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

Hemakumar M Reddy1, Kyung-Ah Cho1,12, Monkol Lek2,3, Elicia Estrella4, Elise Valkanas2,3, Michael D Jones1, Satomi Mitsuhashi5,13, Basil T Darras5, Anthony A Amato6, Hart GW Lidov7, Catherine A Brownstein4,8, David M Margulies4,8, Timothy W Yu4, Mustafa A Salih9, Louis M Kunkel4, Daniel G MacArthur2,3 and Peter B Kang1,5,10,11

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 ANOS1, 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
Limbgirdle 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 allow 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 immunohistostchemistry, 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 States5 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 sequencing7 and 33% in Germany using targeted sequence capture.8

1Division of Pediatric Neurology, Department of Pediatrics, University of Florida College of Medicine, Gainesville, FL, USA; 2Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 3Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA; 4Division of Genetics and Genomics, Boston Children’s Hospital and Harvard Medical School, Boston, MA, USA; 5Department of Neurology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA; 6Department of Pathology, Boston Children’s Hospital and Harvard Medical School, Boston, MA, USA; 7Research Connection and Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children’s Hospital and Harvard Medical School, Boston, MA, USA; 8Department of Pediatrics, Division of Neurology, College of Medicine and King Khalid University Hospital, King Saud University, Riyadh, Kingdom of Saudi Arabia; 9Department of Neurology and Department of Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville, FL, USA and 10Genetics Institute, University of Florida, Gainesville, FL, USA
11Current address: Department of Microbiology, School of Medicine, Ewha Womans University, Seoul 07985, Republic of Korea.
12Current address: Biomedical Informatics Laboratory, Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 299-1193, 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.9,10 The Agilent Sure-Select Human All Exon v2.0, 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/broadinstitute/picard) 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 v4 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 >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 (NHHLBI 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://cgl. vital-it.ch/ama/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 sequence 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 |
|-----------------|---------------------|----------------------|----------------|-----------|------------------------|-----------------|-------------------------------------|------------------|
| LMNA missense   | Dominant            | 3                    | Irish/English/Italian/Caucasian | 2700 | Genetic: CAPN3, CAV3, FKRP, DMD, SMN1 IHC: Dystrophin, sarcoglycans, emerin, merosin | Dystrophic | Arrhythmia/implanted defibrillator | Ambulates with difficulty |
| LMNA missense   | Sporadic, 1 affected ♂ | 2                    | Caucasian | 2700 | Genetic: DMD deletion/ duplication, SMN1 IHC: Dystrophin, merosin, sarcoglycans, desmin | Dystrophic | ECG normal | Unknown |
| COL6A3 missense | Sporadic, 1 affected ♂ | 15                   | Irish/English/Dutch/Caucasian | 300 | Genetic: CAV3, FKRP, DMD IHC: Dystrophin, merosin | Myopathic | N/A | Ambulatory |
| COL6A3 essential splice site | Sporadic, 1 affected ♂ | 2                    | Belish/Hungarian/Caucasian | 300 | Genetic: Athena LGMD panel negative IHC: Dystrophin, merosin | Myopathic | None | Non-ambulatory |
| COL6A3 essential splice site | Sporadic, 1 affected ♀ | 2                    | Polish/Slovakian/Irish/German/Caucasian | 900 | Genetic: CAPN3, CAV3, FKRP IHC: Dystrophin, sarcoglycans, merosin | Dystrophic | N/A | Ambulatory |
| COL6A3 | Sporadic, 1 affected ♀ | 5                    | Irish/Canadian/African American/Italian | 300 | Genetic: CAV3, FKRP IHC: Dystrophin, sarcoglycans, merosin | Dystrophic | None | Non-ambulatory |
| RYR1 missense   | Sporadic, 1 affected ♂ | 3                    | Italian/Caucasian | 30–50 | Genetic: Athena LGMD panel negative IHC: Dystrophin, sarcoglycans, merosin | Myopathic | None | Non-ambulatory |
| VCP missense    | Dominant, 3 generations affected | 42                   | German/Caucasian | 400 | Genetic: DES and 13 cardiomyopathy genes, myofibrillar myopathy/ cardiomyopathy | Dystrophic | N/A | Ambulatory |
| FLNC missense   | Sporadic, 1 affected ♂ | 5                    | Italian/French/Irish/Scottish/English/Caucasian | 120–150 | Genetic: DMD, CAV3, LMNA IHC: Dysferlin | N/A | Transplant for cardiomyopathy | Ambulatory |
| SMCHD1 in-frame deletion of 3 bp D4Z4 deletion at 4q35 | Sporadic, 1 affected ♂ | 12                   | Caucasian/Polish | 600 | Genetic: FSHD1, CAV3, LMNA IHC: Dystrophin | Dystrophic | None | Non-ambulatory |
| CAPN3 compound heterozygous missense | Affected brother and sister | 22                   | German/English/Russian/Czech/Caucasian | 1000 | Genetic: 4q35 deletion confirmed after enrollment IHC: Dysferlin | Dystrophic | N/A | Ambulatory |
| CAPN3 compound heterozygous affecting splicing SGGG compound heterozygous affecting splicing | Affected brother and sister | 23                   | Fr. Canadian/German/Native American/Irish/Caucasian | 1400 | Genetic: FSHD1 IHC: Dystrophin, sarcoglycans, emerin, calpain 3 | Dystrophic | N/A | Ambulatory |
| SGGG compound heterozygous affecting splicing | Affected brother, half-brother | 11                   | Puerto Rican/Hispanic/Pakistani/Asian | 8000–18 000 | Genetic: FSHD1, CAV3, LMNA IHC: Dysferlin | Dystrophic | N/A | Ambulatory |
| SGGG compound heterozygous missense | Consanguineous, 2 affected ♀ ≤ ♀ | 44                   | Caucasian | 18 000 | Genetic: DMD, CAV3, BSG IHC: Alpha-sarcoglycan | Dystrophic | N/A | Ambulatory |
| ANOS compound heterozygous missense | Sporadic, 1 affected ♂ | 34                   | Caucasian | 2000–6000 | Genetic: Dystrophin, sarcoglycans, merosin | Dystrophic | Arrhythmia/ablation | Ambulatory |
| ANOS compound heterozygous missense | Affected brother and sister | 18                   | Swedish/Irish/Norwegian/Native American/Irish/Caucasian | 1000–2000 | Genetic: 13 LGMDs/DMD IHC: Dysferlin | Dystrophic | None | Ambulatory |
| ANOS compound heterozygous missense | 2 affected brothers and an affected paternal cousin | 40                   | Yugoslavian/German/Scandinavian/Irish/Caucasian | 1000–2000 | Genetic: DMD/DMD IHC: Dysferlin | Dystrophic | None | Ambulatory |
| ANOS compound heterozygous missense | Sporadic, 1 affected ♂ | 35                   | Welsh/English/Italian/Caucasian | 1000–2000 | Genetic: 13 LGMDs/DMD IHC: Dysferlin | Dystrophic | None | Ambulatory |
| LAMA2 compound heterozygous missense | Sporadic, 1 affected ♀ | 2                    | NE European/Caucasian | 30 000 | Genetic: EMD, LMNA, FHL1 (has elbow, knee and neck contractures) | N/A | None | Non-ambulatory |
| DMD compound heterozygous missense | Sporadic, 1 affected ♀ | 8                    | Caucasian | 600 | Genetic: Dystrophin, sarcoglycans, merosin | Dystrophic | None | Ambulatory |

Abbreviations: CK, creatine kinase; ECG, electrocardiography; IHC, immunohistochemistry; N/A, not available.
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 mutations were found on Sanger sequencing.

Autosomal dominant \textit{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 \textit{COL6A1} mutations
Family 1092 was found to have novel dominant missense pathogenic mutations in \textit{COL6A1}. This gene is classically associated with Bethlem myopathy and Ullrich congenital muscular dystrophy, but recent reports also link it with LGMD.\textsuperscript{12,13} The \textit{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://geneticlab.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.\textsuperscript{14}

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).
in 1115 was reported previously.\textsuperscript{15} The proband in family 1093 showed a mixed phenotype of LGMD and congenital muscular dystrophy.

**Autosomal dominant RYRI mutations**

RYRI (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>\textsuperscript{T}, NP\_001036188.1 p.Arg4856His) in RYRI 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.\textsuperscript{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>\textsuperscript{T}, 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.\textsuperscript{19} The inclusion body myopathy may present with manifestations similar to LGMD.\textsuperscript{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.\textsuperscript{7} The dominant missense pathogenic mutation FLNC, NM\_001458.4, c.7409C>\textsuperscript{T}, 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 FSMD**

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.\textsuperscript{10} While sequence data were being analyzed, the proband from 1258 informed the research team that he had been diagnosed with FSMD1 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.\textsuperscript{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>\textsuperscript{T} 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+722 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 LGMD mutation found in family 1102, ANO5, NM\_213599.2, c.191dupA, NP\_998764.1, p.Asn64Lys fs Ter15 (rs137854521), is a known pathogenic mutation\textsuperscript{23,24} 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,\textsuperscript{26-30} suggesting that LAMA2 should be included among the causative genes for LGMD2.\textsuperscript{31} Compound heterozygous pathogenic mutations in LAMA2, a previously reported nonsense mutation NM\_000426.3, c.5116C>T, NP\_000417.2, p.Arg1706Ter\textsuperscript{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–Dreiffuss muscular dystrophy\textsuperscript{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)\textsuperscript{33} and a substitution c.-32-13T>G r.spl (rs38634236) that affects splicing.\textsuperscript{34} Both mutations were previously reported.

**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.\textsuperscript{35-37}
Table 2 Genetic analysis for the 22 families with genetic diagnoses

| Family locus | Mutation | Inheritance | References | Exome sequencing | Sanger sequencing | Segregation of mutation |
|--------------|----------|-------------|------------|------------------|------------------|-------------------------|
| 930 LMNA     | NM_170707.3 c.832G>A, NP_733821.1 p.Ala278Pro | Dominant | Kuhn et al. | Proband (♂) | Proband | N/A |
| 1125 LMNA    | NM_170707.3 c.746G>A, NP_733821.1 p.Arg249Gln | De novo dominant | Reddy et al. | Proband (♀) | Trio | Present in proband, absent in both parents |
| 965 COL6A3   | NM_004369.3 c.4156G>A, NP_004360.2 p.Glu1386Lys | Dominant | Walters-Sen et al. | Proband (♀) | Proband | N/A |
| 1093 COL6A3  | NM_004369.3 c.6283-1C>T | De novo dominant, essential splice site | Walters-Sen et al. | Proband (♀) | Trio | Present in proband, absent in both parents |
| 1115 COL6A3  | NM_004369.3 c.6156+1G>A | De novo dominant, essential splice site | Walters-Sen et al. | Proband (♀) | Trio | Present in proband, absent in both parents |
| 1092 COL6A1  | NM_0001848.2 c.868G>A, NP_0001839.2 p.Gly290Arg | De novo dominant heterozygous | Walters-Sen et al. | Proband (♀) | Trio | Present in proband, absent in both parents |
| 596 RYR1     | NM_001042723.1 c.14567G>A, NP_001036188.1 p.Arg4856His | De novo dominant | Tan et al. | Proband (♂) | Trio | Present in proband, absent in both parents |
| 1250 VCP     | NM_007126.3 c.572G>A, NP_009057.1 p.Arg191Gln | Dominant heterozygous | Giusti et al. | Proband (♂) | Proband and affected brother | Present in both proband & affected brother |
| 1399 FLNC    | NM_001458.4 c.7409C>A, NP_001449.3 p.Pro2470His | Dominant heterozygous | Proband (♂) | Proband (♂) | Proband and affected sister | Present in both proband and affected sister |
| 1090 SMCHD1  | NM_012595.2 c.818_820del, NP_056110.2 p.Lys275del | Dominant | Ghaoi et al. | Proband (♂) | Proband, parents, brother (1–9) | |
| 1258 D4Z4    | Deletion of D4Z4 repeats at 4q35 | Heterozygous | Proband (♂) | Proband (♂) | Proband and affected sister | N/A |
| 1197 CAPN3   | NM_000070.2 c.1465C>T, NP_000061.1 p.Asp489Glu | Compound heterozygous | Lampe et al. | Proband (♂) | Proband and affected sister | Present in both proband and affected sister |
| 1365 CAPN3   | NM_000070.2 c.945+5G>A, hg19 chr15:42682999 and c.1746-20C>G | Compound heterozygous affecting splicing | rs201892814 (Emory Genetics Laboratory) | Proband (♂) | Proband, unaffected father and affected brother | Present only in proband and unaffected father absent in all three samples |
| 1049 SGCG    | NM_000231.2 c.195+4_195+7 del CAGTA>G | Compound heterozygous affecting splicing | Proband (♂) | Proband and affected brother | Present in both proband and affected brother |
| 1118 SGCG    | NM_000231.2 c.787G>A, NP_000221.1 p.Glu263Lys | Recessive homozygous | Almomani et al. | Proband (♂) | Proband | N/A |
| 1299 SGCA    | NM_000023.2 c.109G>T, NP_000014.1 p.Val37Leu | Consanguineous recessive homozygous | Proband (♂) | Proband, parents, affected brother (~5) and unaffected sister (~4) | Homozygous in proband and affected brother and heterozygous in parents and unaffected sister | N/A |
| 1102 ANOS    | NM_213599.2 c.191dupA, NP_998764.1 p.Asn64Lys fs Ter15 | Recessive homozygous | Tilgen et al. | Proband (♂) | Proband | N/A |
| 1105 ANOS    | NM_213599.2 c.191dupA, NP_998764.1 p.Asn64Lys fs Ter15 | Compound heterozygous | Tilgen et al. | Proband (♂) | Proband | Homozygous in proband and affected brother and heterozygous in parents and unaffected sister | N/A |

References:
- Tan et al. 14
- Brown et al. 15
- Giusti et al. 16
- Lampe et al. 17
- Mercuri et al. 18
- Sato et al. 19
- Watts et al. 20
- Reddy et al. 21
- Walters-Sen et al. 22
- Kuhn et al. 23
- HM Reddy et al. 24
- Almomani et al. 25
- Tilgen et al. 26
- Sato et al. 27
- Mercuri et al. 28
- Watts et al. 29
- Reddy et al. 30
- Walters-Sen et al. 31
- Kuhn et al. 32
- HM Reddy et al. 33

Notes:
- N/A: Not applicable
- Homozygous in 1, 3, 5 and 6; absent in 2, 4 and 7
- Heterozygous in 1, 2 and 6; absent in 3, 4 and 7

Journal of Human Genetics
nasal muscle biopsies of family 1,255. A rare missense variant (NM_032806.5 c.190G > A, p.Asp64Glu) is found in the proband as well as both parents, whereas a 2 base pair deletion (c.740_741delAA, p.Gly247CysfsTer16) is present in the proband and absent in both parents; the latter appears more likely to be pathogenic. The misssense variant of COL6A1 found in family 1366 is novel (NM_001848.2 c.466G > T, p.Val156Ile), 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 0.000085 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 variant (NM_004006.2 c.9G > A, p.Val30Asp) was confirmed to be benign in family 1398. DNA was only available for the proband in this family, thus analysis of cosegregation patterns was not possible.

### Table 2 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 | Heterozygous | Proband (♂)     | Proband          |
| 1255   | POMGNT2 | NM_032806.5 c.190G > A, NP_116195.2 p.Gly64Ser | 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 | Dominant heterozygous | Proband (♂)     | Proband          |

**Possible mutations**

Suspected but unconfirmed mutations are listed in Table 3. Exome sequencing analysis showed that family 1027 has a homozygous 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, consequently 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.Gly247CysfsTer16) is present in the proband and absent in both parents; the latter appears more likely to be pathogenic. The misssense variant of COL6A1 found in family 1366 is novel (NM_001848.2 c.466G > T, p.Val156Ile), 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.

### 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 | Heterozygous | Proband (♂)     | Proband          |
| 1255   | POMGNT2 | NM_032806.5 c.190G > A, NP_116195.2 p.Gly64Ser | 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 | Dominant heterozygous | Proband (♂)     | Proband          |
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,4,11 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 ANOS1, 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 ANOS1. 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 ANOS 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.41

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 regions43 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 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.

**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 (g) |
| 1365   | USA       | CAPN3 NM_000070.2 c.945+5G>A and c.1746-20C>G | Present in proband (q) (heterozygous) |
| 1309   | Saudi Arabian | Unknown            | Present in proband (g) and unaffected brother (1309-5) |
| 1398   | Saudi Arabian | SGCG NM_000231.2 c.212T>C, NP_000222.1 p.Leu71Ser homozygous | Present in proband (q) (heterozygous), unaffected father (1398-3), and unaffected sister (heterozygous) |
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, including muscular dystrophy. 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 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.
28 Gerasimyan, F., Clement, E., Feng, L. H., Sexey, 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., Alhamad, 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., Pomonio, 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 glycosgen 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:areloration 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 Sloot, N., Bakker, E., den Dunnen, J. T., Breuning, M. H. & Ginjaa, 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., Hanneke, 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., Ginjaa, 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 nesprin proteins. *Brain* 129, 1260–1268 (2006).
46 Lemmers, R. J., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santers, 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)