FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY: CLINICAL AND GENETIC DIAGNOSIS

FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY:

CLINICAL AND GENETIC DIAGNOSIS

By

Kun Li

A Dissertation Submitted to the Faculty of Harvard Medical School in Partial Fulfillment of the Requirements for the Degree of Master of Medical Sciences in Clinical Investigation (MMSCI)

Harvard University, Boston, Massachusetts

May 2022

Area of Concentration: Genetics, Facioscapulohumeral muscular dystrophy, Macrosatellite repeats, Whole genome sequencing, Linkage analysis

Thesis Committee:

Primary Mentor: Prof. Jonathan G. Seidman and Prof. Christine E. Seidman

Content Advisor: Prof. Raju Kucherlapati

Program Representative: Dr. Martina McGrath

External Experts: Prof. Pereira Alexandre

I have reviewed this thesis. It represents work done by the author under my guidance/supervision.

Primary Mentor: Christine E. Seidman, MD and Jonathan G. Seidman, PhD

Contents

Acknowledgement 3

Overview 4

Chapter 1 7

An Unusual Case of Suspected Facioscapulohumeral Muscular Myodystrophy with Cardiac Symptom 7

Abstract 8

Introduction 10

Methods 12

Discussion 18

Strengths and Limitations 20

Conclusion 21

Chapter 2 35

Whole Genome Sequencing and Rare-variant Based Linkage Analysis Enable Diagnosis of Familial Facioscapulohumeral Dystrophy 1 35

Introduction 38

Results 44

Discussion 45

Strengths and Limitations 48

Conclusions 49

Summary of Results and Conclusions 57

Discussion and perspectives 59

Reference 60

Acknowledgement

I will express my appreciation to my mentor Christine E. Seidman, MD and Jonathan G. Seidman, PhD. They always share with me their experience and knowledge without hesitation. I cannot finish this work without their support and mentorship. I would also thank Ping Zhang, MD, PHD, who always care about me and encourage me to pursue knowledge and explore the broader world. Thanks to Daniel Quiat, MD, PhD, Alireza Haghighi, M.D. and DePalma, Steven Robert. It is my pleasure to work and learn with you. I would also like to thank the members of my thesis committee, Prof. Raju Kucherlapati, Dr. Martina McGrath, and Dr. Pereira Alexandre for their guidance and contributions to my thesis projects. So many thanks to Dr. Hein Naomi for providing me with writing and communication services.

Overview

Facioscapulohumeral muscular myodystrophy (FSHD) is an autosomal dominant myodystrophy disorder characterized by initially restricted and a slowly progressive pattern of skeletal muscle weakness and atrophy, with a wide spectrum of disease severity, age at onset, and side-to-side symmetry (1). The prevalence of the disease in different populations ranges from 1:15,000 to 1:20,000, making it the third most prevalent human muscular dystrophy (2).

Typically, FSHD is clinically hallmarked by weakness and atrophy of a specific set of muscles located in the face, periscapular muscles, and upper arms (2). The clinical manifestations are complicated and highly variable with a low detection rate. Usually, the most common initial symptom is periscapular muscles dysfunction induced difficulty of lifting objects or even their arms above shoulder level. Scapular winging, protruding clavicle and rounded shoulders are common in FSHD patients because of shoulder bones protruding out on the back. However, cardiac involvements, which are frequently observed in many other muscular dystrophies, is rare and remains elusive in FSHD (3). Recognition of FSHD with atypical cardiac involvement as main and first symptom requires detailed investigation and remains challenging. Routine cardiac assessment and surveillance have not been recommended in patients without overt symptoms or signs of cardiac disease (4). Therefore, in our study we first want to investigate the detailed clinical course of a FSHD patient whose clinical features mimicking coronary artery disease, providing clues for clinical diagnosis and genetic counseling in this context.

The pathogenic mechanism of FSHD is ectopic production of the double homeobox 4 (DUX4) in skeletal muscle, which is generally associated with the contraction of macrosatellite repeats (D4Z4) in the subtelomeric region of chromosome 4q35 (FSHD1, ~95%) or mutations in SMCHD1 (FSHD2, ~5%) (5, 6). Healthy individuals carry 11-100 highly similar 3.3-kb single D4Z4 units, each containing one copy of entire open reading frame (ORF) of the DUX4 gene, which is only normally expressed in the germline, but is epigenetically suppressed in most somatic tissues including skeletal muscles (2, 7, 8). FSHD1 and 2 share a similar downstream disease-causing mechanism, chromatin relaxation and subsequent DUX4 de-repression, in the setting of a 4qA allele, which contains a permissive locus – DUX4 polyadenylation site (PAS) (9). PAS is of vital importance for pre-mRNA processing and allows for DUX4 pre-mRNA cleavage and extension of polyadenylation to the mRNA (10, 11). Once aberrantly expressed in skeletal muscles, DUX4 triggers multiple cellular toxicities, eventually inducing cell death and muscle dysfunction (12).

The molecular diagnosis of FSHD1 is still based on traditional laboratory methods, such as southern blotting (SB) and molecular combing (MC) used to detect FSHD1-related contractions of D4Z4 on chr 4 (13, 14). These methods are expensive and labor intensive and requires substantial amounts of DNA. Utilizing SB for the diagnosis of FSHD1 has the risk of leading to uncertain results in up to 23% of cases (15). A much more efficient and simpler alternative molecular methods would be preferable. Several efforts, such as next generation sequencing, have been made to establish the

alternative tests. Nevertheless, although high-throughput sequencing of related family members has shown huge significance for investigating genetic disorders, most pedigree-based sequencing research still depend on an ad hoc combination of suboptimal analyses, because of technical sophistication and lack of applicable tools (16). Directly sequencing the D4ZA subtelomeric macrosatellite repeats on chr 4q35 is more complicated and challenging because of several reasons: 1) the high similarity among repeat units and high GC content; 2) the existence of a highly homologous region on the subtelomere of chr 10q26 (5, 13).

Classical linkage analysis is a conventional and predominant statistical approach to assess recombination events between selected common genetic markers and potential pathogenic alleles to search for phenotypic loci, especially for Mendelian disease (16, 17). In the age of next generation sequencing, linkage analysis, often combined with whole genome sequencing (WGS) filtering approaches and tools, has once again exhibited a significant value of using rare variants to investigate the aetiology of genetic disorders (13). And genome-wide association studies (GWASs), also utilizing genetic markers, has been widely adopted for the association tests between common variants and complex traits. Nevertheless, rare variants, which have not been well interrogated by linkage analysis and GWASs, could also play a vital role in explaining a substantial proportion of genetic disorders (18). Therefore, in our second project we aim to explore the effectiveness of rare variants-based linkage analysis for the diagnosis of FSHD.

Chapter 1

An Unusual Case of Suspected Facioscapulohumeral Muscular Myodystrophy with Cardiac Symptom

Kun Li[1,2], Daniel Quiat[1], Fei She[2], Yuanwei Liu[2], Rong He[2], Alireza Haghighi[1], Fang Liu[2], Rui Zhang[2], Steven Robert DePalma[1], Ying Yang[2], Wen Wang[2], Ping Zhang[2], Christine E. Seidman[1], Jonathan G. Seidman[1]

[1]Genetic Center, Harvard Medical School, Boston, MA, USA

[2]Department of Cardiology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

Running Title: Suspected Facioscapulohumeral Muscular Myodystrophy with Cardiac Symptom

Abstract

Background: Facioscapulohumeral muscular myodystrophy (FSHD) is a hereditary myodystrophy typically affecting skeletal muscles of the facial muscles, shoulder girdles, and upper arms. The clinical course of FSHD is highly variable and the heart is rarely involved. We aim to describe the detailed clinical characteristics of an unusual FSHD case.

Methods: We collected data on ten family members from a Chinese FSHD family with their ages, gender, physical examination, investigations, and clinical manifestation details. Radiology, electromyography (EMG), 24-hour Holter monitor, echocardiography, cardiopulmonary exercise test (CPET), audiometry and muscle biopsy were only performed for the proband.

Results: We report the case of a 32-year-old man with recurrent chest pain, elevated myocardial injury markers, and incomplete right bundle branch block (iRBBB). Coronary angiography, echocardiography, and cardiac magnetic resonance (CMR) did not identify any structural or functional abnormality. Medical and family history of muscle weakness suggested the diagnosis of dystrophy with cardiac involvement. Electromyography (EMG), muscle MRI, and muscle biopsy did not provide any indicative evidence for definitive diagnosis. Whole exon sequencing (WES) did not find pathogenic variant. Clinical diagnosis was made according to clinical criteria depending on the presence of typical clinical findings and the absence of other explanations. Cardio-pulmonary exercise test (CPET) revealed reduced excise capacity with maximal exercise 138watt under RAMP protocol with peak oxygen uptake (Peak VO2) was 53 %pred and anaerobic threshold (AT) 43 %pred.

Conclusions: Our study provided detailed information of an unusual FSHD case with cardiac presentation mimicking coronary artery disease. Persistent elevated cardiac enzymes, family history of myopathy could be the prelude of FSHD.

Introduction

Facioscapulohumeral muscular myodystrophy (FSHD) is a relatively common autosomal dominant adult muscular dystrophy with a slowly progressive and regional proximal muscle weakness and atrophy (19). As the third most common muscular dystrophy after Duchenne dystrophy and myotonic dystrophy, its prevalence varies from 1:15,000 to 1:20,000 in different populations (2). It is currently recognized that FSHD is caused by ectopic production of the transcription factor DUX4 in skeletal muscles. There are two subtypes of FSHD: FSHD1 (~95%) associates with the contraction of macrosatellite repeats (D4Z4) in the subtelomeric region of chromosome 4q35 and FSHD2 (~5%) is usually caused by mutations in the SMCHD1 gene (5, 20).

The clinical course and penetrance of FSHD are highly variable. Usually, muscle involvement starts in the muscles of the face, shoulder girdle facial muscles and upper arms that often show asymmetry, followed by weakness of the trunk and lower extremity muscles (21). Clinical manifestations, including age of onset, severity, and progression, could be highly variable, which induces a low detection rate. Compared to other neuromuscular disorders, Beevor's sign, an abnormal upward movement of the umbilicus on flexion of the neck, is frequently observed in FSHD (22). And Coats syndrome which manifests as high-frequency sensorineural hearing loss and retinal vascular disease, is the most common extramuscular complication (23).

Previous studies confirmed minor cardiac abnormalities in FSHD patients, but their nature and clinical course have not been completely depicted. FSHD with cardiac

manifestations is an under-recognized clinical course. Both structural and conduction abnormalities can occur in FSHD, and complete or incomplete right bundle brunch block (RBBB, 12%) and mitral valve prolapse (MVP, 9%) are the most common abnormalities (24). The clinical course of FSHD with cardiac presentations could mimic coronary artery disease with chest pain, myocardial injury markers elevation, and ECG-changes. Discriminating the disease from a patient who manifests as cardiac symptoms is challenging. As a hereditary disorder, the clinical course of FSHD should be fully studied to provide patients and their families with adequate treatment and genetic counseling. The lack of detailed research on cardiac manifestations and the high intra-familial variability makes the differential diagnosis and genetic counseling exceedingly difficult. Herein, we report a FSHD patient with cardiac presentation with detailed information.

Methods

Human subjects: Our patient cohort consisted of 113 patients spanning five different generations suspected to have dystrophy, but without a prior molecular diagnosis. Elements of this study involving human subjects were approved by the Medical Ethics Committee of Beijing Tsinghua Changgung Hospital, Tsinghua University. All participants provided informed written consent before initiating research and data collection.

Clinical assessments

We collected data on ten family members including the proband (VI-5) with their ages, gender, physical examination, investigations, and clinical manifestation details. History-taking, physical examination, and electrocardiogram (ECG) were performed for all included family members. Radiology, electromyography (EMG), 24-hour Holter monitor, echocardiography, cardiopulmonary exercise test (CPET), and audiometry were only performed for the proband. Facial muscles were assessed by completing specific motions, which include whistling, smiling, puckering lips, puffing out cheeks and burying eyelashes. Limb muscles were evaluated by manual muscle testing. Scapular muscle weakness was confirmed according to the standard as follows: 1) scapular wing when abducting arms; 2) unable to lift arms to head level; 3) obvious scapular muscle atrophy (25).

Morphological analyses: Biceps muscle biopsy of the proband were investigated with histological and enzyme histochemical analysis. For histochemical staining, transverse 8 μm cryostat muscle sections were stained with haematoxylin & eosin (H&E), modified

gomori trichome (mGT), Nicotinamide adenosine dinucleotide-tetrazolium reductase (NADH-TR), adenosine triphosphatase (ATP), periodic acid Shiff (PAS), cytochrome oxidase (CCO), oil red O (ORO), acid phosphatase (ACP) and non-specific esterase (NES). For immunohistochemistry analysis included CD4+ T cells, CD8+ T cells, CD20+ B cells, CD68+ macrophages and MHC-I.

Whole exon sequencing (WES): Blood samples were collected for the proband and his mother. Libraries were prepared according to the protocols of Agilent SureSelect QXT Library Prep Kit (5500-0127), and exome capture was conducted based on the protocols of Agilent SureSelect QXT Target Enrichment for Illumina Multiplexed Sequencing version E0. Library DNA quantity and quality were evaluated by Agilent 2200 TapeStation before hybridization. The enriched libraries on the beads were amplified by polymerase chain reaction (PCR) and cleaned with Ampure XP beads (Agencourt, Boston, MA, USA) according to the SureSelect QXT protocol. Paired-end 2×150 bp sequencing was performed on Illumina NovaSeq 6000 System. Genome Analysis Toolkit (GATK) was used to identify variants. Synonymous variants were filtered and excluded. Deleterious single-nucleotide variants (SNVs) were predicted by PolyPhen-2, SIFT and Clinvar. Variants were assessed according to the ACMG (American College of Medical Genetics and Genomics) guideline (26).

Bioinformatic analyses: PhenoPro was utilized for annotating, filtering, and prioritizing the possible disease-causing gene of our family given the variants called from the WES data and a set of phenotypes encoded using Human Phenotype Ontology (HPO) terms assigned to our patients. VCF files (WES) of the proband (VI-5) and his mother (V-2),

and HPO Terms (HPO:0003701 and HPO:0003202) were input into PhenoPro and analyzed according to the published procedure (27).

Statistics Descriptive statistics (median, quartile, and frequency) were used to evaluate each variable, such as onset age and FSHD scores. Statistical analyses were performed with Stata 16.0.

Results

1. Clinical Course

The proband was a 32-year-old man who was admitted to hospital because of recurrent chest pain with persistently elevated myocardial injury markers levels. Coronary angiography from the local hospital showed normal coronary artery. At our visit he informed us he began to develop progressive muscle weakness compared to his peers since he was 18 years old. He was observed to have muscular atrophy in the shoulder, back, and proximal extremities and scapular winging. The proximal muscle strength of his bilateral upper limbs was grade 4, his distal muscle strength was normal, and neurological examination was overall negative, including normal muscle tension, stretch reflexes, and sensation in all parts of the body muscles. He cannot lift his bilateral upper limbs above shoulder level. There was no obvious reduced facial expression, facial movements, and facial diplegia. He reported being able to continue performing physical activities in the IT industry with minimal limitations.

His high-sensitivity troponin T (TnT-hs) was 0.69 ng/mL (reference range, ≤0.14 mg/dL), creatine kinase (CK) was 467U/L (normal range 50–310U/L), CK-MB was 13.65 ng/mL (normal range 0–6.22 ng/mL), and Myoglobin (MYO) was 117.6 ng/mL (normal range 28–72 ng/mL). These myocardial injury markers are persistently elevated during hospitalization (Supplementary Figure 1). No abnormalities in blood routine, liver function, kidney function, and thyroid function.

ECG showed sinus rhythm and incomplete right bundle brunch block (iRBBB) (Supplementary Figure 2). The patient underwent Holter monitoring, which revealed 9% premature ventricular contractions (PVC). No structural or functional abnormalities were revealed by echocardiogram and CMR (Figure 2.A). There was no late gadolinium suggestive of nonspecific fibrosis. CPET revealed reduced excise capacity with maximal exercise 138watt under RAMP protocol. Peak oxygen uptake (Peak VO2) was 53%pred; anaerobic threshold (AT) was 43%pred. No ST-segment changes during exercise.

The MRI of his thigh and shoulder muscles showed degeneration of the supraspinatus tendon without muscle atrophy (Figure 2 B and C). Electromyography did not detect any myotonia or neurogenic changes. Audiometry did not show any auditory impairment. A biceps open biopsy revealed myogenic changes. The muscle fibers were mildly variable in size, scattered with some mild-moderate atrophied muscle fibers and individual small round muscle fibers. Very few muscle fibers were degenerative and necrotic with phagocytosis (Figure 3). Immunohistochemical staining revealed there were very few CD4+ cells and CD68+ cells scattered in the endomysium, while CD8+ cells and CD20+ cells were not observed. MCH-1 and C5b-9 staining were negative (Figure 4).

He reported dozens of family members had developed similar symmetrical or asymmetrical progressive proximal muscle atrophy and weakness and family survey determined the mode of inheritance to be autosomal dominant (Figure 5). The family

lived in a mountain area and refused any other examination except a genetic test. The median age at onset was 26 years and 38% of patients were male. Eighty-eight percent of patients were older than 18-years at symptom onset. Thirteen percent of patients had obvious lower limb weakness and thirteen percent of patients presented with cardiac symptoms. There was no family history of sudden cardiac death. There was no genetic testing of his family members or any known skeletal or cardiac myopathies to the patient's knowledge. WES data of two individuals did not identify any pathogenic or likely pathogenic variant for myopathy and cardiomyopathy. No shared gene variants prioritized by PhenoPro were predicted to be phenotypic correlation with patients in our family and disease-causing (Supplementary table 1-3).

Discussion

We describe the cardiac findings in an unusual case of suspected FSHD from a large family. FSHD is a slowly progressive muscular dystrophy caused by the abnormal production of DUX4 which could induce dysregulation of signaling pathways involved in cell death, muscle differentiation, muscle toxicity oxidative stress responses, immune responses, and protein turnover(28-32). Both FSHD1 and FSHD2 have a similar clinical manifestation. Clinical criteria for the diagnosis of FSHD depends on the existence of typical clinical findings and the absence of other explanations (19). Usually, FSHD is suggested by the presence of facial weakness and weakness of the shoulder scapular stabilizers or foot dorsiflexors, while ptosis, weakness of extraocular muscles or bulbar weakness are absent (19, 33).

In the classical phenotype, FSHD can affect almost all skeletal muscles, but typically spares cardiac muscles. In our cases, the proband had cardiac involvement as main and first symptom. Recurrent chest pain, incomplete RBBB and elevated TnT-hs, mimicking coronary artery disease, made differential diagnosis challenging. Coronary angiography confirmed normal coronary artery and the diagnosis of coronary artery disease (CAD) was excluded. And the presence of muscle weakness and atrophy are not sufficient for the definitive diagnosis of FSHD according to the diagnosis criteria. Whole exon sequencing did not find any pathogenic variants for any known dystrophy. Although we excluded CAD, the differential diagnosis of the disease is challenging under this situation.

Skeletal muscle is rich in multiple enzymes including creatine kinase (CK), lactate dehydrogenase (LDH) and aspartate transaminase (AST). The elevation of these enzymes, which is predominantly due to membrane defects or myonecrosis membrane defects, is a reflection of potential disease process (25). Previously researchers have shown that assessing the levels of these biomarkers may increase significance of other measures to quantify disease progression or regression for FSHD patients (34, 35). Statland et al. reported creatine kinase MB fraction (CK-MB, 6.52 fold change), tissue-type plasminogen activator (PLAT, 1.64 fold change,), myoglobin (2.23 fold change), epidermal growth factor (EGF, 2.33 fold change), chemokine (C-C motif) ligand 2 (1.48 fold change), CD 40 ligand (1.89 fold change), and vitronectin (VTN, 1.28 fold change) were markers that related to FSHD (35). Persistent, asymptomatic elevated CK levels could be the prelude or the only manifestation of FSHD. However, it is not clear if TnT-hs is related to FSHD, especially for patients with cardiac symptom. In our case, TnT-hs were persistently elevated, indicating it could be an important biomarker for cardiac involvement for FSHD patients. In some poor backward areas where sophisticated examinations such as electromyography (EMG), radiology, muscle biopsy, and genetic test are unavailable, biomarkers especially muscular enzymes could play a vital role to help diagnose FSHD.

In many dystrophies, like Duchenne dystrophy and Becker dystrophy, the heart can be seriously affected, even leading to myocardial disease, resulting in cardiomyopathy and heart failure(3, 36). On the contrary, severe cardiomyopathy is uncommon and usually unassociated with FSHD. In our case, CMR and echocardiography did not show any

structural or functional abnormality. Whole exon sequencing did not find pathogenic variant in any known muscular dystrophies and EMG, muscle MRI and muscle biopsy did not provide any positive evidence for definitive diagnosis. Muscle weakness and related family history provided important clue for disease diagnosis and the diagnosis of FSDH was suspected according to the clinical criteria for the diagnosis of FSHD (Table 2).

CPET is a specialized examination for measuring exercise ability by collecting information about the heart and lungs. Not only can it clarify the abnormalities of hemodynamics, but also the changes in the whole body during exercise(37). In our case, CPET successfully detected reduced excise capacity with reduced Peak VO2 and AT. This method can be used as long-term assessments of FSHD. Therefore, CPET is potential method to evaluate disease progression which has the advantage of being noninvasive and ECG monitoring.

Strengths and Limitations

As a rare disease, FSHD has not been fully understood, especially for patient with atypical cardiac involvement. Our study provided detailed information of an unusual FSHD case with cardiac symptom mimicking coronary artery disease. The diagnosis and differential diagnosis are difficult and confusing under such a clinical condition. Our experiences provide precludes for recognizing FSHD.

As a pedigree-based case study, main limitations to this study are the small sample size and the lack of comparison group. Although our report made several new observations

and generated hypotheses about FSHD with cardiac involvement, more data is needed for further investigation. CPET revealed the reduced exercise capacity of our patient, but it is not specific to any particular muscular disease.

Conclusion

We describe an unusual FSHD case with cardiac symptoms as main and first symptoms. FSHD with cardiac presentation is an under-recognized clinical course with symptoms, myocardial injury markers and ECG-changes mimicking coronary artery disease. Persistent elevated TnT-hs, MYO, CK, CK-MB levels could be the prelude of FSHD. Family history investigation could provide important evidence and clues for differential diagnosis and subsequent genetic counseling. We suggest that if a patient shows cardiac presentation with myopathy, especially if they have a family history, FSHD should be included in the differential diagnosis.

Table 1. Clinical histories and phenotypes^ of affected family members.

ID	V-2	V-7	V-13	V-15	V-26	V-30	VI-2	VI-5
Sex	F	F	M	F	M	F	F	M
Age, years	65	51	59	56	56	53	40	32
Age at Onset, y	51	50	14	25	30	23	27	22
Asymmetry	+	+	-	+	-	+	+	-
Facial weakness	-	-	-	-	-	-	-	-
Scapulohumeral weakness	+	+	+	+	+	+	+	+
Scapular wing	-	-	+	+	+	+	-	+
Lower limb weakness	-	-	+	-	-	-	-	-
Cardiac symptoms	-	-	-	-	-	-	-	+
Beevor's sign	-	-	+	-	+	-	-	-

^ Clinical manifestations were identified by physical examinations, performed by the author, in a rural region of China. Modern medical facilities and diagnostic tests were not available.

Table 2. Clinical manifestations of FSHD (19)

Typical features of FSHD	V-2	V-7	V-13	V-15	V-26	V-30	VI-2	VI-5
Onset of weakness affecting the facial muscles or shoulder girdle musculature	+	+	+	+	+	+	+	+
Positive family history	+	+	+	+	+	+	+	+
Asymmetric muscle involvement	+	+	-	+	-	+	+	-
Abdominal weakness	-	-	+	-	+	-	-	-
The presence of retinal vasculopathy or hearing loss in early-onset FSHD	-	-	-	-	-	-	-	-
Not supporting a diagnosis of FSHD								
Ptosis or extraocular muscle involvement	-	-	-	-	-	-	-	-
Lingual involvement or difficulty swallowing	-	-	-	-	-	-	-	-
Prominent contractures	-	-	-	-	-	-	-	-
Cardiomyopathy	-	-	-	-	-	-	-	-
Features on EMG or muscle biopsy suggesting an alternative diagnosis	-	-	-	-	-	-	-	-

Supplementary Table 1. Top 15 candidate disease-causing gene variants prioritized by PhenoPro (VI-5) (27)

References	Gene	Inheritance	#CHROM	POS_start	POS_end	REF	ALT	Func.refGene	QUAL	FILTER	Mutation_type_(annovar)	Mutation_type_(VEP)
rs1195665801	SMPD1	AR	11	6411930	6411942	CCTGGTGCTGGCG	C	exonic	1195.2	PASS	nonframeshift substitution	inframe_deletion
Novel	ATN1	AD	12	7045891	7045894	ACAG	A	exonic	10.3	LowQual;QD	nonframeshift substitution	inframe_deletion
Seen in control databases	RP1L1	AD	8	10467581	10467581	C	CTTTAGTCCCCTCTAACTGCACCCCCTCTTCTTGCAGCCCTTCTTCTGT	exonic	690.8	PASS	nonframeshift substitution	inframe_insertion
rs768950814	CACNA1A	AD	19	13319691	13319691	G	GGGT	exonic	268.9	PASS	nonframeshift substitution	inframe_insertion
rs118026202	DNAH5	AR	5	13911508	13911508	G	A	exonic	1094.4	PASS	nonsynonymous SNV	missense_variant
rs116995413	DNAH5	AR	5	13919366	13919366	G	C	exonic	3466.4	PASS	nonsynonymous SNV	missense_variant
rs754954093	ATXN1	AD	6	16327864	16327870	GTGCTGC	G	exonic	606.4	PASS	nonframeshift substitution	inframe_deletion
rs766029394	ATXN1	AD	6	16327894	16327894	C	CTGA	exonic	66.2	PASS	nonframeshift substitution	inframe_insertion
rs200111316	ATXN1	AD	6	16327900	16327900	C	A	exonic	66.3	PASS	nonsynonymous SNV	missense_variant
rs192917273	XYLT1	AR/Complex	16	17252746	17252746	G	A	exonic	3991.4	PASS	nonsynonymous SNV	missense_variant
rs1396263177	HERC2	AR	15	28447727	28447742	CAGCCAAGCACTGGCA	C	exonic	123.2	PASS	nonframeshift substitution	inframe_deletion
rs146051850	NF1	AD	17	29552200	29552200	A	G	exonic	2735.4	PASS	nonsynonymous SNV	missense_variant
rs114615449	NPHS1	AR	19	36336398	36336398	C	G	exonic	2229.4	PASS	nonsynonymous SNV	missense_variant
rs201822740	NPHS1	AR	19	36340548	36340548	G	T	exonic	3736.4	PASS	nonsynonymous SNV	missense_variant
rs201112075	TRIOBP	AR	22	38120154	38120154	G	A	exonic	335.4	QD	nonsynonymous SNV	missense_variant

Abbreviations: #CHROM, Chromosome number; POS_start, start position; POS_end, end position; AR, Autosomal recessive; AD, Autosomal dominant; REF, Reference; ALT, Alternation; QUAL, Quality.

Supplementary table 2. Top 15 candidate disease-causing gene variants prioritized by PhenoPro (V-2) (27)

References	Gene	Inheritance	#CHROM	POS_start	POS_end	REF	ALT	Func.refGene	QUAL	FILTER	Mutation_type_(annovar)	Mutation_type_(VEP)
rs782422805	ZNF141	AR	4	338197	338197	T	TGA	exonic	81.2	PASS	frameshift substitution	frameshift_variant
rs79869819	ZNF141	AR	4	338200	338200	G	C	exonic	90.9	ABHom	nonsynonymous SNV	missense_variant
rs782668952	ZNF141	AR	4	338201	338203	ATC	A	exonic	78.2	PASS	frameshift substitution	frameshift_variant
Novel	GP1BA	AD/AR	17	4837117	4837195	AAGCCCGACCACCCCAGAGCCCACCTCAGAGCCCGCCCCCAGCCCGACCACCCCCGGAGCCCACCTCAGAGCCCGCCCCC	A	exonic	605.4	PASS	nonframeshift substitution	inframe_deletion
rs750338758	FANCD2	AR	3	10088407	10088408	AG	A	exonic	333.1	PASS	nonframeshift substitution	Benign
rs375350046	FANCD2	AR	3	10088409	10088412	TAAG	T	splicing	309.1	PASS	splicing	Benign
rs768950814	CACNA1A	AD	19	13319691	13319691	G	GGGT	exonic	268.9	PASS	nonframeshift substitution	inframe_insertion
rs118026202	DNAH5	AR	5	13911508	13911508	G	A	exonic	1094.4	PASS	nonsynonymous SNV	missense_variant
rs116995413	DNAH5	AR	5	13919366	13919366	G	C	exonic	3466.4	PASS	nonsynonymous SNV	missense_variant
rs754954093	ATXN1	AD	6	16327864	16327870	GTGCTGC	G	exonic	606.4	PASS	nonframeshift substitution	inframe_deletion
rs182512508	CUBN	AR	10	17127755	17127755	G	C	exonic	1437.2	PASS	nonsynonymous SNV	missense_variant
rs192917273	XYLT1	AR/Complex	16	17252746	17252746	G	A	exonic	3991.4	PASS	nonsynonymous SNV	missense_variant
rs572721907	ABCC9	AD	12	21958154	21958154	G	T	exonic	2456.2	PASS	nonsynonymous SNV	missense_variant
rs542730918	ABCC9	AD	12	21958155	21958155	C	T	exonic	2456.2	PASS	nonsynonymous SNV	missense_variant
rs146051850	NF1	AD	17	29552200	29552200	A	G	exonic	2735.4	PASS	nonsynonymous SNV	missense_variant

Abbreviation: #CHROM, Chromosome number; POS_start, start position; POS_end, end position; AR, Autosomal recessive; AD, Autosomal dominant; REF, Reference; ALT, Alternation; QUAL, Quality.

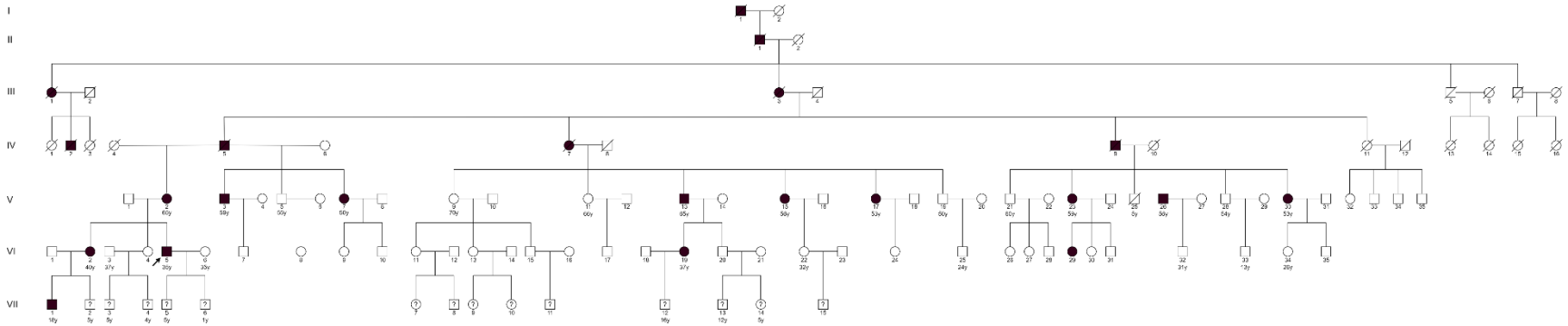
Supplementary Table 3. Top 15 shared candidate disease-causing gene variants between VI-5 and V-2 prioritized by PhenoPro (27)

References	Gene	Inheritance	#CHROM	POS_start	POS_end	REF	ALT	Func.refGene	QUAL	FILTER	Mutation_type_(annovar)	Mutation_type_(VEP)
rs768950814	CACNA1A	AD	19	13319691	13319691	G	GGGT	exonic	268.9	PASS	nonframeshift substitution	inframe_insertion
rs118026202	DNAH5	AR	5	13911508	13911508	G	A	exonic	1094.4	PASS	nonsynonymous SNV	missense_variant
rs116995413	DNAH5	AR	5	13919366	13919366	G	C	exonic	3466.4	PASS	nonsynonymous SNV	missense_variant
rs754954093	ATXN1	AD	6	16327864	16327870	GTGCTGC	G	exonic	606.4	PASS	nonframeshift substitution	inframe_deletion
rs192917273	XYLT1	AR/Complex	16	17252746	17252746	G	A	exonic	3991.4	PASS	nonsynonymous SNV	missense_variant
rs114615449	NPHS1	AR	19	36336398	36336398	C	G	exonic	2229.4	PASS	nonsynonymous SNV	missense_variant
rs201822740	NPHS1	AR	19	36340548	36340548	G	T	exonic	3736.4	PASS	nonsynonymous SNV	missense_variant
Novel	KRT10	AD/AR	17	38975103	38975103	A	AGCTGCCGCCGCCGTATCCGCCGCCGAGCTGCTGCCGCCGCCGTATCCGCCGCCGAGCT	exonic	4743.1	ReadPos	nonframeshift substitution	inframe_insertion
rs186054814	A4GALT	.	22	43089858	43089858	C	T	exonic	3068.4	PASS	nonsynonymous SNV	missense_variant
rs199972616	CORIN	AD	4	47839958	47839958	C	CT	exonic	2290.4	PASS	frameshift substitution	frameshift_variant
rs1304463116	KRT18	AD/AR	12	53343131	53343131	G	GA	exonic	322.4	PASS	frameshift substitution	frameshift_variant

rs147712872	SPTB	AD	14	65260078	65260078	C	T	exonic	2396.2	PASS	nonsynonymous SNV	missense_variant
rs76258939	ABCA4	AD/AR	1	94502888	94502888	A	G	exonic	2437.4	PASS	nonsynonymous SNV	missense_variant
rs1311956591	DIAPH1	AD/AR	5	1.41E+08	1.41E+08	TAAA	T	splicing	142.3	PASS	splicing	Benign
rs200021871	TTN	AD/AR	2	1.8E+08	1.8E+08	T	A	exonic	342.4	PASS	nonsynonymous SNV	missense_variant

Abbreviations: #CHROM, Chromosome number; POS_start, start position; POS_end, end position; AR, Autosomal recessive; AD, Autosomal dominant; REF, Reference; ALT, Alternation; QUAL, Quality.

Figure 1. Pedigree of the family



○ Female, □ Male, ● Females with muscle weakness or atrophy, ■ Males with muscle weakness or atrophy, ? indicates p phenotype is unknown, / indicates deceased.

Figure 2. MRI of the Proband's Heart, Shoulder, and Thigh Muscles.

(A-C) Cardiac magnetic resonance (CMR) did not show cardiac hypertrophy or dilatation. (D-F) T2-W images magnetic resonance imaging (MRI) of thigh muscle. Axial section, coronal section, and sagittal section. (G-I) T2-W images MRI of shoulder muscle. Axial section, coronal section, and sagittal section. No atrophy and fatty-fibrous degeneration of thigh and shoulder muscles.

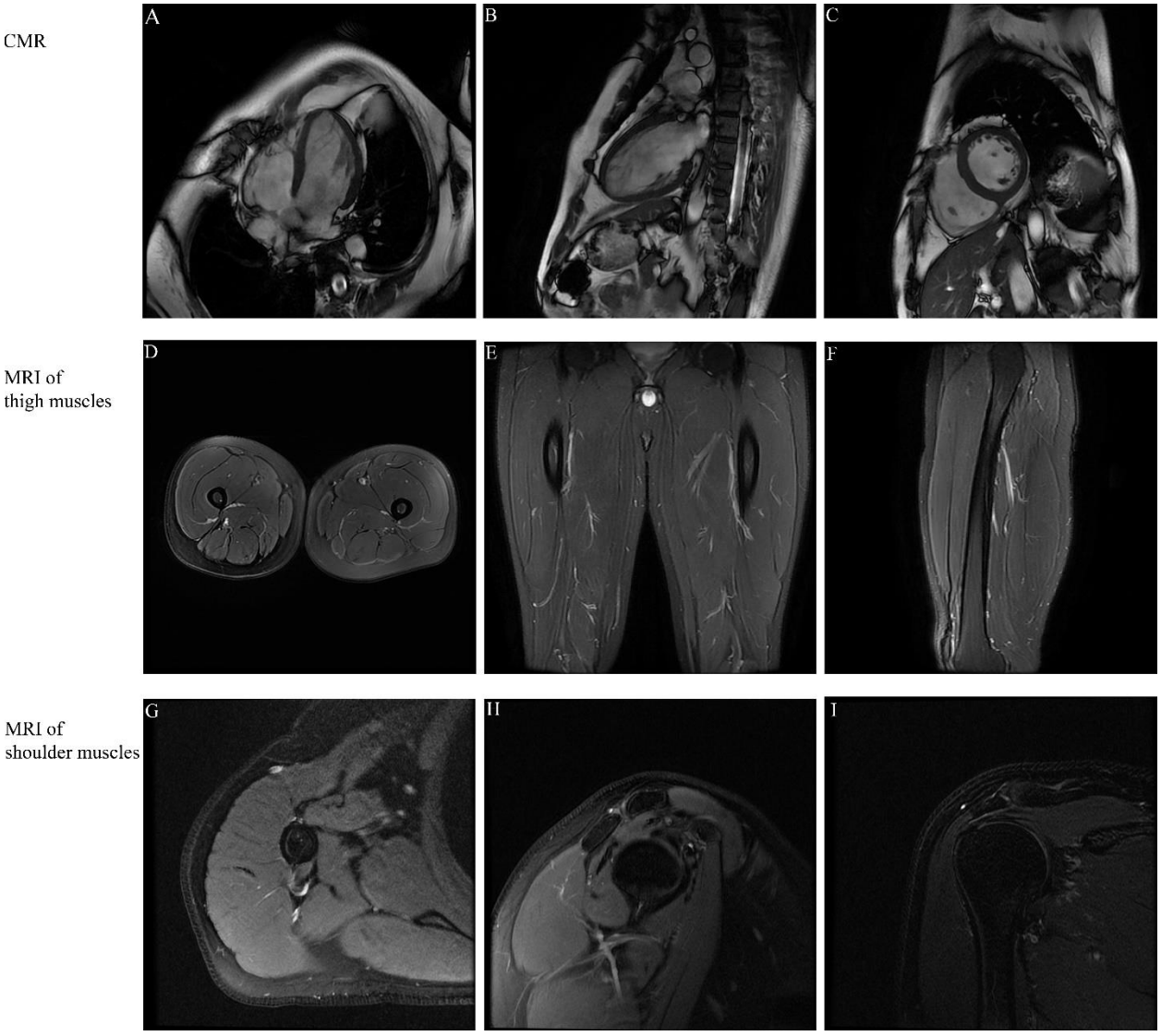


Figure 3. Biceps muscle biopsy with histological and histochemical analysis.

Myogenic changes, variability in fiber size, and scattered with some mild-moderate atrophied muscle fibers and individual small round muscle fibers. Very few muscle fibers were degenerative and necrotic with phagocytosis.

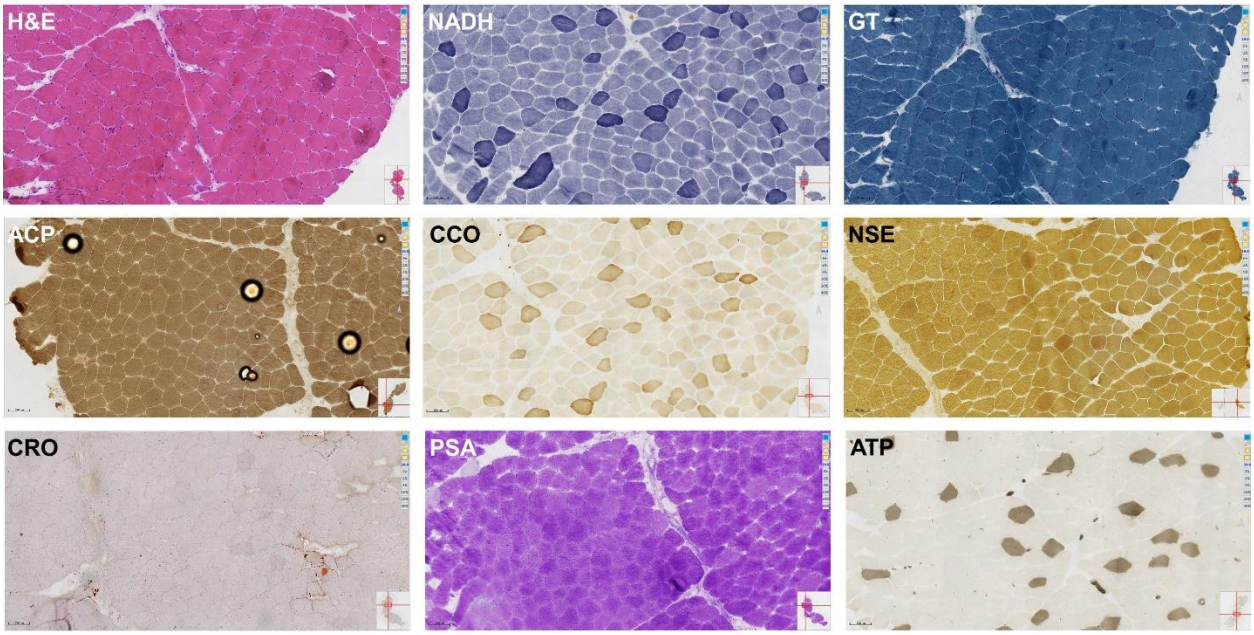


Figure 4. Biceps muscle biopsy with Immunohistochemical analyses. Very few CD4+ cells and CD68+ cells scattered in the endomysium. CD8+ cells and CD20+ cells were not observed. C5b-9 staining was negative.

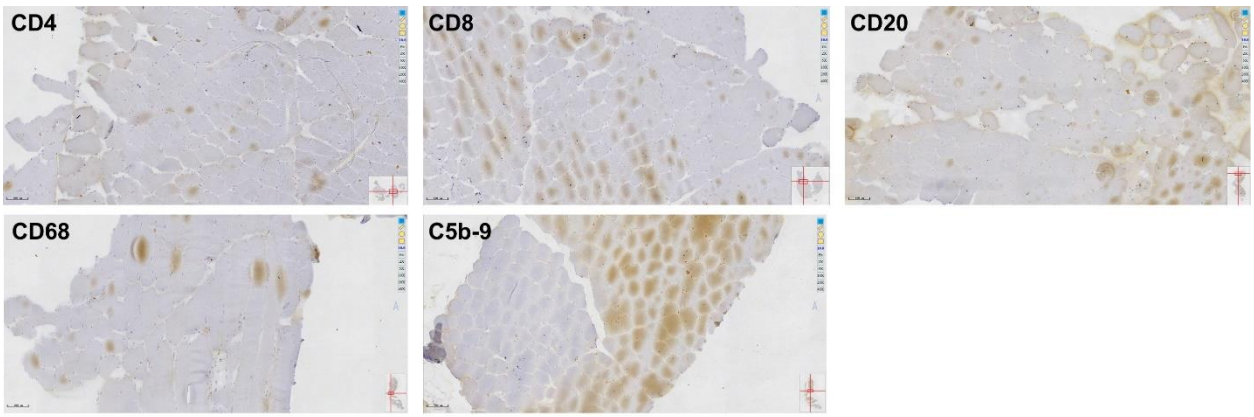
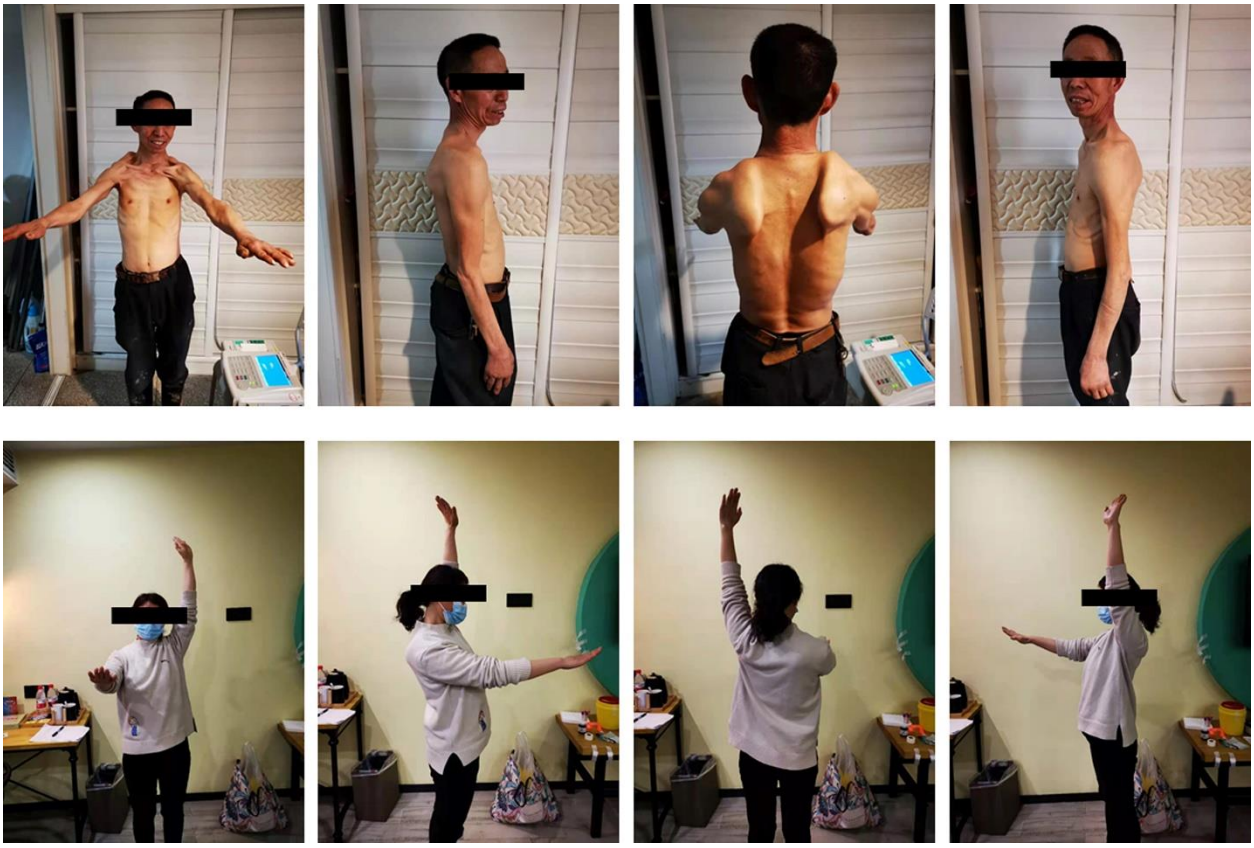
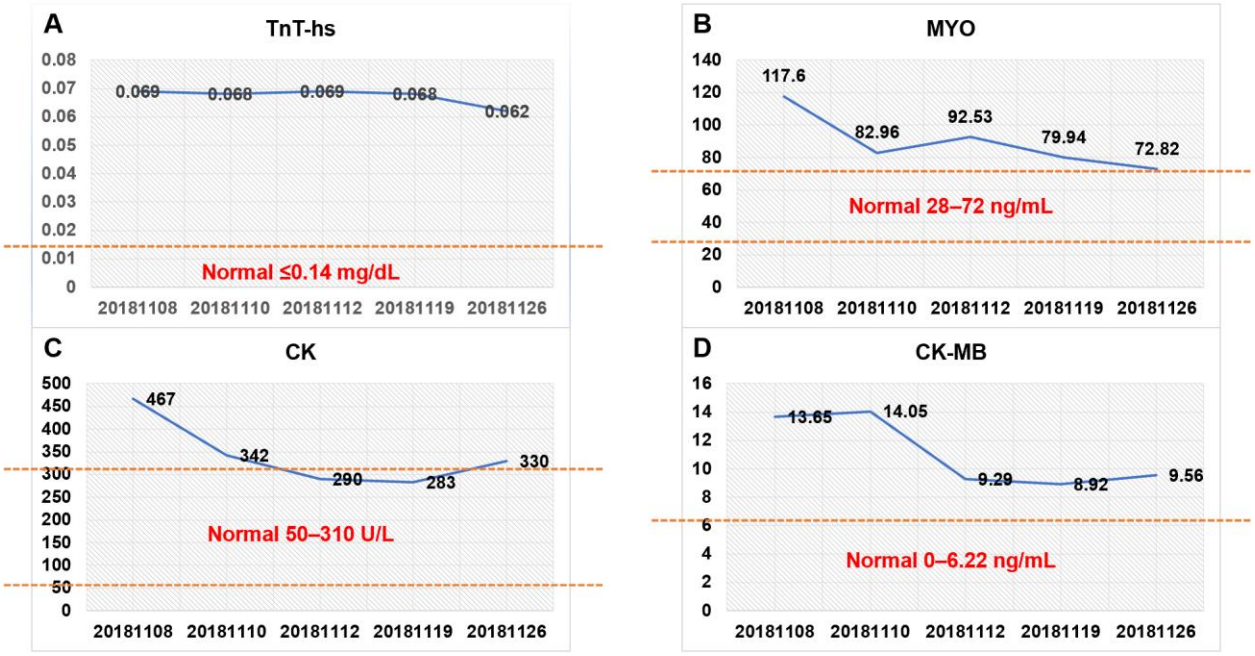


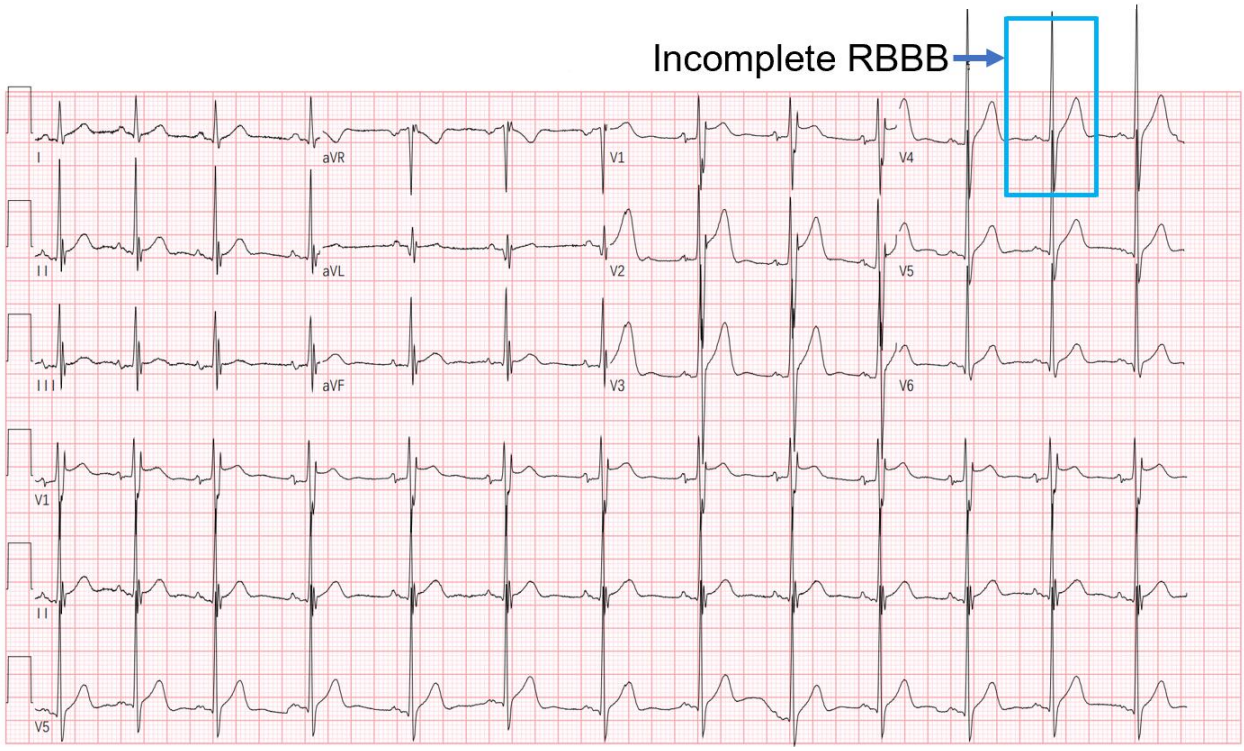
Figure 5. Phenotype of affected family members (V-26 and VI-2). Physical findings include slim body, bilateral atrophy of biceps and pectoralis major, bilateral winged shoulder, and unequal height of shoulders. Note asymmetric findings (greater right involvement) in an affected right-handed woman.



Supplementary Figure 1. Serum biomarkers measured during hospitalization of proband. TnT-hs, high-sensitivity troponin T; MYO, Myoglobin; CK, creatine kinase; CK-MB, creatine kinase MB. Blue line, patient samples, red lines define normal range or values.



Supplementary Figure 2. Electrocardiogram (ECG) of the proband. ECG shows normal heart rate (77 BPM) and cardiac intervals (PR= 142ms, QRS = 100ms, QT/QTc = 360/407 ms) and incomplete right bundle branch block (RBBB).



Chapter 2

Whole Genome Sequencing and Rare-variant Based Linkage Analysis Enable Diagnosis of Familial Facioscapulohumeral Dystrophy 1

Kun Li[1,2], Daniel Quiat[1*], Fei She[2], Yuanwei Liu[2], Rong He[2], Alireza Haghighi[1], Fang Liu[2], Rui Zhang[2], Steven Robert DePalma[1], Ying Yang[2], Wen Wang[2], Ping Zhang[2], Christine E. Seidman[1], Jonathan G. Seidman[1]*

**Kun li and Daniel Quiat contributed equally to this study.*

[1]Genetic Center, Harvard Medical School, Boston, MA, USA

[2]Department of Cardiology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

Running tittle: Using WGS and Rare variants pedigree-based linkage analysis to diagnose FSHD1

Abstract

Background: Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common muscular dystrophies. Over 95% of FSHD cases are caused by deletion of the subtelomeric macrosatellite repeats (D4Z4) on human chromosome 4q35. Directly sequencing these repetitive regions is difficult owing to the high similarity among repeat units and high GC content. The existing FSHD1 diagnostic approaches, including Southern blot and Molecular combing, are not widely used for the reason of time-consuming, labor-intensive, and high equipment requirements. We aim to develop an efficient and accurate procedure for the diagnosis of FSHD.

Methods: Whole genome sequencing was performed for ten individuals from a large FSHD family. Parametric linkage analysis was performed via Merlin (v1.1.2) with the following parameters: dominant model, an estimated population allele frequency of 1E-5, and penetrance of 90%. 4q and 10q haplotypes were characterized by alignment to the chm13 reference genome. A BLAT of the pLAM sequence were performed to identify the A/B haplotypes. We realigned the 250-bp paired end whole genome sequencing reads to the chm13 reference genome using BWA-MEM and then measured the read count for reads containing either the 4q-specific D4Z4 sequence of 4q-specific pLAM sequence and normalized counts to the read depth of the p-arm of chromosome 4.

Results: Rare variants-based linkage analysis identified one single 1.7 MB haplotype on chromosome 4q35.2, presenting in affected individuals and absent from unaffected family members. Parametric linkage analysis resulted in a LOD score of 3.228 for the region. All pedigree samples contained 4q-pLAM sequence suggesting at least one

copy of an 4qA permissive haplotype. Normalized counts of reads containing 4q-specific pLAM sequence were comparable between the FSHD patients, while 4q-specific D4Z4 repeat sequence demonstrated fewer reads in pedigree samples.

Conclusion: Family WGS and rare variants-based linkage analyses, combined with read depth analysis, provide the possibility of detecting the D4Z4 repeat contraction for FSHD1 patients. 4q pLAM specific sequence could be detected by matching to T2T-CHM13 reference.

Introduction

Facioscapulohumeral dystrophy type 1 (FSHD1) is an autosomal dominant myodystrophy with progressive and regional skeletal muscle involvement and variable disease penetrance, which is induced by epigenetic de-repression of the macrosatellite tandem repeats D4Z4 in the subtelomeric region of chromosome 4q (38). The contraction of D4Z4 is pathogenic only when permissive allele A containing the polyadenylation signal (PAS) in the flanking pLAM region exist (9). Although FSHD is usually not life-threatening, it can generate significant morbidity and even loss of working capacity, and almost 20% of patient will need a wheelchair after 50 years old (19).

The existing molecular diagnosis for FSHD1, including southern blot (SB) and molecular combing (MC), are laborious and not widely applied and lots of attempts have been made to construct an alternative diagnostic tool or procedure. In recent decades, high-throughput sequencing of related individuals has become a promising approach for researching inherited disorders. However, because of technical limitations and lack of reliable tools, most of the family-based sequencing research depends on a series of ad hoc filtering criteria (16, 39).

Linkage analysis is a predominant tool utilizing genetic recombination information occurred in the pedigree for locus mapping (16). In recent years, genome-wide association studies (GWASs) using genetic markers has become the preferrable because of its power in studying the association of complex traits and common variants. However, it is realized that as the use of WGS technology becomes more widespread,

linkage analysis, combined with WGS based rare variants, could be applied for a substantial proportion of inherited disorders(40). Although there are several studies that use rare variants for pedigree analysis (41, 42), these approaches only capture association information without incorporating linkage or variant-classification information (40).

Furthermore, the genetic diagnosis of FSHD1 is challenging because it is caused by contraction of macrosatellite repeats (D4Z4) with high GC content and the existence of a highly homologous region on the subtelomere of chr 10q26. Currently used genome reference (GRCh38) were constructed by sequencing bacterial artificial chromosomes (BACs) with several Genome Reference Consortium (GRC) assembly gaps because of incompatible structural polymorphisms on their flanks (43). Besides, plenty of other polymorphic and repetitive regions remain incorrectly assembled or even unfinished (43, 44). These limitations could lead to an underrepresentation of repetitive sequences and a mosaic of haplotypes, making the diagnosis of FSHD1 more difficult (45).

Recently, T2T-CHM13 reference assembled leveraging PacBio HiFi and Oxford Nanopore ultralong-read sequencing, overcomes this technological barrier that has hidden 8% of the genome from sequence-based analysis, including all centromeric regions and the entire short arms of five human chromosomes(43, 46). These technologies and tools were expected to drive future improvement in human genomic health and disease. In this study, we also assessed if 4q35.2 region and pLAM sequence could be successfully mapped by T2T-CHM13 reference.

Methods

Subject and sample collection: Our primary subject cohort consists of eight patients and two unaffected individuals from a lager clinical diagnosed Chinese FSHD family (Figure 1), but without a prior molecular diagnosis. We performed WGS on these participants with available DNA samples. Elements of this study involving human subjects were approved by the Medical Ethics Committee of Beijing Tsinghua Changgung Hospital, Tsinghua University. All participants provided informed written consent before initiating research and data collection.

Whole Genome Sequencing: The NadPrep® EZ DNA Library Preparation Module (for Illumina®) kit was used for library construction, followed by high-throughput, high-depth sequencing on the Illumina Novaseq platform. The genomic DNA is randomly broken into fragments with a length of about 250 bp by endonuclease, and after end repair and A-tail addition, DNA libraries with specific tags are connected to the two ends of the fragments respectively. Finally, PCR amplification of the library was performed. After the library was constructed, the Agilent 2200 Tape Station was used to detect the insert size of the library. qPCR was subsequently adopted to accurately quantify the effective concentration of the library to ensure the quality of the library. Illumina Novaseq platform PE150 sequencing was performed according to the effective concentration of the qualified library and data output requirements. The sequencing results were decontaminated, and the adapters were removed; the statistical results include the number of sequencing reads, data yield, sequencing error rate, Q20 content, Q30 content, and GC content.

Bioinformatic analysis: Reads were aligned to the hg38 reference genome using the Burrows-Wheeler Aligner (BWA-MEM) and processed in accordance with Genome Analysis Toolkit (GATK, Broad Institute) Workflow best practices(47). Briefly, sequence alignments underwent BaseQuality Score Recalibration (BQSR, GATK) and marking of duplicates (MarkDuplicates, Picard), followed by individual level variant calling with HaplotypeCaller (GATK). Individual samples were then joint-genotyped, and variants underwent Variant Quality Score Recalibration (VQSR, GATK). Final SNP and indel calls were filtered with the 'PASS' filter from VQSR LOD scoring, read depth (DP) >=10, and genotype quality (GQ) >=20, and subsequently phased with Eagle (v2.4.1) using the 1000 Genomes Project Phase 3 reference panel.

We ran GATK-SV to with a single sample mode that uses a subset of the 1000 genomes reference samples to call structural variants. We used this mode and ran each of the samples individually. Any structural problem > 50 bp were analyzed.
We next identified haplotypes that were identical-by-descent in affected individuals and absent from unaffected relatives. Under the assumption of an autosomal dominant model based on the pedigree, we used bcftools (v1.10) to filter variant calls for rare heterozygous biallelic SNPs with a frequency in East Asian subjects (EAS) in gnomAD (v3) less than 1E-3.

We then used these haplotype-defining SNP markers to perform a parametric linkage analysis using Merlin (v1.1.2) with the following parameters: dominant model, an estimated population allele frequency of 1E-5, and penetrance of 90%.

The 4q35.2 region mapped in this pedigree has been previously associated with FSHD. FSHD1 is a repeat mediated disease that occurs when an individual carries a 'permissive' haplotype carrying a contraction of tandem D4Z4 repeat elements (< 10?). The complex repeat elements in this region are exemplified by the multiple alternate hg38 reference genome contigs that overlap this area. Recently, a complete telomere-to-telomere reference genome (T2T-CHM13) was constructed by assembly of long-read whole genome sequence, and more accurately captures genomic regions containing complex sequence repeats (PMID 35357953). We hypothesized that FSHD-associated haplotypes could be genotyped, at least in part, by alignment to the chm13 reference genome. We first sought to characterize the 4q and 10q haplotypes present in the reference genome. Evaluation of the simple sequence length polymorphism (SSLP) on the centromertic side of the D4Z4 repeat elements identified 4q-161 and 10q-166 alleles in the reference sequence.

As the pLAM sequence differentiates between A/B haplotypes, we next performed a BLAT of the pLAM sequence (obtained from GenBank sequence AF117653.3) against the chm13 reference and identified complete pLAM sequences on 4q and 10q. We noted multiple SNPs within the pLAM that differentiate 4q and 10q pLAM sequences, with the sequence 'TAAAATGCCCCCTCCCTGTGGATCCTATAGAAGATTTGCATCTTTTGTGTGATG' specific for 4q pLAM. These data suggest homozygous 4A161 and 10A166 haplotypes in the chm13 reference sequence. We next used a previously reported sequence

unique to 4q D4Z4 repeat elements (PMID: 34071558) to quantify the number of D4Z4 repeat elements on chromosome 4q in chm13 reference and identified 32 D4Z4 repeats containing the ‘ACGCCTCCGCCTCCGCGCGG’ motif. Based on these data the chm13 reference genome is homozygous for the 4A161 permissive FSHD haplotype and carries a non-pathogenic number of D4Z4 repeat elements.

After characterizing the FSHD-associated haplotypes in the chm13 reference we aimed to use these data to genotype FSHD alleles in the pedigree and control subjects. We first realigned the 250-bp paired end whole genome sequencing reads to the chm13 reference genome using BWA-MEM. We measured the read count for reads containing either the 4q-specific D4Z4 sequence of 4q-specific pLAM sequence and normalized counts to the read depth of the p-arm of chromosome 4.

Results

WGS data passing QC were successful obtained for all participants. No diseasing-causing structural problem > 50 bp including duplications, deletions, inversions, translocations, and large copy number variants (CNVs) was identified. Totally 15417 rare heterozygous biallelic markers with allele frequency of 1E-5 under the assumption of a dominant model and a 90% penetrance were used for genome-wide linkage analysis. Parametric linkage analysis with Merlin (v1.1.2) identified only one candidate region on chromosome 4 with a LOD score of 3.228, suggesting compelling evidence of genetic linkage (Figure 2). Totally four SNP markers with LOD score > 3 identified by linkage analysis (Table 1). LOD scores were -2 for all other chromosome which indicated convincing evidence against linkage (Supplementary figure 1).

The one single 1.7 MB haplotype containing D4Z4 repeats predisposing to FSHD1 located on chromosome 4q35.2 (Figure 3.A), presenting in affected individuals and absent from unaffected family members (Figure 3.B).

Normalized read depth was comparable between controls and KLMA family members. All pedigree samples contained 4q-pLAM sequence suggesting at least one copy of an 4qA permissive haplotype (Figure 4.A). Totally 7 (87.5%) affected individuals showed lower read depth compared to controls. Normalized counts of reads containing 4q-specific D4Z4 repeat sequence demonstrated fewer reads in pedigree samples (Figure 4.B).

Discussion

High-throughput next generation sequencing (NGS) technologies have been applied to perform genetic diagnosis of peculiar inherited disorders and have many advantages over other approaches. For instance, targeted panel, WES, WGS, and mitochondrial DNA sequencing are common and popular testing for germline mutations(48). Nevertheless, the currently used human reference genome (GRCh38) only covers the euchromatic fraction, with the heterochromatic regions remain unfinished (43). The remaining 8% of the genome regions, such as homologous regions, repetitive regions, and GC-rich regions, are not reliably interpretable by the current NGS methods and standard bioinformatics algorithms (49). An increasing number of researches of microsatellite repeats have verified their functional significance in genome organization and disease development (50). Therefore, it is of vital importance to develop approaches for detecting changes in macrosatellite tandems. FSHD1 is a classical case associated with the contraction of macrosatellite repeats (D4Z4). Southern blotting is the traditional diagnostic test still used. However, disadvantages are obvious, including not all equipment are easily available, is quite expensive and labor intensive, and requires copious amounts of DNA. In addition, some FSHD1 cases are caused by the presence of hybrid alleles, mosaicism and D4Z4 array rearrangements, making the result difficulty to interpret. Molecular combing can resolve some of the limitations by providing direct visualization of the D4Z4 repeats on chr 4q and chr 10q alleles, and even the precise size of D4Z4 repeats(13). However, current application of molecular combing is limited because of the low throughput, and special equipment and chemistry is required.

To simplify the diagnostic approaches and tools, efficiently detecting the 4q35 D4Z4 array sizing and its A haplotyping, we developed a procedure using WGS and rare variants-based linkage analysis (Figure 5).

Usually, GWASs is the most common study that utilizes the association between common SNP markers and complex traits to explore the genetic basis of a disease and has become the preferred association mapping approach. However, recent high-throughput sequencing techniques of WES and WGS now allow comprehensive detection of rare variants, which bring innovative approaches and possibilities for disease etiology study. Family based rare-variant methods has advantages over an ad hoc combination of suboptimal analyses to identify pathogenic regions or variants. Aggregating the rare variants in each gene can also reduce the problems of multiple comparison and increase the statistical power (16).

In our study, we first use WGS data and rare variants-based linkage analysis to identify the haplotype presenting in affected individuals and absent from unaffected family members. This method successfully confirmed the haplotype containing the D4Z4 macrosatellite repeats on chromosome 4q35.2 with a high LOD score, suggesting compelling evidence of genetic linkage.

Previous reports revealed that sometimes southern blotting cannot detect the non-permissive chr 10 haplotype which looks similar to permissive 4q35 haplotypes (38). In

our study, we then performed the second step of confirming the existence of 4qA permissive by detecting the 4q pLAM specific sequence using T2T-CHM13 reference. Compared to the current human reference genome GRCh38, the T2T-CHM13 reference genome adds approximately 200 million base pairs (Mbp) of sequence and universally improves variant identification and read mapping by unlocking the complex regions including macrosatellite repeats (51).

Although the introduction of high-throughput NGS technologies into the clinical practice considerably improved the detection of many suspected genetic disorders, it is still unfortunately not applicable for directly characterizing the repetitive traits of DNA (5, 52). In our research, we found that normalized counts of reads containing 4q-specific D4Z4 repeat sequence showed fewer reads in affected individuals. This result did not directly demonstrate the existence of D4Z4 contraction, but it provides researchers with preclue for suspecting the diagnosis of FSHD, combined with the pedigree's clinical phenotype.

In our study, we use rare variants with frequency 1E-5 for linkage analysis and four SNPs were identified with robust evidence of linkage. Although these SNPs are rare in population, they are common variants in our FSHD1 family. Theoretically, these markers can be utilized for pre-clinical test for screening children with phenotype unknown. This will be the most cost effective and efficient method of disease screening for this family.

Although linkage analysis successfully verified the disease-causing region of 4q35 containing D4Z4 repeats, but we will still do not the exact repeat number of D4Z4 and has the risk of being misled because of incomplete penetrance. Nanopore sequencing is a single-molecule sequencing technology which has the advantage of direct and real-time sequencing the human genome with ultra-long reads and has the ability of exploring epigenetic modifications (53). Satomi et al.(54) has successfully applied nanopore technology to sequence a bacterial artificial chromosome (BAC) clone containing 13 D4Z4 repeats with estimated sequence accuracy of 99.8%. However, although nanopore sequencing has the potential of determining the pathogenicity by directly sequencing the complete D4Z4 array, it has not been used in real FSHD cases. As a novel approach for investigating subtelomeric region containing DNA repetitive traits, it is promising to bring us new knowledge for the diagnosis and mechanism of FSHD1 (55).

Strengths and Limitations

Our research developed an efficient procedure for the diagnosis of FSHD1 with high accuracy in the context of NGS era. We demonstrated the power of rare variants-based linkage analysis in mendelian diseases with high penetrance.

The main consideration of our study is that linkage analysis can only identify the chromosomal region of the disease-causing gene, and usually it is the first step for pathogenic variant exploring. Sample size can limit this approach as we need to sequence adequate family to achieve optimal power. In our study, we successful

mapped the disease-causing region indicating the diagnosis of FSHD1 without knowing the exact number of D4Z4 repeat units. Although further analysis showed the read depth of D4Z4 repeats was lower in most of the affected individuals, it did not provide accurate number of the macrosatellite repeat units. And an overlap of read depth existed between affected and unaffected individuals. A more accurate approach is needed to evaluate the contraction and methylation of the D4z4 repeats, as D4Z4 repeats number will influence the severity and risk of disease.

Conclusions

Our workflow includes genotyping by WGS, rare variants-based linkage analyses, 4q pLAM specific sequence and read depth analysis (Figure 5). This workflow used 4q pLAM specific sequences present in the newly released T2T-CHM13 human reference sequence. Family data based whole genome sequencing, rare variants-based linkage analyses and read depth analysis, detected a D4Z4 repeat contraction on chromosome 4q35. Although the precise number of remaining D4Z4 repeats is uncertain, these data support a diagnosis for FSHD1 in the family. With the decreasing cost of performing pedigree based WGS, we expect that this workflow may be widely applied to other rare disorders involving sophisticated genomic structural problems.

Table 1. Four SNPs with high LOD score identified by linkage analysis

#CHROM	References	POS	REF	ALT	Zygosity
chr4	rs367635682	188855116	T	C	Heterozygous
chr4	rs138969910	188903131	T	A	Heterozygous
chr4	.	189894661	C	T	Heterozygous
chr4	rs910574891	189975788	A	G	Heterozygous

Abbreviations: SNPs, single nucleotide polymorphisms; #CHROM, Chromosome number; POS, Position; REF, Reference; ALT, Alternation.

Figure 1. Family pedigree

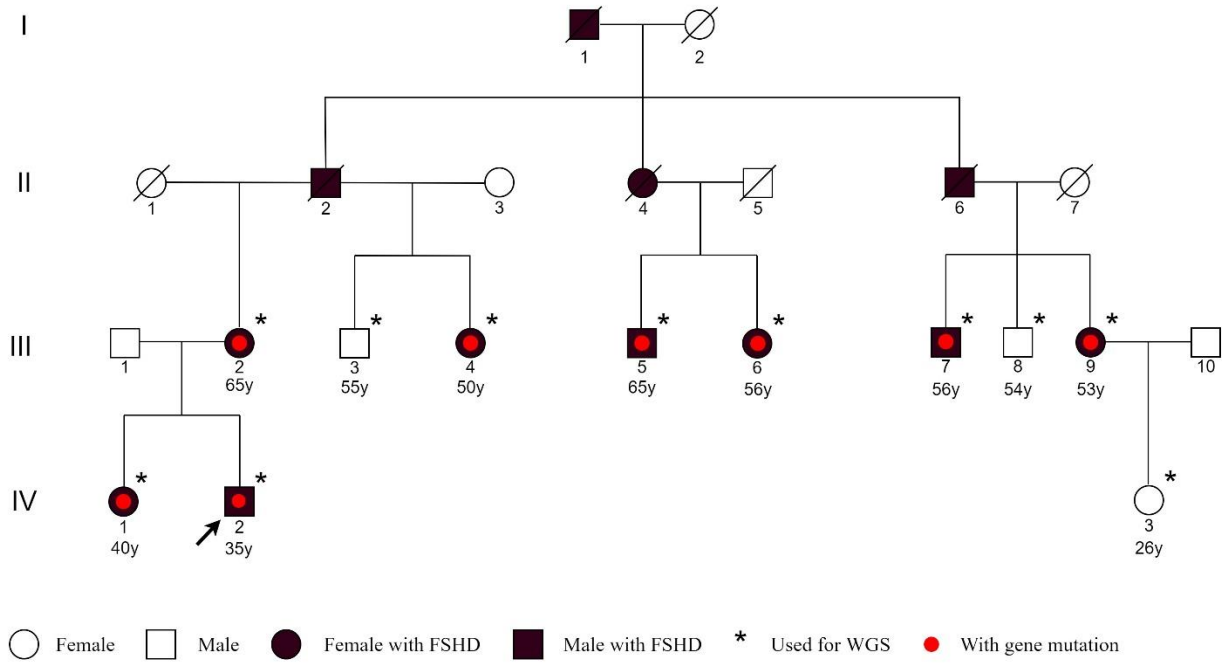
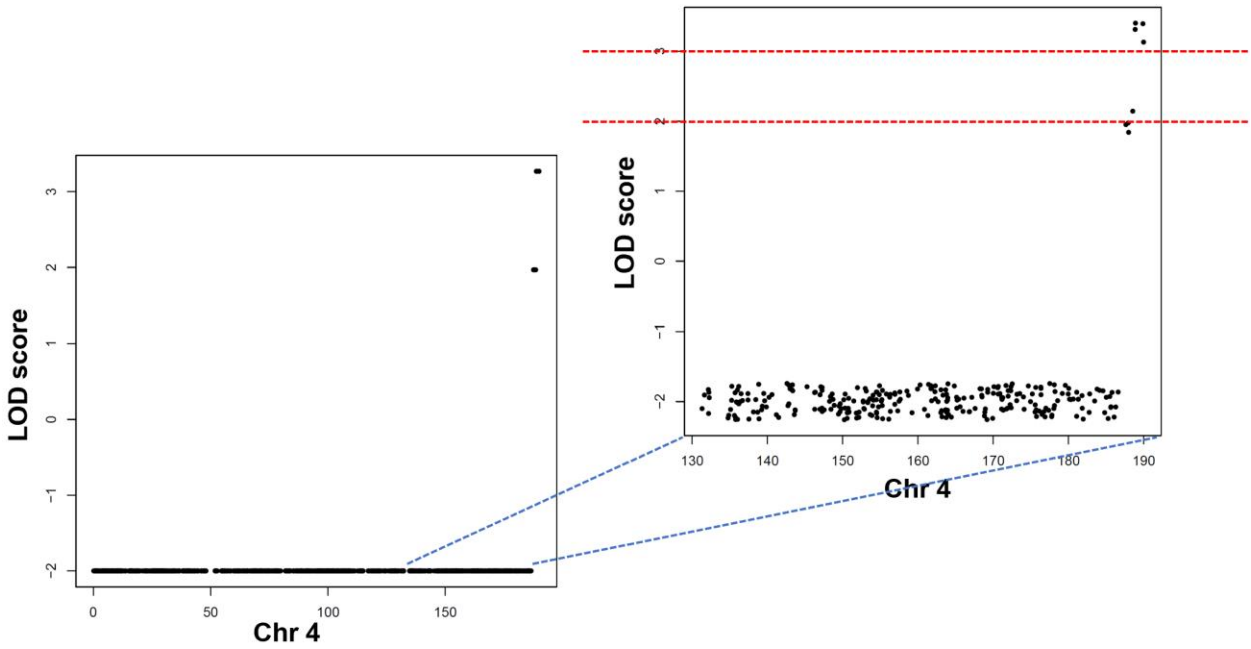


Figure 2. Rare variant analyses define candidate disease locus on chromosome 4q35.2. LOD scores across all autosomes were calculated. Analyses of chromosome 4 (left lower panel) indicated high scores at 4q35.2. This region is shown in detail on the upper right panel.



Black dots denote SNPs studied at the chromosome position indicated. Red lines denote LOD scores of 2-3.

Figure 3. A Chr4.35 locus contains disease gene. A. A map of the chromosome 4q35.2. B. One single 1.7MB haplotype on chromosome 4q35.2 is present in affected individuals and absent from unaffected family members. Abbreviations: UA, unaffected; A, affected.

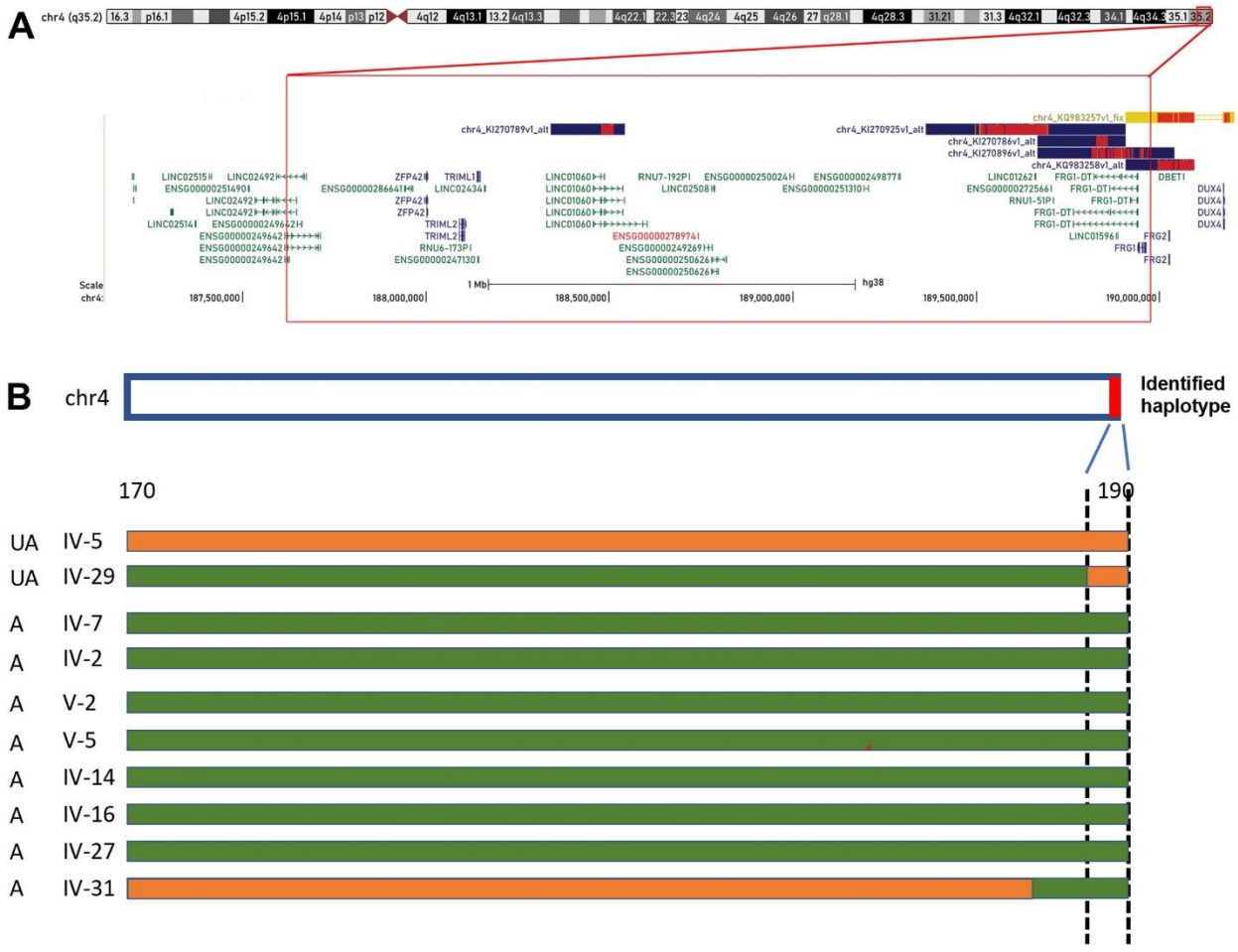


Figure 4. Normalized read depth analyses. (A) Normalized read depth of sequences at 4q-pLAM sequence was comparable between controls and KLMA family members; (B) Seven (87.5%) affected individuals showed lower read depth compared to controls.

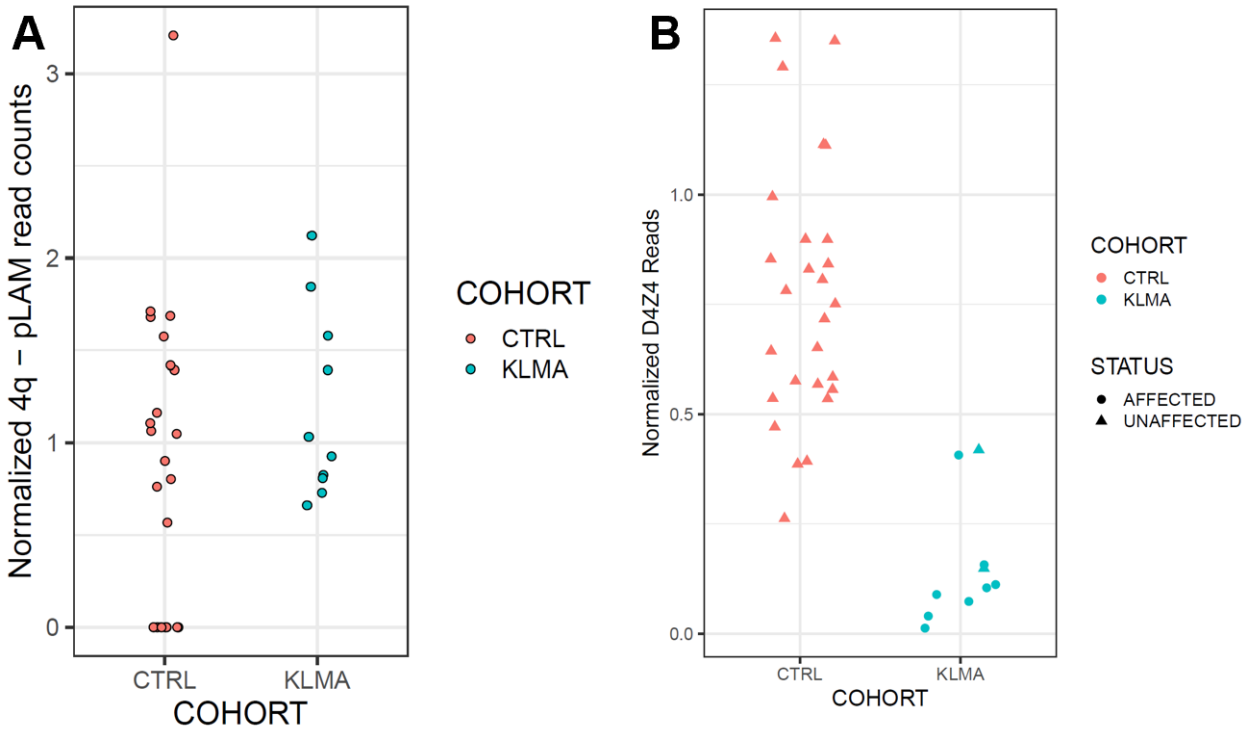
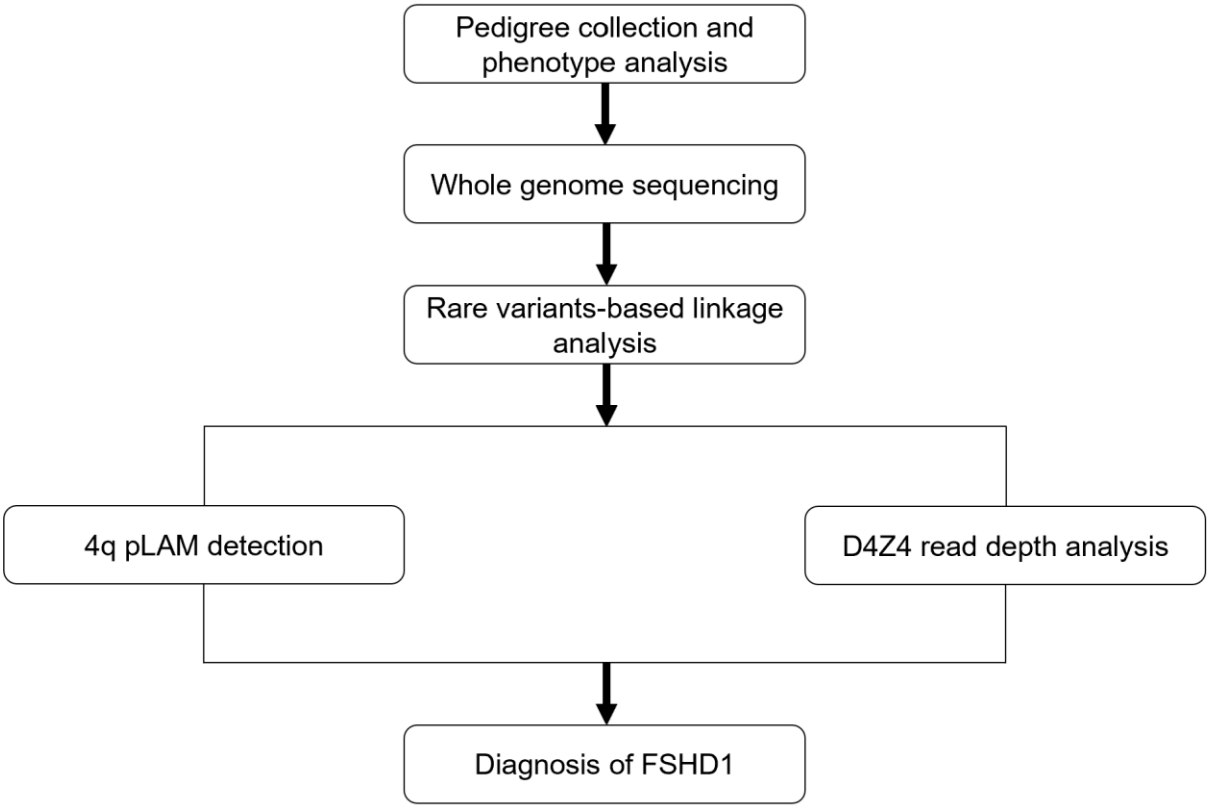
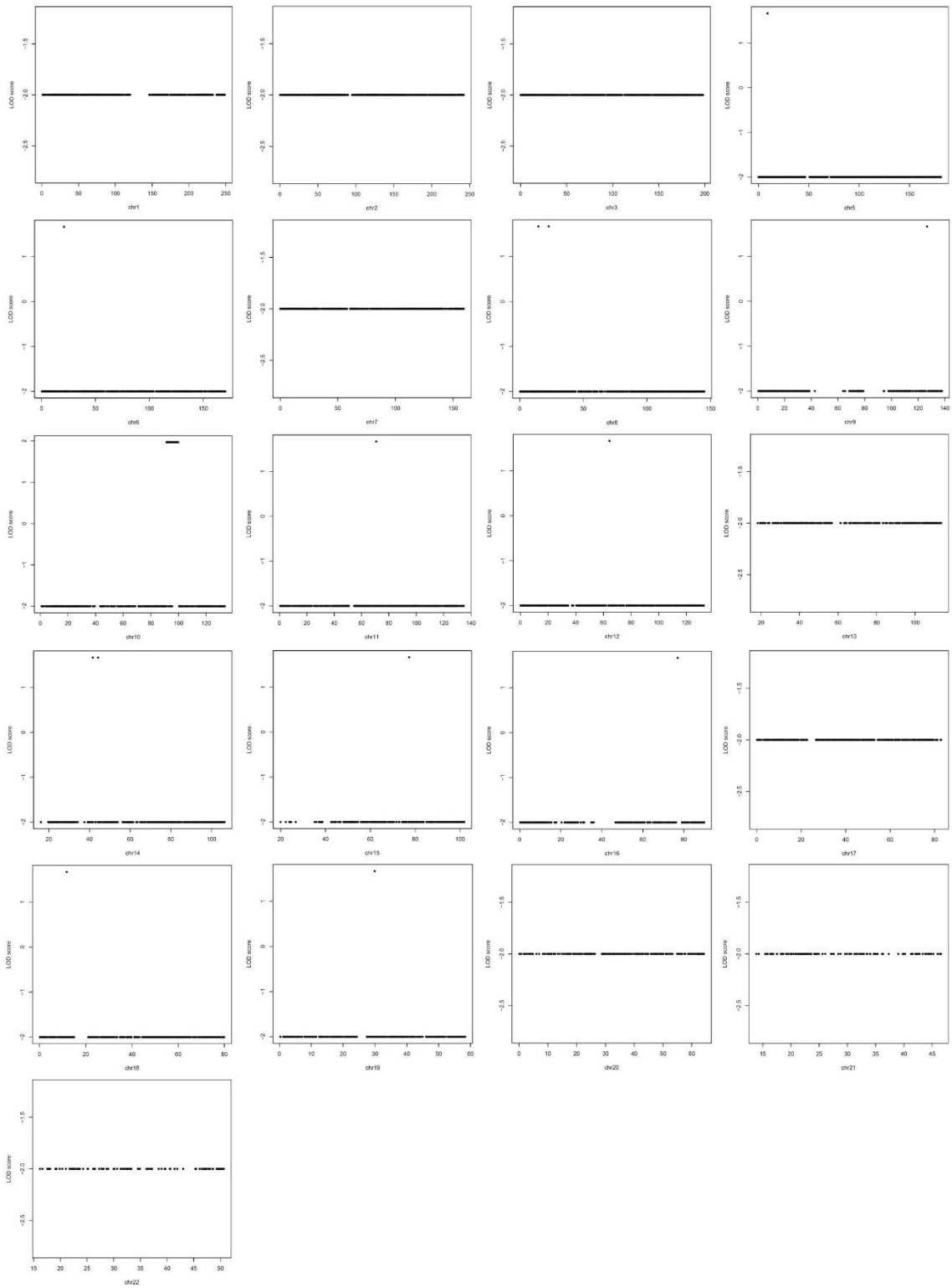


Figure 5. Proposed flow chart for genetic diagnosis of FSHD1



Supplementary figure 1. Genome-wide LOD scores by Merlin (v1.1.2) except chromosome 4



Summary of Results and Conclusions

According to two unmet needs, our work focuses on two overarching goals: First, describe the cardiac manifestations of a FSHD case, and provide adequate genetic counseling for such patients. Second, explore the efficiency and accuracy of rare variants-based linkage analysis for the genetic diagnosis and assessment of FSHD.

In our first project, we described the detailed clinical features of a FSHD Patient whose onset mimicking coronary artery disease, with recurrent chest pain, myocardial injury markers elevation and iRBBB. Apart from CK, other cardiac enzymes including TnT-hs, MYO and CK-MB, could also be persistently elevated in FSHD patients. Family history investigation could help clinical diagnosis in the context of atypical case and subsequent genetic counseling. These results indicate that if a patient shows cardiac presentation accompanied by myopathy, especially if they have a family history, FSHD should be included in the differential diagnosis.

CPET revealed reduced excise capacity with maximal exercise 138watt under RAMP protocol. Peak oxygen uptake (Peak VO2) was 53 %pred; anaerobic threshold (AT) was 43 %pred. These results showed CPET could be a potential approach to evaluate and excise capacity disease progression.

FSHD is clinical diagnosed in project 1. In project 2, we aim to develop an efficient and informative diagnostic procedure for FSHD, because the limitations of current molecular diagnosis depending on southern blot or molecular combing. We hypothesized that pedigree WGS data and rare variants-based linkage analysis can identify disease-causing genes in a FSHD1 pedigrees.

In our research, we successfully identify one single 1.7MB haplotype on chromosome 4q35.2 using 15417 rare heterozygous biallelic markers. Parametric linkage analysis with Merlin (v1.1.2) resulted in a LOD score of 3.228 for the region on 4q35.2 suggesting convincing evidence of genetic linkage. By alignment to T2T-CHM13 reference, we confirmed the existence of 4qA permissive by detecting the 4q pLAM specific sequence. Normalized counts of reads containing 4q-specific D4Z4 repeat sequence demonstrated fewer reads in pedigree samples. All evidence indicates the diagnosis of FSHD1.

Discussion and perspectives

The present studies describe an unusual FSHD case and his family, and subsequently establish a unique procedure for the genetic diagnosis and evaluation of FSHD using rare variants-based linkage analysis.

First, despite a paucity of evidence on the clinical characteristic, through description of the detailed clinical course, our results provide some clue for clinical diagnosis and genetic counseling for FSHD family. A patient shows cardiac presentation accompanied by myopathy, especially if they have a family history, FSHD should be included in the differential diagnosis. More clinical data will be needed for future study.

Second, we provided an efficient and accurate procedure for the genetic diagnosis and evaluation of FSHD. We performed WGS for the family and rare variants-based linkage analysis and read depth analysis, providing a simple and efficient path for the diagnosis of FSHD1. T2T-CHM13 reference could more accurately capture genomic regions containing complex sequence repeats.

Reference

1. Tawil R, Van Der Maarel SM. Facioscapulohumeral muscular dystrophy. Muscle Nerve. 2006;34(1):1-15.
2. Tawil R. Facioscapulohumeral muscular dystrophy. Handb Clin Neurol. 2018;148:541-8.
3. Hermans MC, Pinto YM, Merkies IS, de Die-Smulders CE, Crijns HJ, Faber CG. Hereditary muscular dystrophies and the heart. Neuromuscul Disord. 2010;20(8):479-92.
4. Tawil R, Kissel JT, Heatwole C, Pandya S, Gronseth G, Benatar M. Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the Practice Issues Review Panel of the American Association of Neuromuscular & Electrodiagnostic Medicine. Neurology. 2015;85(4):357-64.
5. Zampatti S, Colantoni L, Strafella C, Galota RM, Caputo V, Campoli G, et al. Facioscapulohumeral muscular dystrophy (FSHD) molecular diagnosis: from traditional technology to the NGS era. Neurogenetics. 2019;20(2):57-64.
6. Yao Z, Snider L, Balog J, Lemmers RJ, Van Der Maarel SM, Tawil R, et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Hum Mol Genet. 2014;23(20):5342-52.
7. Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, Nelson AM, et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. 2010;6(10):e1001181.
8. Le Gall L, Sidlauskaite E, Mariot V, Dumonceaux J. Therapeutic Strategies Targeting DUX4 in FSHD. J Clin Med. 2020;9(9).
9. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 2010;329(5999):1650-3.
10. Peart N, Wagner EJ. A distal auxiliary element facilitates cleavage and polyadenylation of Dux4 mRNA in the pathogenic haplotype of FSHD. Hum Genet. 2017;136(9):1291-301.
11. Wang LH, Tawil R. Current Therapeutic Approaches in FSHD. J Neuromuscul Dis. 2021;8(3):441-51.
12. Daxinger L, Tapscott SJ, van der Maarel SM. Genetic and epigenetic contributors to FSHD. Curr Opin Genet Dev. 2015;33:56-61.
13. Vasale J, Boyar F, Jocson M, Sulcova V, Chan P, Liaquat K, et al. Molecular combing compared to Southern blot for measuring D4Z4 contractions in FSHD. Neuromuscul Disord. 2015;25(12):945-51.
14. Salort-Campana E, Nguyen K, Lévy N, Pouget J, Attarian S. [Clinical and molecular diagnosis of facioscapulohumeral dystrophy type 1 (FSHD1) in 2012]. Rev Neurol (Paris). 2013;169(8-9):573-82.
15. Yanoov-Sharav M, Leshinsky-Silver E, Cohen S, Vinkler C, Michelson M, Lerman-Sagie T, et al. Genetic counseling and testing for FSHD (facioscapulohumeral muscular dystrophy) in the Israeli population. J Genet Couns. 2012;21(4):557-63.
16. Hu H, Roach JC, Coon H, Guthery SL, Voelkerding KV, Margraf RL, et al. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. Nat Biotechnol. 2014;32(7):663-9.
17. Borecki IB, Province MA. Linkage and association: basic concepts. Adv Genet. 2008;60:51-74.
18. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95(1):5-23.
19. Statland JM, Tawil R. Facioscapulohumeral Muscular Dystrophy. Continuum (Minneap Minn). 2016;22(6, Muscle and Neuromuscular Junction Disorders):1916-31.
20. Sidlauskaite E, Le Gall L, Mariot V, Dumonceaux J. DUX4 Expression in FSHD Muscles: Focus on Its mRNA Regulation. J Pers Med. 2020;10(3).
21. Statland J, Tawil R. Facioscapulohumeral muscular dystrophy. Neurol Clin. 2014;32(3):721-8, ix.

22. Awerbuch GI, Nigro MA, Wishnow R. Beevor's sign and facioscapulohumeral dystrophy. Arch Neurol. 1990;47(11):1208-9.
23. Statland JM, Sacconi S, Farmakidis C, Donlin-Smith CM, Chung M, Tawil R. Coats syndrome in facioscapulohumeral dystrophy type 1: frequency and D4Z4 contraction size. Neurology. 2013;80(13):1247-50.
24. Ducharme-Smith A, Nicolau S, Chahal CAA, Ducharme-Smith K, Rehman S, Jaliparthy K, et al. Cardiac Involvement in Facioscapulohumeral Muscular Dystrophy (FSHD). Front Neurol. 2021;12:668180.
25. Zhang Y, Huang JJ, Wang ZQ, Wang N, Wu ZY. Value of muscle enzyme measurement in evaluating different neuromuscular diseases. Clin Chim Acta. 2012;413(3-4):520-4.
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
27. Li Z, Zhang F, Wang Y, Qiu Y, Wu Y, Lu Y, et al. PhenoPro: a novel toolkit for assisting in the diagnosis of Mendelian disease. Bioinformatics. 2019;35(19):3559-66.
28. Mah JK, Chen YW. A Pediatric Review of Facioscapulohumeral Muscular Dystrophy. J Pediatr Neurol. 2018;16(4):222-31.
29. Ansseau E, Laoudj-Chenivesse D, Marcowycz A, Tassin A, Vanderplanck C, Sauvage S, et al. DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. PLoS One. 2009;4(10):e7482.
30. Ansseau E, Eidahl JO, Lancelot C, Tassin A, Matteotti C, Yip C, et al. Homologous Transcription Factors DUX4 and DUX4c Associate with Cytoplasmic Proteins during Muscle Differentiation. PLoS One. 2016;11(1):e0146893.
31. Xiao T, Yang H, Gan S, Wu L. A pediatric case report and literature review of facioscapulohumeral muscular dystrophy type1. Medicine (Baltimore). 2021;100(47):e27907.
32. Rashnonejad A, Amini-Chermahini G, Taylor NK, Wein N, Harper SQ. Designed U7 snRNAs inhibit DUX4 expression and improve FSHD-associated outcomes in DUX4 overexpressing cells and FSHD patient myotubes. Mol Ther Nucleic Acids. 2021;23:476-86.
33. He JJ, Lin XD, Lin F, Xu GR, Xu LQ, Hu W, et al. Clinical and genetic features of patients with facial-sparing facioscapulohumeral muscular dystrophy. Eur J Neurol. 2018;25(2):356-64.
34. Petek LM, Rickard AM, Budech C, Poliachik SL, Shaw D, Ferguson MR, et al. A cross sectional study of two independent cohorts identifies serum biomarkers for facioscapulohumeral muscular dystrophy (FSHD). Neuromuscul Disord. 2016;26(7):405-13.
35. Statland J, Donlin-Smith CM, Tapscott SJ, van der Maarel S, Tawil R. Multiplex Screen of Serum Biomarkers in Facioscapulohumeral Muscular Dystrophy. J Neuromuscul Dis. 2014;1(2):181-90.
36. Emery AE. The muscular dystrophies. Lancet. 2002;359(9307):687-95.
37. Adachi H. Cardiopulmonary Exercise Test. Int Heart J. 2017;58(5):654-65.
38. Zernov NV, Guskova AA, Skoblov MY. FSHD1 Diagnosis in a Russian Population Using a qPCR-Based Approach. Diagnostics (Basel). 2021;11(6).
39. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science. 2010;328(5978):636-9.
40. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 2015;16(5):275-84.
41. Schaid DJ, McDonnell SK, Sinnwell JP, Thibodeau SN. Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. Genet Epidemiol. 2013;37(5):409-18.

42. Oualkacha K, Dastani Z, Li R, Cingolani PE, Spector TD, Hammond CJ, et al. Adjusted sequence kernel association test for rare variants controlling for cryptic and family relatedness. Genet Epidemiol. 2013;37(4):366-76.
43. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44-53.
44. Eichler EE, Clark RA, She X. An assessment of the sequence gaps: unfinished business in a finished human genome. Nat Rev Genet. 2004;5(5):345-54.
45. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 2017;27(5):849-64.
46. Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. Genome Res. 2017;27(5):677-85.
47. Poplin R, Ruano-Rubio V, DePristo M, Fennell T, Carneiro M, Van der Auwera G, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv; 2017.
48. Kanzi AM, San JE, Chimukangara B, Wilkinson E, Fish M, Ramsuran V, et al. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. Front Genet. 2020;11:544162.
49. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. Arch Pathol Lab Med. 2017;141(11):1544-57.
50. Dumbovic G, Forcales SV, Perucho M. Emerging roles of macrosatellite repeats in genome organization and disease development. Epigenetics. 2017;12(7):515-26.
51. Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, et al. A complete reference genome improves analysis of human genetic variation. Science. 2022;376(6588):eabl3533.
52. Adams DR, Eng CM. Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. N Engl J Med. 2018;379(14):1353-62.
53. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. Dev Growth Differ. 2019;61(5):316-26.
54. Mitsuhashi S, Nakagawa S, Takahashi Ueda M, Imanishi T, Frith MC, Mitsuhashi H. Nanopore-based single molecule sequencing of the D4Z4 array responsible for facioscapulohumeral muscular dystrophy. Sci Rep. 2017;7(1):14789.
55. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36(4):338-45.