears. *Curr Mol Med*. 2014;14:1052–68.
- 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.
- Lamperti C, Fabbri G, Vercelli L, D'Amico R, Frusciantè R, Bonifazi E, et al. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: the FSHD clinical score. *Muscle Nerve*. 2010;42:213–7.
- 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*. 2016;263:1204–14.
- Bergamaschi S, Bouquet P, Giacomuzzi D, Guerra F, Po L, Vincini M. MELIS: an incremental method for the Lexical annotation of domain ontologies. *Int J Semant Web Inf Syst. IGI Global*; 2007. <https://doi.org/10.4018/jswis.2007070103>.
- Magnotta L, Gagliardelli L, Simonini G, Orsini M, Bergamaschi S. Momis dashboard: A powerful data analytics tool for industry 4.0. *Adv Transdiscipl Eng*. 2018;7:1074–81.
- Beneventano D, Bergamaschi S, Gagliardelli L, Po L. Open data for improving Youth Policies. In: IC3K 2015 – proceedings of 7th international jt. conference knowledge discovery and knowledge engineering knowledge management. 2015. [www.opendatabarometer.org](http://www.opendatabarometer.org).
- 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*. The American Society of Human Genetics; 2012;90:628–35. <https://doi.org/10.1016/j.ajhg.2012.02.019>.
- Ruggiero L, Mele F, Manganello F, Bruzzese D, Ricci G, Vercelli L, et al. Phenotypic Variability Among Patients With D4Z4 Reduced Allele Facioscapulohumeral Muscular Dystrophy. *JAMA Netw open*. 2020;3:e204040.
- 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 Nat Res*. 2020;10:1–12. <https://doi.org/10.1038/s41598-020-78578-7>.
- 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*. 2021;268:356–66. <https://doi.org/10.1007/s00415-020-10144-7>.
- 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*. 2020;21:2635.
- 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*. 2020;27:25–9.
- Taruscio D, Kodra Y, Ferrari G, Vittozzi L, Annicchiarico G, Attolini E, et al. The Italian National Rare Diseases Registry. *Blood Transfus*. 2014;12:s606.
- V H, J L, S M, A R, G J, T B, et al. The Canadian Neuromuscular Disease Registry 2010–2019: a decade of facilitating clinical research through a nationwide, pan-neuromuscular disease registry. *J Neuromuscul Dis*. 2021;8:53–61.
- Guien C, Blandin G, Lahaut P, Sanson B, Nehal K, Rabarimeriarajaona S, et al. The French National Registry of patients with Facioscapulohumeral muscular dystrophy. *Orphanet J Rare Dis*. 2018;13:1–10. <https://doi.org/10.1186/s13023-018-0960-x>.
- H N, M K. [Registry of muscular dystrophy (Remudy). Construction of the patient self-report registry and collaboration with overseas network]. *Rinsho Shinkeigaku. Rinsho Shinkeigaku*; 2011;51:901–2.
- Stanimirovic D, Murko E, Battelino T, Groselj U. Development of a pilot rare disease registry: a focus group study of initial steps towards the establishment of a rare disease ecosystem in Slovenia. *Orphanet J Rare Dis*. 2019;14.
- Alonso V, Abaitua I, Zurriaga Ó, Astray J, Errezola M, Aldana-Espinal JM, et al. National rare diseases registry in Spain: pilot study of the Spanish Rare Diseases Registries Research Network (SpainRDR). *Orphanet J Rare Dis*. *BioMed Central*; 2014;9:P5.

32. Hamel J, Johnson N, Tawil R, Martens WB, Dilek N, McDermott MP, et al. Patient-reported symptoms in facioscapulohumeral muscular dystrophy (PRISM-FSHD). *Neurology*. 2019;93:e1180.
33. Krasnianski M, Eger K, Neudecker S, Jakubiczka S, Zierz S. Atypical phenotypes in patients with facioscapulohumeral muscular dystrophy 4q35 deletion. *Arch Neurol*. 2003;60:1421–5.
34. 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. *Hum. Genet*. 2013. p. 1077–130.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

