Journal of Human Genetics (2017) 62, 243–252

www.nature.com/jhg

ORIGINAL ARTICLE

The sensitivity of exome sequencing in identifying pathogenic mutations for LGMD in the United States

Hemakumar M Reddy[1], Kyung-Ah Cho[1,12], Monkol Lek[2,3], Elicia Estrella[4], Elise Valkanas[2,3], Michael D Jones[1], Satomi Mitsuhashi[5,13], Basil T Darras[5], Anthony A Amato[6], Hart GW Lidov[7], Catherine A Brownstein[4,8], David M Margulies[4,8], Timothy W Yu[4], Mustafa A Salih[9], Louis M Kunkel[4], Daniel G MacArthur[2,3] and Peter B Kang[1,5,10,11]

The current study characterizes a cohort of limb-girdle muscular dystrophy (LGMD) in the United States using whole-exome sequencing. Fifty-five families affected by LGMD were recruited using an institutionally approved protocol. Exome sequencing was performed on probands and selected parental samples. Pathogenic mutations and cosegregation patterns were confirmed by Sanger sequencing. Twenty-two families (40%) had novel and previously reported pathogenic mutations, primarily in LGMD genes, and also in genes for Duchenne muscular dystrophy, facioscapulohumeral muscular dystrophy, congenital myopathy, myofibrillar myopathy, inclusion body myopathy and Pompe disease. One family was diagnosed via clinical testing. Dominant mutations were identified in *COL6A1*, *COL6A3*, *FLNC*, *LMNA*, *RYR1*, *SMCHD1* and *VCP*, recessive mutations in *ANO5*, *CAPN3*, *GAA*, *LAMA2*, *SGCA* and *SGCG,* and X-linked mutations in *DMD*. A previously reported variant in *DMD* was confirmed to be benign. Exome sequencing is a powerful diagnostic tool for LGMD. Despite careful phenotypic screening, pathogenic mutations were found in other muscle disease genes, largely accounting for the increased sensitivity of exome sequencing. Our experience suggests that broad sequencing panels are useful for these analyses because of the phenotypic overlap of many neuromuscular conditions. The confirmation of a benign *DMD* variant illustrates the potential of exome sequencing to help determine pathogenicity.
Journal of Human Genetics (2017) **62,** 243–252; doi:10.1038/jhg.2016.116; published online 6 October 2016

INTRODUCTION

Limb-girdle muscular dystrophy (LGMD) is a broad and increasingly heterogeneous category of inherited muscle diseases.[1] LGMD typically causes progressive proximal muscle weakness and has been associated with classic histological abnormalities on muscle biopsy. As genetic discoveries in LGMD proliferate, it has become clear that the clinical and histological presentations, as well as outcomes, may vary widely between subtypes and among different affected individuals. However, these variations are not consistent enough to enable clinicians to identify subtypes based on phenotype alone. Two major subcategories are recognized based on inheritance patterns: LGMD type 1 (LGMD1) is dominantly inherited and LGMD type 2 (LGMD2) is recessively inherited. To date, 8 dominant forms (LGMD1A–H) and 23 recessive forms (LGMD2A–W) have been described, each corresponding to a different causative gene.[2] Onset of symptoms may occur at almost any age, with the exception of infancy, which would indicate the presence of a congenital muscular dystrophy. Traditional approaches of identifying pathogenic mutations by immunohistochemistry, western blotting and Sanger sequencing of selected genes can yield genetic diagnoses in 35% of families.[3] Clinical exome sequencing in general has been reported to have a diagnostic rate of 25%,[4] whereas recent studies of exome sequencing for neuromuscular disease show a 46% diagnostic rate in the United States[5] and 73% in a highly consanguineous population from Iran.[6] Diagnostic rates in LGMD have recently been reported to be 45% in Australia using exome sequencing[7] and 33% in Germany using targeted sequence capture.[8]

[1]Division of Pediatric Neurology, Department of Pediatrics, University of Florida College of Medicine, Gainesville, FL, USA; [2]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; [3]Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA; [4]Division of Genetics and Genomics, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; [5]Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; [6]Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [7]Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; [8]Research Connection and Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; [9]Department of Pediatrics, Division of Neurology, College of Medicine and King Khalid University Hospital, King Saud University, Riyadh, Kingdom of Saudi Arabia; [10]Department of Neurology and Department of Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville, FL, USA and [11]Genetics Institute, University of Florida, Gainesville, FL, USA
[12]Current address: Department of Microbiology, School of Medicine, Ewha Womans University, Seoul 07985, Republic of Korea.
[13]Current address: Biomedical Informatics Laboratory, Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan.
Correspondence: Dr PB Kang, Division of Pediatric Neurology, Department of Pediatrics, University of Florida College of Medicine, PO Box 100296, Gainesville, FL 32610, USA.
E-mail: pbkang@ufl.edu
Received 16 April 2016; revised 31 August 2016; accepted 2 September 2016; published online 6 October 2016

The results of exome sequencing in LGMD for a large cohort from the United States has not previously been published.

We analyzed 55 families from the United States, each of which has one or more individuals with the clinical diagnosis of LGMD. Pathogenic mutations were identified in 22 of 55 families using exome sequence analysis in concert with clinical findings and Sanger sequence confirmation. Our results correlate with the results of studies performed in other countries, and yield interesting observations about approaches to genetic diagnosis in muscular dystrophy.

MATERIALS AND METHODS

Recruitment of families and sample collection

Patients with the clinical diagnosis of LGMD who did not have a genetic diagnosis after clinical evaluation (including some clinical genetic testing), as well as their available informative family members, were recruited for this study. Onset of symptoms for all probands was over 1 year. A total of 55 families were enrolled via an institutionally approved research protocol at Boston Children's Hospital. One of the authors (EE), a certified genetic counselor, personally enrolled most of the subjects and reviewed risks and benefits in detail during the consent process. Clinical data collected included medical and family histories, physical examinations, laboratory results, clinical genetic test results and clinical muscle biopsy data, which were stored in a secure Filemaker Pro v.10 database (see Supplementary Figure 1 for sample form). Peripheral blood or saliva samples were collected from probands and informative relatives for DNA extraction. Any clinical information that indicated specific gene candidates, such as deficiencies of protein expression on immunohistochemistry, was taken into account when analyzing the exome sequencing data.

Whole-exome sequencing

The Genomics Platform at the Broad Institute was used to perform whole-exome sequencing of DNA samples representing selected subjects from 45 of the 55 families; the full sequencing protocol has been published for LGMD cohorts from other countries.[7,9] The Agilent Sure-Select Human All Exon v.2.0, 44 Mb baited target and the Broad in-solution hybrid selection process were used to target exons in genomic DNA. At least 250 ng of DNA with concentrations of at least 2 ng μl^{-1} were submitted for each sample. The hybrid selection libraries cover $>80\%$ of targets at 20x or more, with a mean target coverage of >80x. Exome sequencing data was processed through a pipeline based on Picard (https://github.com/broadinstitute/picard), using base quality score recalibration and local realignment at known insertions and deletions. The Burrows–Wheeler Aligner (https://github.com/lh3/bwa) mapped reads to the human genome build 37 (hg19) reference sequence. The variant call set was uploaded on to xBrowse (https://atgu.mgh.harvard.edu/xbrowse/) and an analysis limited to the candidate gene list was performed using the various inheritance patterns. The main report contains variants restricted to nonsense, frameshift, essential splice site and missense variants and filtered on variant site and genotype quality.

DNA samples from the remaining 10 of the 55 families underwent whole-exome sequencing at the Genomic Diagnostic Laboratory and analyzed by the Interpretive Genomic Services team at Boston Children's Hospital as described previously.[10] Briefly, exome capture was performed using the Agilent v.4 Human Exome Kit. Library sequencing was performed on an Illumina HiSeq (Illumina Inc., San Diego, CA, USA), generating 31 million paired-end reads (100 bpX2) and a mean target coverage of 27×, with 81% of the target covered by $\geqslant 10$ reads. Alignment, variant calling and annotation were performed with a custom informatics pipeline using Burrows–Wheeler Aligner, Picard (http://picard.sourceforge.net), Genome Analysis Toolkit and ANNOVAR. The human genome reference used for these studies was hg19/GRCh37. Single-nucleotide changes, microdeletions and microinsertions were reported and annotated using the NCBI and UCSC reference sequences and online genome databases (NHLBI Exome Sequencing Project with ~5400 exomes, 1000 Genomes Project, dbSNP135, Complete Genomics 52).

A total of 30 exomes were sequenced from the 22 diagnosed families. Seventeen families had only proband samples available for sequencing. Trios (proband and parents) underwent exome sequencing in three families, whereas the proband and an additional informative family member were sequenced in each of the remaining two families. As the exome sequencing was performed on a research basis, incidental findings of pathogenic mutations for unrelated diseases were not systematically sought, identified or reported.

In silico analysis

The candidate variants were identified by xBrowse (Broad Institute of MIT and Harvard, Cambridge, MA, USA) and other software. The 1000 Genomes Project (http://www.1000genomes.org) and the Exome Aggregation Consortium (ExAC) databases (http://exac.broadinstitute.org) were used to determine if the candidate variants were known single-nucleotide polymorphisms (SNPs). Candidate variants that were known SNPs were required to have a minor allele frequency <0.0001 to be considered for further analysis. SNPs with a minor allele frequency >0.0001 were determined to be non-pathogenic. The UCSC browser (https://genome.ucsc.edu/) was used to determine candidate variant amino-acid conservation among species through evolution from lamprey to humans. Species conservation was determined using the likelihood ratio test of significantly conserved amino-acid positions (LRT) and PhyloP (http://ccg.vital-it.ch/mga/hg19/phylop/phylop.html). Pathogenicity of these variants was predicted by using SIFT (http://sift.jcvi.org), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2), Mutation Taster (http://mutationtaster.org) and FATHMM (http://fathmm.biocomputer.org.uk). Variants affecting conserved amino acids that were reported to be pathogenic by at least two of the four prediction programs were selected for further analysis. In light of the limitations on the accuracy of these programs,[11] outputs from these prediction algorithms were used only for screening purposes with a deliberately liberal threshold, and were not used to make final determinations of pathogenicity.

Sanger sequencing confirmation

PCR amplification of selected candidate variants from exome sequence analysis was performed using standard PCR primers. Amplicons were assessed via agarose gel electrophoresis, then purified by treating 5 μl of PCR product with 2 μl of Exonuclease and Shrimp Alkaline Phosphatase (Exo-SAP-IT; Affymetrix) and submitted to the Molecular Genetics Core Facility at Boston Children's Hospital or the Interdisciplinary Center for Biotechnology Research at the University of Florida for sequencing using the ABI Prism BigDye Terminator cycle sequencing protocols (Applied Biosystems, Perkin-Elmer Corp., Foster City, CA, USA). Sequence data were generated in an ABI Prism 3130 or 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA), formatted by ABI Sequencing Analysis software v.5.2 and KB Basecaller, and analyzed using Sequencher v.5.2.3 or earlier versions (GeneCodes Corporation, Ann Arbor, MI, USA). Sanger sequencing was performed in affected family members and other informative family members to confirm pathogenic mutations and track cosegregation patterns. The only widespread screening performed via Sanger sequencing was for *FKRP* in 18 families who had exome sequencing on an older platform that did not have good coverage of that gene.[7]

RESULTS

Overview

Clinical features and details of clinical diagnostic testing are summarized in Table 1. Most of the probands had clinical muscle biopsies, and none of the muscle biopsies led to a genetic diagnosis before enrollment. Analysis of exome sequencing data yielded the identification of pathogenic mutations in 21 of the 55 families, with one additional family among the 55 receiving a clinical genetic diagnosis during the course of the study (Figure 1 and Table 2). The 22 families with diagnoses included 11 with dominant mutations, 10 with recessive mutations and 1 with an X-linked *DMD* mutation. Novel pathogenic mutations were identified in eight families; four of these novel mutations were heterozygous mutations. Two other families have pathogenic mutations reported in public databases, including LOVD (http://www.lovd.nl), Emory Genetics Laboratory (http://geneticslab.emory.edu/emvclass/emvclass.php) and GeneDx

Table 1 Clinical features of the 22 families with genetic diagnoses

Family	Mutation	Inheritance pattern	Age at onset (years)	Ethnicity/race	CK levels	Genetic testing and IHC	Muscle histology	Cardiac/ respiratory complications	Ambulatory status
930	*LMNA* missense	Dominant	3	Irish/English/Italian/ Caucasian	2700	Genetic: *CAPN3*, *CAV3*, *FKRP*, *DMD*, *SMN1* IHC: Dystrophin, sarcoglycans, emerin, merosin	Dystrophic	Arrythmia/ implanted defibrillator	Ambulates with difficulty
1125	*LMNA* missense	Sporadic, 1 affected ♂	2	Caucasian	2700	Genetic: *DMD* deletion/ duplication, *SMN1* IHC: Dystrophin, merosin, sarcoglycans, desmin	Dystrophic	ECG normal	Unknown
965	*COL6A3* missense	Sporadic	15	Irish/English/Dutch/ Caucasian	300	Genetic: *CAV3*, *FKRP*, *DMD* Athena LGMD panel negative	Myopathic	N/A	Ambulatory
1093	*COL6A3* essential splice site	Sporadic, 1 affected ♀	2	Belish/Hungarian/ Caucasian	300	IHC: Dystrophin, merosin	Myopathic	None	Non-ambulatory —hip dysplasia at birth
1115	*COL6A3* essential splice site	Sporadic, 1 affected ♀	2	Polish/Slovakian/Irish/ German/Caucasian	900	Genetic: *CAPN3*, *CAV3*, *FKRP*	Dystrophic	N/A	Ambulatory
1092	*COL6A1* missense	Sporadic, 1 affected ♀	5	Irish/Canadian/ African American/ Italian	300	Genetic: *CAV3*, *FKRP* IHC: Dystrophin, sarcoglycans, merosin	Dystrophic	None	Non-ambulatory
596	*RYR1* missense	Sporadic, 1 affected ♂	3	Italian/Caucasian	30–50	Genetic: Athena LGMD panel negative IHC: Dystrophin, sarcoglycans, merosin	Myopathic	None	Non-ambulatory
1250	*VCP* missense	Dominant, 3 generations affected	42	German/Caucasian	400	IHC: Dystrophin, dysferlin	Dystrophic	N/A	Ambulatory
1399	*FLNC* missense	Sporadic, 1 affected ♂	5	Italian/French/Irish/ Scottish/English/ Caucasian	120–150	Genetic: *DES* and 13 cardiomyopathy genes, myofibrillar myopathy/ cardiomyopathy	N/A	Transplant for cardiomyopathy	Ambulatory
1090	*SMCHD1* in-frame deletion of 3 bp	Dominant	12	Caucasian/Polish	600	Genetic: *FSHD1*, *CAV3*, *LMNA*	Dystrophic	None	Non-ambulatory
1258	*D4Z4* deletion at 4q35	Sporadic, 1 affected ♂	22	German/English/ Russian/Czech/ Caucasian	1000	Genetic: 4q35 deletion confirmed after enrollment	Dystrophic	N/A	Ambulatory
1197	*CAPN3* compound heterozygous missense	Affected brother and sister	14	Italian/French/Irish/ Scottish/English/ Caucasian	Markedly elevated	IHC: Dysferlin	Dystrophic	N/A	Non-ambulatory
1365	*CAPN3* compound heterozygous affecting splicing	Affected brother and sister	23	Fr. Canadian/German/ Native American/ Irish/Caucasian	1400	Genetic: *FSHD1* IHC: Dystrophin, sarcoglycans, caveolin 3, calpain 3	Dystrophic	N/A	Ambulatory
1049	*SGCG* compound heterozygous affecting splicing	Affected brother, sister and half-brother	5	African American/ Black	8000–18 000	Genetic: *DMD*, sarcoglycans, *FKRP*, *DM1*, *DM2*	Dystrophic in affected brother	N/A	Unknown
1118	*SGCG* homozygous missense		11	Puerto Rican/ Hispanic	3000–4000	IHC: Dystroglycan	Dystrophic	Dilated cardiomyopathy	Non-ambulatory
1299	*SGCA* homozygous missense	Consanguineous, 2 affected ♂s	44	Pakistani/Asian	18 000	Genetic: *DMD*, *CAV3*, *BSG* IHC: Alpha-sarcoglycan	Dystrophic	N/A	Ambulatory
1102	*ANO5* homozygous frameshift	Sporadic, 1 affected ♂	34	Caucasian	2000–6000	IHC: Dystrophin, sarcoglycans, merosin	Dystrophic	Arrhythmia/ ablation	Ambulatory
1105	*ANO5* heterozygous frameshift and heterozygous essential splice site	Affected brother and sister	18	Swedish/Irish/ Norwegian/Native American/Italian/ Caucasian	1000–2000	LGMDs, mitochondrial disease, DMD	Dystrophic	None	Ambulatory
1395	*ANO5* heterozygous frameshift and heterozygous nonsense	2 affected brothers and an affected paternal cousin	40	Yugoslavian/German/ Scandinavian/Italian/ Caucasian	1000–2000	Genetic: 13 *LGMDs*/*DMD*	Dystrophic	None	Ambulatory
1117	*GAA* heterozygous essential splice site heterozygous missense	Sporadic, 1 affected ♂	35	Welsh/English/ Caucasian	1000–2000	IHC: Dystrophin, sarcoglycans, merosin	Dystrophic (no increased glycogen observed)	Atrial fibrillation	Ambulatory
1409	*LAMA2* heterozygous nonsense and heterozygous essential splice	Sporadic, 1 affected ♀	2	NE European/ Caucasian	30 000	Genetic: *EMD*, *LMNA*, *FHL1* (has elbow, knee and neck contractures)	N/A	None	Non-ambulatory
1107	*DMD* hemizygous nonsense	Sporadic, 1 affected ♂	8	Caucasian	600	IHC: Dystrophin, sarcoglycans, merosin	Dystrophic	None	Ambulatory

Abbreviations: CK, creatine kinase; ECG, electrocardiography; IHC, immunohistochemistry; N/A, not available.

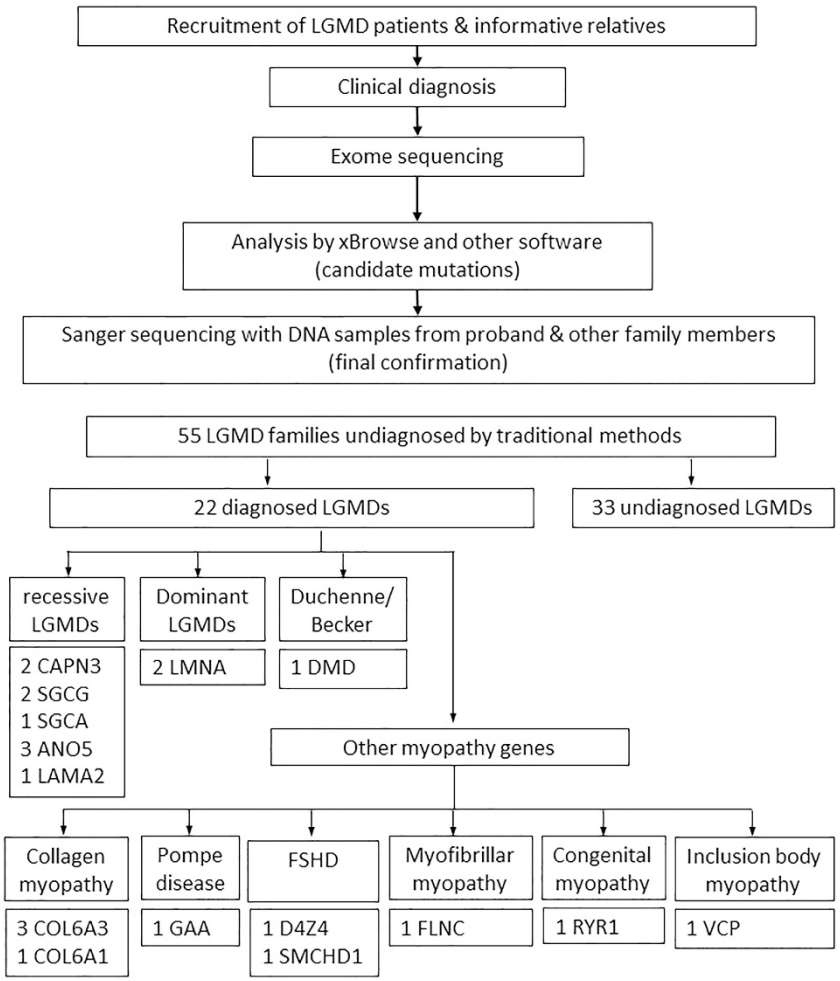

Figure 1 Flow chart of analytic process for the cohort of families with LGMD (top), along with a breakdown of genetic diagnoses by category (bottom).

but not published; one pathogenic mutation was in both categories. Sanger sequencing confirmed these mutations in all probands and also confirmed expected cosegregation patterns for available family members. Cosegregation was confirmed in 13 of the 22 families, whereas the remaining nine had only proband DNA samples available. The families with only proband samples available included 7 with previously reported pathogenic mutations and 2 with novel pathogenic mutations (one family had compound heterozygous pathogenic mutations that included a previously reported nonsense mutation and a novel essential splice site mutation). No *FKRP* mtuations were found on Sanger sequencing.

Autosomal dominant *LMNA* mutations

Two unrelated individuals representing families 930 and 1125 were found to have LGMD1B with pathogenic mutations in *LMNA*. Both affected individuals had onset in the toddler years, elevated serum creatine kinase levels and dystrophic muscle biopsies.

Autosomal dominant *COL6A1* mutations

Family 1092 was found to have novel dominant missense pathogenic mutations in *COL6A1*. This gene is classically associated with Bethlem myopathy and Ullrich congenital muscular dystrophy, but recent reports also link it with LGMD.[12,13] The *COL6A1* NM_001848.2 c.868G>A, NP_001839.2 p.Gly290Arg (rs121912939) pathogenic mutation in family 1092 has been reported by GeneDx (http://www.genedx.com/test-catalog/disorders/limb-girdle-muscular-dystrophy-lgmd/, with NCBI submission accession number: SCV000196773.1) and Emory Genetics Laboratory (http://geneticslab.emory.edu/index.html, with NCBI submission accession numbers: SCV000224895.1, SCV000224896.1 and SCV000111716.3) as being pathogenic. A dominant missense pathogenic mutation c.868G>C that causes the identical p.Gly290Arg amino-acid substitution has been reported in Ullrich congenital muscular dystrophy.[14]

Autosomal dominant *COL6A3* mutations

Similarly, pathogenic mutations in *COL6A3*[15] are known to cause Ullrich congenital muscular dystrophy and Bethlem myopathy, but the association with LGMD has only been reported recently.[7] We identified pathogenic mutations in *COL6A3* in three families. A dominant mutation in *COL6A3* (NP_004360.2, p.Glu1386Lys) identified in family 965 was previously reported as being pathogenic[15] and that amino-acid residue was highly conserved across species. Proband 965-1 had neither distal laxity nor a tendency towards keloid formation, and a thigh magnetic resonance imaging did not show findings specific for Ullrich congenital muscular dystrophy or Bethlem myopathy.[16] The pathogenic mutations of *COL6A3* identified in two other families (1093 and 1115) are *de novo* essential splice site mutations. The pathogenic *COL6A3* mutation (NM_004369.3, c.6283-1C>T) in family 1093 is novel, whereas the NM_004369.3, c.6156+1G>A *de novo* pathogenic mutation observed

in 1115 was reported previously.[15] The proband in family 1093 showed a mixed phenotype of LGMD and congenital muscular dystrophy.

Autosomal dominant *RYR1* mutations

RYR1 (ryanodine receptor 1) mutations are known to cause a congenital myopathy, central core disease. A *de novo* dominant missense pathogenic mutation (NM_001042723.1, c.14567G>A, NP_001036188.1 p.Arg4856His) in *RYR1* was found in the proband of family 596. This mutation has been reported to cause a congenital neuromuscular disease with uniform type 1 fibers and an association with central core disease.[17,18]

Autosomal dominant *VCP* mutations

Pathogenic *VCP* mutations are known to cause amyotrophic lateral sclerosis and inclusion body myopathy. The mutation identified in family 1250 (*VCP* NM_007126.3, c.572G>A, NP_009057.1, p.Arg191Gln (rs121909334)) was previously reported in familial amyotrophic lateral sclerosis and in patients with an unusual syndrome of inclusion body myopathy, Paget disease of bone and frontotemporal dementia.[19] The inclusion body myopathy may present with manifestations similar to LGMD.[19]

Autosomal dominant *FLNC* mutations

Pathogenic mutations in gamma filamin (*FLNC*) usually cause myofibrillar myopathy with distal weakness, but a recent report showed that they may cause an LGMD phenotype.[7] The dominant missense pathogenic mutation *FLNC* NM_001458.4, c.7409C>A, NP_001449.3, p.Pro2470His identified in 1399 is novel, has not been reported in any population database and was predicted to be pathogenic by SIFT, PolyPhen, MutTaster and FATHMM. The proband of family 1399 showed an LGMD phenotype with cardiomyopathy, accompanied by features of myofibrillar myopathy, similar to other individuals reported to have pathogenic *FLNC* mutations.

Autosomal dominant FSHD

The dominant pathogenic mutation in *SMCHD1*, identified in family 1090, causes an in-frame deletion of amino-acid lysine at position 275 and has been reported previously.[10] While sequence data were being analyzed, the proband from 1258 informed the research team that he had been diagnosed with FSHD1 based on clinical genetic testing of the D4Z4 region on chromosome 4q35. He had asymmetric weakness in the right chest and arm, but no facial weakness.

Autosomal recessive *CAPN3* mutations

Compound heterozygous pathogenic mutations in *CAPN3* were identified in families 1197 and 1365. The missense mutations found in family 1197 were previously reported as homozygous mutations in different families.[20,21] Both heterozygous pathogenic mutations of *CAPN3* found in family 1365 affect splicing, and a western blot of protein extracted from a muscle tissue biopsy showed reduced calpain 3 expression. The *CAPN3* NM_000070.2, c.1746-20C>G (rs201892814) pathogenic mutation was reported previously by the Emory Genetics Laboratory (http://www.ncbi.nlm.nih.gov/clinvar/variation/92408/, with NCBI submission accession number: SCV000109927.4), and c.945+5G>A is a novel pathogenic mutation that shifts a splice site downstream, extending the exon. The latter was found to have a minor allele frequency of 0.0000082 (i.e., singleton) in the ExAC database.

Autosomal recessive sarcoglycan mutations

A consanguineous family, 1299, had a pathogenic homozygous recessive missense mutation (NM_000023.2, c.109G>T, NP_000014.1, p.Val37Leu) in *SGCA*; this mutation has not been previously reported. Two pathogenic mutations in *SGCG*, a previously reported heterozygous deletion of four nucleotides (AGTA) at NM_000231.2, c.195+4_195+7[22] and a novel heterozygous substitution of c.195+1G>C (rs200502077), were found in family 1049. The latter is an essential splice site mutation. A muscle biopsy was performed on the proband, but tissue from this biopsy was not available for the current study.

Autosomal recessive *ANO5* mutations

Pathogenic mutations in *ANO5*, which cause LGMD2L, were found in three families (1102, 1105 and 1395). The homozygous recessive mutation found in family 1102, *ANO5* NM_213599.2, c.191dupA, NP_998764.1, p.Asn64Lys fs Ter15 (rs137854521), is a known pathogenic mutation[23–25] that generates a stop codon 15 amino-acid residues downstream of the mutation. The two other families (1105 and 1395) also have this mutation but in a heterozygous state; the other allele has novel mutations: a nonsense mutation c.835C>T, p.Arg279Ter in family 1395 and a splicing mutation c.2235+5G>A in family 1105. The pathogenic *ANO5* mutations were confirmed for cosegregation in their respective families.

Autosomal recessive *LAMA2* mutations

Pathogenic mutations in *LAMA2* have been identified as the cause of merosin-deficient congenital muscular dystrophy. Several studies have reported that partial merosin deficiency by *LAMA2* mutations and some forms of *LAMA2* mutations are known to manifest as LGMD phenotypes,[26–30] suggesting that *LAMA2* should be included among the causative genes for LGMD2.[31] Compound heterozygous pathogenic mutations in *LAMA2*, a previously reported nonsense mutation NM_000426.3, c.5116C>T, NP_000417.2, p.Arg1706Ter[28] and a novel splice site mutation c.8703+1G>A r.spl, were identified in family 1409. The phenotype of the proband, 1409-1, was reviewed again and was confirmed to meet the criteria for LGMD. The proband had some contractures and onset was in early childhood but was not early enough to be classified as congenital muscular dystrophy. Mutations in LAMA2 have recently been associated with Emery–Dreifuss muscular dystrophy[32] and this diagnosis has also been a consideration for the proband. However, the subject was a young adult at the most recent evaluation and ongoing cardiac monitoring has revealed little to no evidence for overt cardiac complications to date.

Autosomal recessive *GAA* mutations

Compound heterozygous pathogenic mutations in *GAA*, known to cause Pompe disease, were found in family 1117. These were a missense mutation NM_000152.3, c.1841C>A, NP_000143.2, p.Thr614Lys (rs369531647)[33] and a substitution c.-32-13T>G r.spl (rs386834236) that affects splicing.[34] Both mutations were reported previously.

X-linked mutations

One family was found to have an X-linked pathogenic mutation in the dystrophin gene (*DMD*). The pathogenic nonsense mutation *DMD* NM_004006.2, c.9G>A, NP_003997.1, p.Trp3Ter, found in family 1107, was reported previously.[35–37]

Table 2 Genetic analysis for the 22 families with genetic diagnoses

Family	Gene/ locus	Mutation	Inheritance	References (SNP)/novel	Exome sequencing	Sanger sequencing	Segregation of mutation
930	*LMNA*	NM_170707.3 c.832G>C, NP_733821.1 p.Ala278Pro hg19 chr1:156104999	Dominant	Kuhn *et al.*[8]	Proband (♂)	Proband	N/A
1125	*LMNA*	NM_170707.3 c.746G>A, NP_733821.1 p.Arg249Gln hg19 chr1:156104702	*De novo* dominant	Reddy *et al.*[9] (rs59332535)	Proband (♂)	Trio	Present in proband, absent in both parents
965	*COL6A3*	NM_004369.3 c.4156G>A, NP_004360.2 p.Glu1386Lys hg19 chr2:238280504	Dominant	Walters-Sen *et al.*[11]	Proband (♂)	Proband	N/A
1093	*COL6A3*	NM_004369.3 c.6283-1C>T hg19 chr2:238268032	*De novo* dominant, essential splice site	Novel	Proband (♀) and parents	Trio	Present in proband, absent in both parents
1115	*COL6A3*	NM_004369.3 c.6156+1G>A hg19 chr2:238270381	*De novo* dominant, essential splice site	Walters-Sen *et al.*[11]	Proband (♀) and parents	Trio	Present in proband, absent in both parents
1092	*COL6A1*	NM_001848.2 c.868G>A, NP_001839.2 p.Gly290Arg hg19 chr21:47409531	*De novo* dominant heterozygous	Novel (rs121912939) (GeneDx, Emory Genetics Laboratory)	Proband (♀)	Trio	Present in proband, absent in both parents
596	*RYR1*	NM_001042723.1 c.14567G>A, NP_001036188.1 p.Arg4856His hg19 chr19:39071080	*De novo* dominant	Tan *et al.*[12] and Brown *et al.*[13] (rs63749869)	Proband (–) and parents	Trio	Present in proband, absent in both parents
1250	*VCP*	NM_007126.3 c.572G>A, NP_009057.1 p.Arg191Gln hg19 chr9:35065252	Dominant heterozygous	Giusti *et al.*[14] (rs121909334)	Proband (♂)	Proband and affected brother (1250-4)	Present in both proband & affected brother
1399	*FLNC*	NM_001458.4 c.7409C>A, NP_001449.3 p.Pro2470His hg19 chr7:128496823	Dominant heterozygous	Novel	Proband (♂)	Proband	N/A
1090	*SMCHD1*	NM_015295.2 c.818_820del, NP_056110.2 p.Lys275del hg19 chr18:2688690_2688692del	Dominant	Ghaoui *et al.*[7]	Proband (♂), father	Proband, parents, brother (1–9)	
1258	*D4Z4*	Deletion of D4Z4 repeats at 4q35	Heterozygous		Proband (♂)	Proband	N/A
1197	*CAPN3*	NM_000070.2 c.1465C>T, NP_000061.1 p.Arg489Trp hg19 chr15:42693949 and c.1715G>A, p.Arg572Gln hg19 chr15:42695170	Compound heterozygous	Lampe *et al.*[15] and Mercuri *et al.*[16] as two separate mutations (rs121434544)	Proband (♂)	Proband and affected sister	Present in both proband and affected sister
1365	*CAPN3*	NM_000070.2 c.945+5G>A hg19 chr15:42682299 and c.1746-20C>G hg19 chr15:42695919	Compound heterozygous affecting splicing	Novel rs201892814 (Emory Genetics Laboratory)	Proband (♀)	Proband, unaffected father and affected brother	Present only in proband and affected brother absent in unaffected father present in all three samples
1049	*SGCG*	NM_000231.2 c.195+4_195+7 del CAGTA>C hg19 chr13:23778027 and c.195+1G>C hg19 chr13:23778029	Compound heterozygous affecting splicing	Sato *et al.*[17] novel (rs200502077)	Proband (♀), affected brother	Proband and affected brother	Present in both proband and affected brother
1118	*SGCG*	NM_000231.2 c.787G>A, NP_000222.1 p.Glu263Lys hg19 chr13:23898591	Recessive homozygous	Almomani *et al.*[38] (rs104894423)	Proband (♀)	Proband	N/A
1299	*SGCA*	NM_000023.2 c.109G>T, NP_000014.1 p.Val37Leu hg19 chr17:48244800	Consanguineous recessive homozygous	Novel	Proband (♂)	Proband, parents, affected brother (–5) and unaffected sister (–4)	Homozygous in proband and affected brother and heterozygous in parents and unaffected sister
1102	*ANO5*	NM_213599.2 c.191dupA, NP_998764.1 p.Asn64Lys fs Ter15 hg19 chr11:22242653	Recessive homozygous	Tilgen *et al.*,[18] Watts *et al.*[19] and Richard *et al.*[20] (rs137854521)	Proband (♂)	Proband	N/A
1105	*ANO5*	NM_213599.2 c.191dupA, NP_998764.1 p.Asn64Lys fs Ter15 hg19 chr11:22242653 and c.2235+5G>A hg19 chr:22294540	Compound heterozygous	Tilgen *et al.*,[18] Watts *et al.*[19] and Richard *et al.*[20] (rs137854521) novel	Proband (♂)	1: proband; 2–3: parents; 6: affected sister; 4, 5 and 7: unaffected siblings	Heterozygous in 1, 3, 5 and 6; absent in 2, 4 and 7 Heterozygous in 1, 2 and 6; absent in 3, 4 and 7

Table 2 (Continued)

Family	Gene/ locus	Mutation	Inheritance	References (SNP)/novel	Exome sequencing	Sanger sequencing	Segregation of mutation
1395	*ANO5*	NM_213599.2 c.191dupA, NP_998764.1 p.Asn64Lys fs Ter15 hg19 chr11:22242653 and c.835C>T, p.Arg279Ter hg19 chr11:22261187	Compound heterozygous	Tilgen *et al.*,[18] Watts *et al.*[19] and Richard *et al.*[20] (rs137854521) novel	Proband (♂)	1, 2 and 4	
1117	*GAA*	NM_000152.3 c.-32-13T>G r.spl hg19 chr17:78078341 and NM_000152.3 c.1841C>A, NP_000143.2 p.Thr614Lys hg19 chr17:78086463	Compound heterozygous	Geranmayeh *et al.*[28] (rs386834236) Naom *et al.*[27] (rs369531647)	Proband (♂)	Proband	N/A
1409	*LAMA2*	NM_000426.3 c.5116C>T, NP_000417.2 p.Arg1706Ter hg19 chr6:129712680 and c.8703+1G>A r.spl hg19 chr6:129826501	Compound heterozygous	Bolduc *et al.*[23] novel	Proband (♀)	Proband	N/A
1107	*DMD*	NM_004006.2 c.9G>A, NP_003997.1 p.Trp3Ter hg19 chrx:33229421	X-linked nonsense	Gavassini *et al.*,[29] Marques *et al.*[30] and Lokken *et al.*[31]	Proband (♂)	Proband	N/A

Abbreviations: N/A, not applicable; SNP, single-nucleotide polymorphism.

Table 3 Summary of three additional families with possible pathogenic mutations

Family	Gene	Possible mutation	Inheritance	Exome sequencing	Sanger sequencing
1027	*MYOT*	NM_006790.2 c.1345delC, NP_006781.1 p.Pro449Gln fs Ter16 (rs780331457) hg19 chr5:137222922	Heterozygous	Proband (♀)	Proband
1255	*POMGNT2*	NM_032806.5 c.190G>A, NP_116195.2 p.Gly64Ser (rs548769646) hg19 chr3:43122734 c.740_741delAA, p.Phe247CysfsTer16 hg19 chr3:43122182	Heterozygous, also found in both parents *de novo* heterozygous	Proband (♂)	Proband and parents
1366	*COL6A1*	NM_001848.2 c.466G>T, NP_001839.2 p.Val156Leu hg19 chr21:47406477	Dominant heterozygous	Proband (♂)	Proband

Possible mutations

Suspected but unconfirmed mutations are listed in Table 3. Exome sequencing analysis showed that family 1027 has a heterozygous dominant variant in *MYOT* (NM_006790.2 c.1345delC, NP_006781.1 p.Pro449Gln fs Ter16 (rs780331457). Mutations in *MYOT* are known to cause LGMD1A, but DNA is only available on the proband for this family, hence it is difficult to confirm this variant as a pathogenic mutation. It is a novel variant that is not found in the 1000 Genomes database and with minor allele frequency of 0.00004942 in the ExAC database. The amino-acid residue is also very well conserved. We found compound heterozygous variants of *POMGNT2* (*GTDC2*) in family 1255. A rare missense variant (NM_032806.5 c.190G>A, NP_116195.2 p.Gly64Ser (rs548769646)) is found in the proband as well as both parents, whereas a 2 base pair deletion (c.740_741delAA, p.Phe247CysfsTer16) is present in the proband and absent in both parents; the latter appears more likely to be pathogenic. The missense variant of *COL6A1* found in family 1366 is novel (NM_001848.2 c.466G>T, NP_001839.2 p.Val156Leu), and the affected amino-acid residue is conserved from lamprey through human. The mutation in family 1366 is not found in the 1000 Genomes database and has a very low minor allele frequency of 0.0000085 in the ExAC database (http://exac.broadinstitute.org/). It is predicted to be pathogenic by three of four prediction programs analyzed. The phenotype of the proband in family 1366 showed some overlap with congenital muscular dystrophy. DNA was only available for the proband in this family, thus analysis of cosegregation patterns was not possible.

A recurrent *DMD* variant confirmed to be benign

A *DMD* NM_004006.2 c.8762A>G, NP_003997.1 p.His2921Arg (rs1800279) variant suspected of being benign[37–40] was identified in the probands of four families (1258, 1309, 1365 and 1398). In each family, the variant was confirmed to be benign due to causative mutations found in other genes (Table 4). Two of the families, 1309 and 1398, were from Saudi Arabia and were not included in the 55 families for the main analysis noted above, but are mentioned here as further evidence of the benign nature of this variant. The proband of family 1258 is male, but his muscle biopsy showed normal dystrophin staining and he was diagnosed with FSHD1, as noted above. Both the male proband and an unaffected brother in family 1309 had the hemizygous *DMD* variant in question. The female proband of family 1365 had confirmed *CAPN3* compound heterozygous missense mutations as well as the heterozygous *DMD* mutation. Family 1398 was found to have a known homozygous *SGCG* NM_000231.2, c.212T>C, XP_005266562.1, p.Leu71Ser mutation that cosegregates with phenotype in this family. The proband of this family is female and was found to have the heterozygous *DMD* variant, whereas the unaffected father was found to have the hemizygous *DMD* variant. The minor allele frequency for

Table 4 Summary of four families with the benign p.H2921R *DMD* variant

Family	*Ethnicity*	*Pathogenic mutation*	*DMD p.H2921R variant*
1258	USA	FSHD1	Present in proband (♂)
1365	USA	CAPN3 NM_000070.2 c.945+5G>A and c.1746-20C>G	Present in proband (♀) (heterozygous)
1309	Saudi Arabian	Unknown	Present in proband (♂) and unaffected brother (1309-5)
1398	Saudi Arabian	*SGCG* NM_000231.2 c.212T>C, NP_000222.1 p.Leu71Ser homozygous	Present in proband (♀) (heterozygous), unaffected father (1398-3), and unaffected sister (heterozygous)

this variant (rs1800279) in the ExAC database is 0.02629, which is not compatible with a pathogenic mutation.

DISCUSSION

Among the 55 families studied, exome sequencing analysis identified pathogenic mutations in 21, whereas clinical genetic testing revealed the diagnosis for an additional family. The overall success rate of 40% is comparable to recent previous reports of exome sequencing analysis for LGMD and neuromuscular diseases in non-consanguineous populations.[5,7,8] Traditional genetic, biochemical and histopathological examinations yield diagnoses in 30–40% of LGMD cases,[3,41] and targeted sequence capture has similar yields.[8] Exome sequencing has improved the diagnostic yield to the 40–45% range, both in our cohort and in the literature,[5,7,8] likely due in part to the use of trios and family studies. As the subjects had varying degrees of clinical evaluation before enrollment, including clinical genetic testing, a similar approach would be expected to have an even higher yield in the clinical setting for patients who had not had prior genetic testing or were screened appropriately for pathogenic mutations not amenable to sequencing technologies. Several families had pathogenic mutations in *CAPN3*, sarcoglycans and *ANO5*, common LGMD genes for which clinical genetic testing is readily available. The absence of any subjects with pathogenic *DYSF* mutations is notable, as well as the under-representation of common genes aside from *ANO5*. The depth of clinical evaluations varied among these families. Many patients with pathogenic mutations in common LGMD genes were likely diagnosed on clinical genetic testing and this cohort does not represent those individuals. Most of the subjects who had extensive LGMD genetic testing before enrollment underwent those evaluations before the association of *ANO5* with LGMD was first described in 2010.

Among the pathogenic mutations identified in our cohort, six were found in loci not traditionally classified as being associated with LGMD (e.g., *DMD*, *GAA*, *SMCHD1*, *VCP*, *FLNC* and the D4Z4 region of 4q35), suggesting that these genes could account for at least some of the increased diagnostic yield, as recently noted.[7] These findings, along with the decreasing use of muscle biopsy in clinical settings, indicate that diagnostic genetic testing panels based on targeted sequence capture for LGMD should include a broad array of muscle disease genes, not only ones that meet the strict definition of LGMD. The diversity of causative genes also illustrates the importance of accurate clinical phenotyping for both clinical and research purposes. There is significant phenotypic overlap between LGMD and diseases that are not traditionally considered to be LGMD, such as Pompe disease, and although the subjects in our cohort with non-LGMD causative genes could not be distinguished from the others based on clinical presentation, there may be other cases where this is possible. Of note, given the availability of a treatment for Pompe disease, the individual with the *GAA* mutations had clinical confirmation in compliance with our IRB protocol so that treatment options could be offered.

This study confirmed that *DMD* NM_004006.2, c.8762A>G, NP_003997.1, p.His2921Arg is a non-pathogenic benign variant, as it was found in multiple unaffected individuals in the hemizygous state, and affected individuals were also found to have confirmed pathogenic mutations in other genes. The variant has been increasingly suspected of being benign.[37–40] The additional findings in our study illustrate one of the benefits of accumulating databases of exome sequences. Although the amount of data is significantly larger, requiring sophisticated computational approaches to analyze completely, the array of identified variants for each individual tested is more complete, which over time will permit more definitive assignments of pathogenicity, fewer 'variants of unknown significance' and correction of reported mutations that may not truly be pathogenic.[42]

These diagnostic outcomes have been consistent across multiple exome sequencing studies performed on disease categories that are genetically heterogeneous, as LGMD is. This suggests that the previous estimate that 85% of pathogenic mutations are found in coding regions[43] may be too high. However, the subjects selected for the current study and similar studies were ones who had previously had clinical evaluations, including genetic testing, suggesting that the yield would be higher had the cohorts not been prescreened. In addition, certain types of pathogenic mutations affecting coding regions are not easily detected with current exome sequencing technologies. For example, single and multiple exon deletions and duplications comprise the majority of pathogenic mutations in Duchenne and Becker muscular dystrophy, trinucleotide repeat expansions cause the most common form of myotonic dystrophy, and the D4Z4 macrosatellite deletion on 4q35 that is associated with facioscapulohumeral muscular dystrophy type 1 is also not easily detected on exome sequencing. A number of our subjects who had phenotypes suggestive of these specific types of muscle disease had appropriate clinical genetic testing, but a patient with an atypical presentation of facioscapulohumeral muscular dystrophy type 1 was enrolled in our research and received a clinical genetic diagnosis of LGMD owing to his phenotype. Careful phenotyping of individuals and family members proves to be very important to help keep the investigator on the proper course to ultimately lead to a molecular diagnosis.

Ethical issues persist in the collection of exome and genome-wide sequencing data with respect to the potential for the identification of incidental pathogenic mutations. These mutations are often hidden in the mountains of data generated, as research laboratories and clinical laboratories typically extract only those variants that lie in a specific, limited set of genes of interest. Incidental variants would only be found if they were actively sought during variant analysis. Another problem is that if some pathogenic mutations may not lead to symptomatic disease for decades, what would be an optimal time to discover and report such mutations. Various national and international organizations are actively discussing this issue. One solution is to provide patients and research subjects access to their electronic sequencing data, so that they may, if they choose, seek additional analysis by other facilities and investigators without having to have the sequencing repeated.

Further analysis continues on the families in whom pathogenic mutations were not identified in the current study. Some of the probands had clinical muscle biopsies performed, and when available, biopsy reports and slides were reviewed to confirm the absence of pathogenic findings. The possibility of digenic compound heterozygous mutations will be considered, as has been described for specific diseases,[44] including muscular dystrophy.[45,46] To extend the current study, we plan to perform whole-genome sequencing and other genetic analyses on selected families in an attempt to detect larger pathogenic mutations such as copy number variants, inversions and large-scale deletions such as the D4Z4 macrosatellite contraction. The rare pathogenic mutation in a noncoding region will be difficult to identify and confirm, even with the assistance of whole-genome sequencing, given the collective size of the intronic regions and the number of variants that will be identified for each affected individual. Exceptions may be found in regions with known functions such as microRNA binding sites, where pathogenic mutations have been confirmed in a handful of cases. And, there is of course the promise that novel disease genes remain to be identified. We are currently examining candidate mutations in several potential novel genes that have been identified on the exome sequencing analysis. Though such genes are becoming more difficult to discover and confirm, it is unlikely that we have identified all the genes associated with LGMD, and the number of cases that remain without a genetic diagnosis provide a tantalizing clue that more such genes are out there.

The current analysis of whole-exome data from a sizeable cohort of families affected by LGMD in the United States has yielded similar overall findings to those reported in other countries. Most of the pathogenic mutations identified were in known LGMD genes, but a few were in muscle disease genes that are not strictly considered to be LGMD, indicating that clinical genetic testing panels should include a broad array of genes to maximize the yield. A previously reported pathogenic mutation in *DMD* was found to be a benign variant in multiple families, providing an example of how candidate mutations in both known and novel disease genes should be scrutinized carefully. The number of cases without a genetic diagnosis remains stubbornly high, even after exome sequencing, suggesting that there are unusual pathogenic mutations in known genes and all manner of pathogenic mutations in novel disease genes that have yet to be identified.

CONFLICT OF INTEREST

TWY and LMK have received personal compensation from Claritas Genomics.

ACKNOWLEDGEMENTS

We thank all the study participants for their valuable contributions to this study. The work was supported by NIH R01 NS080929 (HMR, KAC, EE and PBK), NIH R01 GM104371 (DGM), Department of Pediatrics at the University of Florida College of Medicine (KAC and PBK), Muscular Dystrophy Association Research Grant 186796 (PBK) and the Bernard F and Alva B Gimbel Foundation (LMK). Exome sequencing was supported by Medical Sequencing Program Grant U54HG003067 from the National Human Genome Research Institute. MAS was supported by the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia via research group project number RGP-VPP-301.

1 Mitsuhashi, S. & Kang, P. B. Update on the genetics of limb girdle muscular dystrophy. *Semin. Pediatr. Neurol.* **19**, 211–218 (2012).
2 Nigro, V. & Savarese, M. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myol.* **33**, 1–12 (2014).
3 Moore, S. A., Shilling, C. J., Westra, S., Wall, C., Wicklund, M. P., Stolle, C. *et al.* Limb-girdle muscular dystrophy in the United States. *J. Neuropathol. Exp. Neurol.* **65**, 995–1003 (2006).
4 Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y. *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* **312**, 1870–1879 (2014).
5 Ankala, A., da Silva, C., Gualandi, F., Ferlini, A., Bean, L. J., Collins, C. *et al.* A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Ann. Neurol.* **77**, 206–214 (2015).
6 Fattahi, Z., Kalhor, Z., Fadaee, M., Vazehan, R., Parsimehr, E., Abolhassani, A. *et al.* Improved diagnostic yield of neuromuscular disorders applying clinical exome sequencing in patients arising from a consanguineous population. *Clin. Genet.* (e-pub ahead of print 28 May 2016; doi:10.1111/cge.12810).
7 Ghaoui, R., Cooper, S. T., Lek, M., Jones, K., Corbett, A., Reddel, S. W. *et al.* Use of whole-exome sequencing for diagnosis of limb-girdle muscular dystrophy: outcomes and lessons learned. *JAMA Neurol.* **72**, 1424–1432 (2015).
8 Kuhn, M., Glaser, D., Joshi, P. R., Zierz, S., Wenninger, S., Schoser, B. *et al.* Utility of a next-generation sequencing-based gene panel investigation in German patients with genetically unclassified limb-girdle muscular dystrophy. *J. Neurol.* **263**, 743–750 (2016).
9 Reddy, H. M., Hamed, S. A., Lek, M., Mitsuhashi, S., Estrella, E., Jones, M. D. *et al.* A homozygous nonsense mutation in SGCA is a common cause of LGMD in Assiut, Egypt. *Muscle Nerve* **54**, 690–695 (2016).
10 Mitsuhashi, S., Boyden, S. E., Estrella, E. A., Jones, T. I., Rahimov, F., Yu, T. W. *et al.* Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. *Neuromuscul. Disord.* **23**, 975–980 (2013).
11 Walters-Sen, L. C., Hashimoto, S., Thrush, D. L., Reshmi, S., Gastier-Foster, J. M., Astbury, C. *et al.* Variability in pathogenicity prediction programs: impact on clinical diagnostics. *Mol. Genet. Genomic Med.* **3**, 99–110 (2015).
12 Tan, D., Yang, H., Yuan, Y., Bonnemann, C., Chang, X., Wang, S. *et al.* Phenotype–genotype analysis of Chinese patients with early-onset LMNA-related muscular dystrophy. *PLoS ONE* **10**, e0129699 (2015).
13 Brown, C. A., Lanning, R. W., McKinney, K. Q., Salvino, A. R., Cherniske, E., Crowe, C. A. *et al.* Novel and recurrent mutations in lamin A/C in patients with Emery–Dreifuss muscular dystrophy. *Am. J. Med. Genet.* **102**, 359–367 (2001).
14 Giusti, B., Lucarini, L., Pietroni, V., Lucioli, S., Bandinelli, B., Sabatelli, P. *et al.* Dominant and recessive COL6A1 mutations in Ullrich scleroatonic muscular dystrophy. *Ann. Neurol.* **58**, 400–410 (2005).
15 Lampe, A. K., Dunn, D. M., von Niederhausern, A. C., Hamil, C., Aoyagi, A., Laval, S. H. *et al.* Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J. Med. Genet.* **42**, 108–120 (2005).
16 Mercuri, E., Lampe, A., Allsop, J., Knight, R., Pane, M., Kinali, M. *et al.* Muscle MRI in Ullrich congenital muscular dystrophy and Bethlem myopathy. *Neuromuscul. Disord.* **15**, 303–310 (2005).
17 Sato, I., Wu, S., Ibarra, M. C., Hayashi, Y. K., Fujita, H., Tojo, M. *et al.* Congenital neuromuscular disease with uniform type 1 fiber and RYR1 mutation. *Neurology* **70**, 114–122 (2008).
18 Tilgen, N., Zorzato, F., Halliger-Keller, B., Muntoni, F., Sewry, C., Palmucci, L. M. *et al.* Identification of four novel mutations in the C-terminal membrane spanning domain of the ryanodine receptor 1: association with central core disease and alteration of calcium homeostasis. *Hum. Mol. Genet.* **10**, 2879–2887 (2001).
19 Watts, G. D., Wymer, J., Kovach, M. J., Mehta, S. G., Mumm, S., Darvish, D. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat. Genet.* **36**, 377–381 (2004).
20 Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N. *et al.* Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* **81**, 27–40 (1995).
21 Richard, I., Roudaut, C., Saenz, A., Pogue, R., Grimbergen, J. E., Anderson, L. V. *et al.* Calpainopathy—a survey of mutations and polymorphisms. *Am. J. Hum. Genet.* **64**, 1524–1540 (1999).
22 Bonnemann, C. G., Wong, J., Jones, K. J., Lidov, H. G., Feener, C. A., Shapiro, F. *et al.* Primary gamma-sarcoglycanopathy (LGMD 2C): broadening of the mutational spectrum guided by the immunohistochemical profile. *Neuromuscul. Disord.* **12**, 273–280 (2002).
23 Bolduc, V., Marlow, G., Boycott, K. M., Saleki, K., Inoue, H., Kroon, J. *et al.* Recessive mutations in the putative calcium-activated chloride channel Anoctamin 5 cause proximal LGMD2L and distal MMD3 muscular dystrophies. *Am. J. Hum. Genet.* **86**, 213–221 (2010).
24 Hicks, D., Sarkozy, A., Muelas, N., Koehler, K., Huebner, A., Hudson, G. *et al.* A founder mutation in Anoctamin 5 is a major cause of limb-girdle muscular dystrophy. *Brain* **134**, 171–182 (2011).
25 Penttila, S., Palmio, J., Suominen, T., Raheem, O., Evila, A., Muelas Gomez, N. *et al.* Eight new mutations and the expanding phenotype variability in muscular dystrophy caused by ANO5. *Neurology* **78**, 897–903 (2012).
26 Sewry, C. A., Naom, I., D'Alessandro, M., Sorokin, L., Bruno, S., Wilson, L. A. *et al.* Variable clinical phenotype in merosin-deficient congenital muscular dystrophy associated with differential immunolabelling of two fragments of the laminin alpha 2 chain. *Neuromuscul. Disord.* **7**, 169–175 (1997).
27 Naom, I., D'Alessandro, M., Sewry, C. A., Philpot, J., Manzur, A. Y., Dubowitz, V. *et al.* Laminin alpha 2-chain gene mutations in two siblings presenting with limb-girdle muscular dystrophy. *Neuromuscul. Disord.* **8**, 495–501 (1998).

28 Geranmayeh, F., Clement, E., Feng, L. H., Sewry, C., Pagan, J., Mein, R. *et al.* Genotype–phenotype correlation in a large population of muscular dystrophy patients with LAMA2 mutations. *Neuromuscul. Disord.* **20**, 241–250 (2010).
29 Gavassini, B. F., Carboni, N., Nielsen, J. E., Danielsen, E. R., Thomsen, C., Svenstrup, K. *et al.* Clinical and molecular characterization of limb-girdle muscular dystrophy due to LAMA2 mutations. *Muscle Nerve* **44**, 703–709 (2011).
30 Marques, J., Duarte, S. T., Costa, S., Jacinto, S., Oliveira, J., Oliveira, M. E. *et al.* Atypical phenotype in two patients with LAMA2 mutations. *Neuromuscul. Disord.* **24**, 419–424 (2014).
31 Lokken, N., Born, A. P., Duno, M. & Vissing, J. LAMA2-related myopathy: frequency among congenital and limb-girdle muscular dystrophies. *Muscle Nerve* **52**, 547–553 (2015).
32 Nelson, I., Stojkovic, T., Allamand, V., Leturcq, F., Bécane, H. M., Babuty, D. *et al.* Laminin α2 deficiency-related muscular dystrophy mimicking Emery–Dreifuss and collagen VI related diseases. *J. Neuromuscul. Dis.* **2**, 229–240 (2015).
33 Kroos, M., Pomponio, R. J., van Vliet, L., Palmer, R. E., Phipps, M., Van der Helm, R. *et al.* Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. *Hum. Mutat.* **29**, E13–E26 (2008).
34 Huie, M. L., Chen, A. S., Tsujino, S., Shanske, S., DiMauro, S., Engel, A. G. *et al.* Aberrant splicing in adult onset glycogen storage disease type II (GSDII): molecular identification of an IVS1 (–13T–>G) mutation in a majority of patients and a novel IVS10 (+1GT–>CT) mutation. *Hum. Mol. Genet.* **3**, 2231–2236 (1994).
35 Gurvich, O. L., Maiti, B., Weiss, R. B., Aggarwal, G., Howard, M. T. & Flanigan, K. M. DMD exon 1 truncating point mutations: amelioration of phenotype by alternative translation initiation in exon 6. *Hum. Mutat.* **30**, 633–640 (2009).
36 Flanigan, K. M., Dunn, D. M., von Niederhausern, A., Howard, M. T., Mendell, J., Connolly, A. *et al.* DMD Trp3X nonsense mutation associated with a founder effect in North American families with mild Becker muscular dystrophy. *Neuromuscul. Disord.* **19**, 743–748 (2009).
37 Flanigan, K. M., Dunn, D. M., von Niederhausern, A., Soltanzadeh, P., Gappmaier, E., Howard, M. T. *et al.* Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. *Hum. Mutat.* **30**, 1657–1666 (2009).
38 Almomani, R., van der Stoep, N., Bakker, E., den Dunnen, J. T., Breuning, M. H. & Ginjaar, I. B. Rapid and cost effective detection of small mutations in the DMD gene by high resolution melting curve analysis. *Neuromuscul. Disord.* **19**, 383–390 (2009).
39 Lenk, U., Hanke, R. & Speer, A. Carrier detection in DMD families with point mutations, using PCR-SSCP and direct sequencing. *Neuromuscul. Disord.* **4**, 411–418 (1994).
40 Hofstra, R. M., Mulder, I. M., Vossen, R., de Koning-Gans, P. A, Kraak, M., Ginjaar, I. B. *et al.* DGGE-based whole-gene mutation scanning of the dystrophin gene in Duchenne and Becker muscular dystrophy patients. *Hum. Mutat.* **23**, 57–66 (2004).
41 Mahmood, O. A., Jiang, X. & Zhang, Q. Limb-girdle muscular dystrophy subtypes: first-reported cohort from northeastern China. *Neural Regen. Res.* **8**, 1907–1918 (2013).
42 Di Fruscio, G., Garofalo, A., Mutarelli, M., Savarese, M. & Nigro, V. Are all the previously reported genetic variants in limb girdle muscular dystrophy genes pathogenic? *Eur. J. Hum. Genet.* **24**, 73–77 (2016).
43 Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl Acad. Sci. USA* **106**, 19096–19101 (2009).
44 Cooper, D. N., 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.* **132**, 1077–1130 (2013).
45 Muntoni, F., Bonne, G., Goldfarb, L. G., Mercuri, E., Piercy, R. J., Burke, M. *et al.* Disease severity in dominant Emery Dreifuss is increased by mutations in both emerin and desmin proteins. *Brain* **129**, 1260–1268 (2006).
46 Lemmers, R. J., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santen, G. W. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)