Journal of Human Genetics (2015) 60, 305–312

www.nature.com/jhg

ORIGINAL ARTICLE

New splicing mutation in the choline kinase beta (*CHKB*) gene causing a muscular dystrophy detected by whole-exome sequencing

Jorge Oliveira[1,2,3], Luís Negrão[4], Isabel Fineza[5], Ricardo Taipa[6], Manuel Melo-Pires[6], Ana Maria Fortuna[3,7], Ana Rita Gonçalves[1,3], Hugo Froufe[8], Conceição Egas[8], Rosário Santos[1,3,10,11] and Mário Sousa[2,3,9,11]

Muscular dystrophies (MDs) are a group of hereditary muscle disorders that include two particularly heterogeneous subgroups: limb-girdle MD and congenital MD, linked to 52 different genes (seven common to both subgroups). Massive parallel sequencing technology may avoid the usual stepwise gene-by-gene analysis. We report the whole-exome sequencing (WES) analysis of a patient with childhood-onset progressive MD, also presenting mental retardation and dilated cardiomyopathy. Conventional sequencing had excluded eight candidate genes. WES of the trio (patient and parents) was performed using the ion proton sequencing system. Data analysis resorted to filtering steps using the GEMINI software revealed a novel silent variant in the choline kinase beta (*CHKB*) gene. Inspection of sequence alignments ultimately identified the causal variant (*CHKB*:c.1031 +3G>C). This splice site mutation was confirmed using Sanger sequencing and its effect was further evaluated with gene expression analysis. On reassessment of the muscle biopsy, typical abnormal mitochondrial oxidative changes were observed. Mutations in *CHKB* have been shown to cause phosphatidylcholine deficiency in myofibers, causing a rare form of CMD (only 21 patients reported). Notwithstanding interpretative difficulties that need to be overcome before the integration of WES in the diagnostic workflow, this work corroborates its utility in solving cases from highly heterogeneous groups of diseases, in which conventional diagnostic approaches fail to provide a definitive diagnosis.
Journal of Human Genetics (2015) **60,** 305–312; doi:10.1038/jhg.2015.20; published online 5 March 2015

INTRODUCTION

Muscular dystrophies (MD) are a set of highly heterogeneous muscle diseases sharing distinctive dystrophic features on muscle histology, namely degenerative fibers and fibrosis, variable myofiber size and abnormally internalized nuclei.[1] MD are classically subdivided into five major groups: congenital MD (CMD), Duchenne/Becker MD, Emery–Dreifuss MD, facioscapulohumeral MD and limb-girdle MD (LGMD). Two of these groups, CMD and LGMD, are particularly heterogeneous from both a clinical and a genetic perspective. Until now 31 different loci have been reported for the LGMDs and 28 genes in CMD.[2,3] Notably, at least seven genes are shared between these two groups, and this list is increasing over time. Concerning CMD, mutations in the laminin-α2 (*LAMA2*) gene are a major cause of the disease, corresponding to ~58% of patients referred for genetic studies in this disease group.[4] In patients with inconclusive imunohistochemistry studies, the diagnostic workup may be more complex and is usually steered by other muscle biopsy features, clinical evaluation, electromyogram, and brain and muscle imaging. Specific clinical signs are important to consider, namely the distribution and progression of muscle weakness, the presence of brain structural changes (with or without mental retardation) and other features such as the presence of contractures, rigid spine and hyperlaxidity.[3] All these features have been used to classify CMD into further subtypes and to establish robust diagnostic algorithms. Besides laminin-α2 deficiency or changes for α-dystroglycan, both detected by imunohistochemistry, other clinically relevant histological findings (such as structural changes) are rare.

Nishino *et al.*[5] reported a subtype of CMD with mitochondrial structural abnormalities, presumably of autosomal recessive inheritance. The four patients described shared some clinical features with

[1]Unidade de Genética Molecular, Centro de Genética Médica Dr Jacinto de Magalhães, Centro Hospitalar do Porto E.P.E., Porto, Portugal; [2]Departamento de Microscopia, Laboratório de Biologia Celular, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal; [3]Unidade Multidisciplinar de Investigação Biomédica (UMIB), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal; [4]Consulta de Doenças Neuromusculares, Hospitais da Universidade de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal; [5]Unidade de Neuropediatria, Centro de Desenvolvimento da Criança Luís Borges, Hospital Pediátrico de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal; [6]Unidade de Neuropatologia, Centro Hospitalar do Porto, Porto, Portugal; [7]Serviço de Genética Médica, Centro de Genética Médica Dr Jacinto Magalhães, Centro Hospitalar do Porto E.P.E., Porto, Portugal; [8]Unidade de Sequenciação Avançada, Biocant, Parque Tecnológico de Cantanhede, Núcleo 04, Cantanhede, Portugal; [9]Centro de Genética da Reprodução Professor Alberto Barros, Porto, Portugal and [10]UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal
[11]These authors contributed equally to this work.
Correspondence: Dr R Santos, Unidade de Genética Molecular, Centro de Genética Médica Dr Jacinto de Magalhães, Centro Hospitalar do Porto E.P.E., Praça Pedro Nunes, 88, 4099-028 Porto, Portugal.
E-mail: rosario.santos@chporto.min-saude.pt
Received 21 October 2014; revised 14 January 2015; accepted 2 February 2015; published online 5 March 2015

other CMD forms, but had marked mental retardation without structural brain changes. Disease progression was slow and some patients developed dilated cardiomyopathy. Creatine kinase (CK) levels were mildly elevated. Muscle phenotype included dystrophic features and a striking finding was observed with succinate dehydrogenase and cytochrome *c* oxidase staining, revealing reduction of mitochondrial content in the center of the myofiber and enlarged mitochondria at the periphery, assuming a 'megaconial' appearance.[5] This entity was later demonstrated to be caused by mutations in the choline kinase beta (*CHKB*) gene, leading to loss of activity for this enzyme in rostrocaudal MD mice[6] and in humans in a very limited number of patients reported to date.[7–11] CHKB together with two isoforms of choline kinase alfa (CHKA) phosphorylates choline into phosphocholine, the initial step of the Kennedy's pathway. This step is essential in the biosynthesis of phosphatidylcholine—an important component of the phospholipidic membrane layer of eukaryotes—and also a key precursor for many signaling molecules with relevant cellular functions such as cell growth signaling and proliferation.[12] The disease mechanism has been attributed to the decreased levels of phosphatidylcholine in the mitochondria, leading to its dysfunction and subsequently mitophagy and clearance in the skeletal muscle.[13,14] Immunoprecipitation studies performed in the mouse liver suggested that chk activity was mainly due to heterodimeric CHKA/CHKB (60%), although some of the activities were due to homodimeric forms.[6] However, in human skeletal muscle CHKA is not expressed; therefore, the CHKB isoform is essential to sustain normal phosphatidylcholine levels.[7]

The development and application of new sequencing technologies, frequently coined next-generation sequencing (NGS), is currently accelerating research in the field of clinical genetics, essentially given its capacity to generate and analyze more genomic sequences at a lower cost and at a faster pace.[15] Using this technology it is possible to analyze several specific genes simultaneously, comprising a disease panel, or at the limit it may include all the exonic regions of the human genome—whole-exome sequencing (WES) or even the entire genome (individual genome sequencing). Whole-exome or -genome sequencing may also contribute to the identification of new genes associated with genetic diseases. Nevertheless, its introduction into routine diagnostics still faces some challenges, particularly considering the amount of data to be analyzed, and the possibility of obtaining false-positive and -negative results should not be neglected.

In this work we report the use of WES in a MD patient, which contributed to the identification of a novel splicing mutation in the *CHKB* gene. Our study corroborates the potential of WES to identify genetic defects involved in rare neuromuscular diseases and illustrates the difficulties that may be encountered using this technology, especially regarding data analysis.

MATERIALS AND METHODS

Patient

The patient is the first child of a nonconsanguineous healthy couple; pregnancy and neonatal period were uneventful. She was first observed by neuropediatrics at the age of 4 years because of motor problems: independent walking at 2.5 years, but with progressive lower limb weakness and difficulties in climbing stairs and running. She is presently unable to rise from a chair, the arms' reach do not rise to the shoulder level and she walks with braces. The CK values, ascertained over a period of 30 years, were slightly elevated, ranging from 299 to 1857 $U\,l^{-1}$ (average 849 $U\,l^{-1}$; normal $<200\,U\,l^{-1}$). Electromyography performed at the of age 7 showed a myopathic pattern without evidence of fibrillation potentials or myotonic discharges. She also presented behavioral abnormalities and cognitive deficit, which prevented her from attaining the basic academic goals.

At the end of the first decade dilated cardiomyopathy was diagnosed, adequately controlled with medication (digoxin and furosemide). Standard genetic screening tests due to delay in psychomotor development (including speech delay), performed at the age of 22, were all normal. The patient had no signs of ichthyosis. Informed consent was obtained and the research was approved by the ethics committee from the Centro Hospitalar do Porto.

Sanger sequencing

Eight genes known to be defective in MDs were previously analyzed by conventional (Sanger) sequencing, namely: *ANO5, CAPN3, LARGE, FKRP, FKTN, POMGNT1, POMT1* and *POMT2*. Genomic sequencing included all coding regions and adjacent intronic sequences. Results were analyzed using the SeqScape software V2.5 (Life Technologies, Foster City, CA, USA). Variants were described according to the Human Genome Variation Society guidelines for mutation nomenclature (version 2.0).[16] Large heterozygous deletions and duplications were also screened in several genes using the multiplex ligation-dependent probe amplification technique. Different commercial kits from MRC-Holland (Amsterdam, the Netherlands) were used: *SGCA, SGCB, SGCD, SGCG* and *FKRP* (P116); *FKTN, LARGE* and *POMT2* (P326); *POMT1* and *POMGNT1* (P061). Genomic DNA (150 ng) from the patient and at least three normal controls were used in the procedure. Multiplex ligation-dependent probe amplification products were resolved by capillary electrophoresis on an ABI 3130xl genetic analyzer (Life Technologies). Data analysis was conducted using the GeneMarker Software V1.5 (SoftGenetics LLC, State College, PA, USA). After using the population normalization method, data were plotted using the probe ratio option.

Exome-sequencing

Seventy-five nanograms of high-quality DNA from each individual (patient and both parents) were amplified in 12 primer pools of 200-bp amplicons with the Ion AmpliSeq Exome Library Preparation kit (Life Technologies). Samples were barcoded with the Ion Xpress Barcode Adapter (Life Technologies) in order to pool two exomes per chip. The libraries were evaluated for quality with High Sensitivity DNA Kit in Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantified using the Ion Library Quantitation Kit (Life Technologies). Library fragments were clonally amplified with emulsion PCR using the Ion PI Template OT2 200 kit v2 and the Ion OneTouch 2 System (Life Technologies), and the positive-ion sphere particles enriched in the Ion OneTouch ES machine (Life Technologies). Finally, the positive spheres were loaded in Ion PI chip v2 and sequenced in the Ion Proton System (Life Technologies). All procedures were carried out according to the manufacturer's instructions.

Exome mapping and variant calling

The data generated from the three exomes (patient and parents) were processed with the Torrent Suite Software 4.1 (Life Technologies). Reads were mapped against the human reference genome hg19 using Torrent Mapping Alignment Program version 4.0.6 (Life Technologies). Variant calling was performed by running Torrent Variant Caller plugin version 4.0, using the optimized parameters for exome-sequencing recommended for AmpliSeq sequencing (Life Technologies). The variant call format (VCF) files from the trio were combined using vcftools version 0.1.9.0.[17] and all variants were annotated and prioritized with GEMINI.[18] Variant filtering was performed using a list of 108 candidate genes known to be implicated in hereditary myopathies (Supplementary Data I). Each gene was screened for variants matching the recessive disease model, being: (i) homozygous in the patient and parents carrying the same variant in heterozygosity ('autosomal recessive') or (ii) two heterozygous variants in the patient (compound heterozygosity, 'comp_hets' function) and each having a distinct parental origin. Candidate variants were manually checked on the Binary Alignment Map files through Integrative Genomics Viewer (IGV) version 2.3[19] and GenomeBrowse V2.0.0 (Golden Helix Inc., Bozeman, MT, USA).

Confirmation of variants with sanger sequencing

To verify the candidate variant detected in the *CHKB* gene of the three individuals, specific primers were designed (9 F-5′-GTGGAGTCTGGAAGG

AATGGC-3′; 9 R-5′-TTTAACTTCTCCCCCACTGTCC-3′) to amplify the region of interest using PCR. Resulting PCR products were purified using Illustra ExoStar (GE Healthcare, Little Chalfont, UK) and subjected to a new cycling reaction prepared with BigDye Terminator kit V3.1 chemistry (Life Technologies). Sequencing reactions were analyzed in an ABI 3130xl genetic analyzer (Life Technologies).

Gene expression analysis

Total RNA was obtained from the peripheral blood (patient and a control) using the PerfectPure RNA Blood Kit (5 PRIME, Hamburg, Germany) and from muscle (only available from a control) with the PerfectPure RNA Fibrous Tissue kit (5 PRIME). RNA was converted into cDNA using the High Capacity cDNA Reverse Transcription kit (Life Technologies). For the analysis of *CHKB* transcripts, PCR amplification was performed using primers annealing to the regions corresponding to exons 7–11 (6/7 F-5′-TCCAGGAAGGGAACA TCTTG-3′; 11 R-5′-GGTGGAGTCAGGATGAGGAG-3′, on the basis of the reference sequence with accession number NM_005198.4). As an experimental control, the 3′ region of the transcripts for protein-O-mannosyltransferase 1 (*POMT1*) was simultaneously amplified, using the following primers: 18 F-5′-CGGCGAAGAAATGTCCATGAC-3′ and 3′-UTR-R-5′-GAGCTTTTCAATGA GACCCCC-3′ (on the basis of reference sequence with accession number NM_007171.3). Resulting PCR products were resolved in a 2% agarose gel; visible bands were excised and purified using the PureLink Quick Gel Extraction kit (Invitrogen by Life Technologies). Products were sequenced as described above.

Muscle histology

Two muscle biopsies were performed, at 9 and 20 years of age. In the first biopsy, histochemical studies were not performed and paraffin sections showed nonspecific changes. The second biopsy revealed striking variation in fiber size, with type 1 and type 2 fiber atrophy, necrotic fibers, several basophilic fibers and endomysial fibrosis. Imunohistochemistry showed normal labeling for sarcoglycans, dystrophin and laminin-α2. succinate dehydrogenase and nicotinamide adenine dinucleotide staining was uneven in central areas of the myofiber. The conclusion at that time (1995) was that these findings pointed toward a possible MD.

RESULTS

Preceding molecular genetic analysis

The patient was followed by neuropediatrics and later neurology, with the clinical suspicion of LGMD or possibly a dystroglycanopathy. Eight candidate genes were initially analyzed using Sanger sequencing as part of the routine genetic diagnostic workup. Several variants were detected in these studies, including a novel silent heterozygous variant (NM_024301.4:c.474C>G) in the fukutin-related protein (*FKRP*) gene, which prompted further characterization at the complementary DNA (cDNA) level. As no splicing defects were identified, this unclassified variant was considered nonpathogenic. These studies were complemented by screening for large genomic deletions or duplications at several loci by multiplex ligation-dependent probe amplification; however, again no causal mutations were detected. These continuous molecular studies were inconclusive and unable to identify the genetic defect responsible for the patient's phenotype. The high number of candidate genes that were screened and the detailed clinical characterization consistent with a genetically determined MD led us to consider WES as an appropriate alternative approach to identify the underlying defect.

Exome-sequencing metrics

Overall quality parameters of the exome trio are summarized in Table 1. On average, each exome generated a total of 37 million sequence reads, 99% of which successfully aligned against the human reference genome. Accordingly, the exome's target regions were on average covered 105.6 times and 93% of these had a minimum coverage depth of 20-fold. These values are relatively high and above the threshold considered appropriate for this study. Approximately 83% of all sequenced bases had a Phred quality score above Q20 (that is, a 99% accuracy of the base call). The size of the total captured regions was ~58 Mbp. For each sequenced exome one Binary Alignment Map and VCF files were generated. An overview of the variants called for each exome is described in Table 2. WES identified 48 847 single-nucleotide variants in the index case and 48 870 and 47 878 single-nucleotide variants in each parent. More than 2300 insertions/deletions were also detected in each individual exome. The patient's exome contains 1820 variants not previously reported in single nucleotide polymorphism database (dbSNP) (http://www.ncbi.nlm.nih.gov/SNP/).

Exome analysis and variant filtering

To deal with the large amount of variants identified using WES (>50 000 variants in each individual exome), the analysis approach was based on the patient's clinical phenotype and the most probable inheritance pattern for the disease. The first filtering strategy focused on the known genetic causes for hereditary myopathies (including MDs), where a total of 108 candidate genes were selected (Supplementary Data I), on the information available in the Muscle Gene Table website (http://musclegenetable.fr/, accessed April 2014). The GEMINI database framework was used to annotate and filter the large number of sequence variants listed in the three generated VCF files. This software has different built-in analysis tools that are useful for identifying *de novo* mutations, but also to filter variants meeting autosomal recessive or autosomal-dominant inheritance patterns in family-based studies.[18] As no other relatives of our patient were reported as being affected with a muscle disease, the autosomal recessive model was used, which in GEMINI is subdivided into two

Table 1 Exome-sequencing quality metrics

Individual	*Patient*	*Mother*	*Father*
Total reads (number)	41 776 885	38 365 995	31 441 741
Aligned reads passed filtering (number)	41 271 480	37 922 459	31 102 979
Aligned reads passed filtering (%)	99	99	99
Mean coverage of target region (*x*)	117.3	109.5	89.96
Bases with *Q*>20 (%)	83.02	83.37	83.60
Total captured region size (Mb)	57.74	57.74	57.74
Captured regions with coverage >20 (%)	93.19	93.42	91.17

Sequencing and mapping metrics are presented per individual.

Table 2 Exome variant metrics

Individual	*Patient*	*Mother*	*Father*
Total number of SNVs	48 847	48 870	47 878
Total number of indels	2473	2375	2332
Homozygous variants	19 510	19 231	18 958
Heterozygous variants	31 810	32 014	31 252
Novel variants[a]	1820	1749	1610

Abbreviation: SNV, single-nucleotide variant.
Variants (with quality ⩾50) are presented per individual. Variant calling files from each sequenced individual were combined and analyzed with GEMINI.
[a]Not listed in dbSNP (no 'rs' number attibuted).

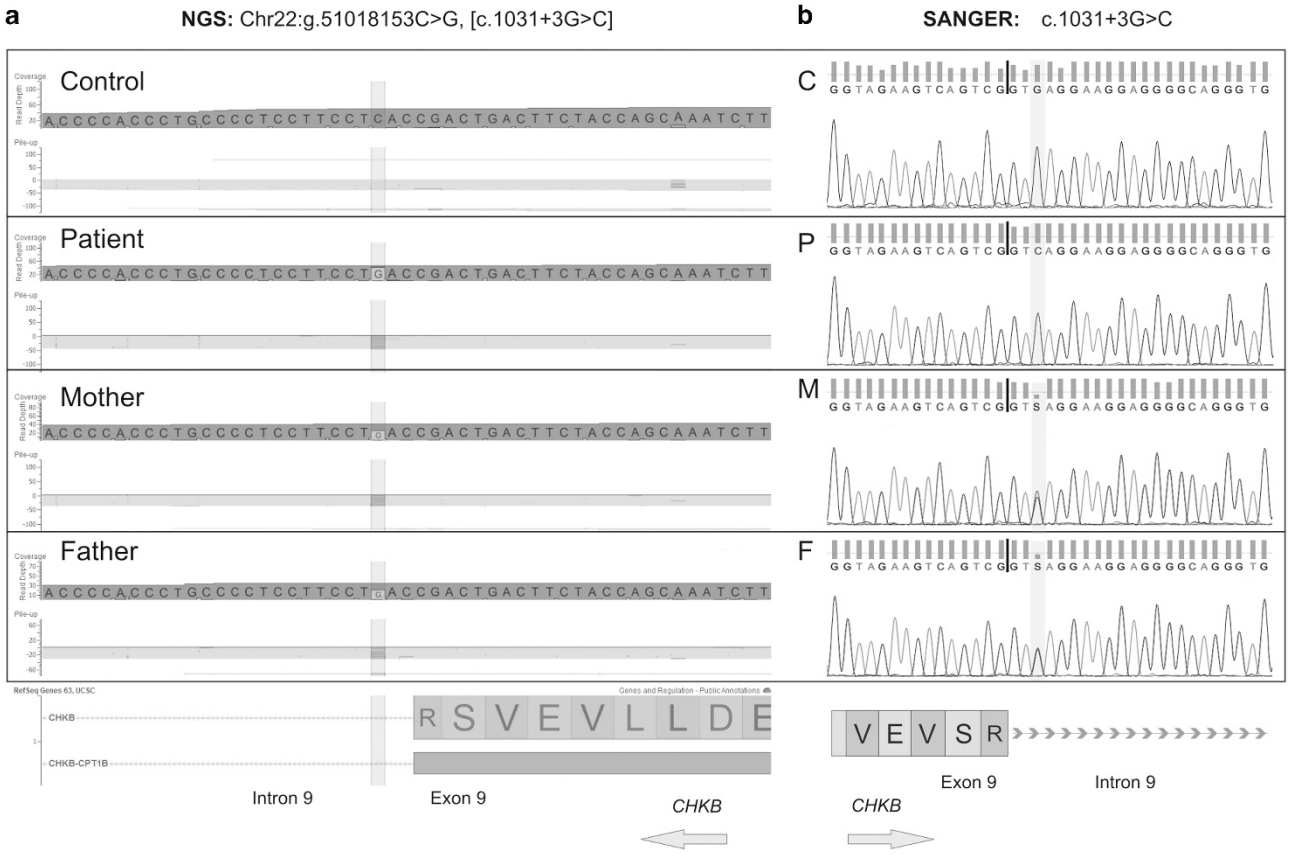

Figure 1 Novel splice mutation (c.1031+3G>C) identified in the *CHKB* gene. (**a**) Sequencing results from next-generation sequencing (NGS) showing this variant in the patient and in both parents. For this region of the gene, only reversely orientated reads were generated. (**b**) Although annotated in the variant call format (VCF) file as being present in heterozygosity, Sanger sequencing confirmed this variant to be homozygous in the patient (P); the patient's mother (M) and father (F) are both heterozygous carriers. cDNA reference sequence: NM_005198.4. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

different functions: 'comp_hets' and 'autosomal_recessive'. Considering that parental consanguinity was not reported, the first tool used was 'comp_hets'. This function is suitable to identify potential compound heterozygote genotypes. A list of compound heterozygous variants in the selected 108 genes was generated for the patient (Supplementary Data II). A total of 29 heterozygous variants in 11 different genes were identified, 24 are considered polymorphisms (frequency higher than 1.0%) and five had not been documented in dbSNP. After manually checking the read alignments on those genomic coordinates (for the trio and by comparing with a control exome), three variants were considered sequencing artifacts. One heterozygous missense variant (NM_004369.3:c.1927C>A, p.Leu643Ile) was detected in the α3 chain of the collagen VI gene (*COL6A3*); however, it was seen to have been inherited from the father, who was homozygous for this change. A second variant (NM_001267550.1: c.33513_33515dup) located in the titin gene (*TTN*), predictably resulting in the insertion of a new codon, was considered nonpathogenic as it was previously reported in the Exome Variant Server (http://evs.gs.washington.edu/EVS/) with an overall frequency of 1.3%.

As no suitable candidates were found, a second filtering approach ('autosomal_recessive') was used. This identifies canonical recessive mutations by generating a list of homozygous variants present in the patient and inherited from each parent, heterozygotes for those changes. In this analysis fewer variants were identified ($n=14$) present in a total of eight different genes (one variant each in *CHKB*, *CNTN1*, *DNM2*, *ETFDH*, *ITGA7*, *NEB* and *TIA1* and seven variants in *SYNE1;* Supplementary Data II). Thirteen variants were already listed in publically available sequence variant databases (frequency higher than 1.0%) and one was a novel

Table 3 Data details for the main candidate variant

Individual	*Patient*	*Mother*	*Father*
Chromosome	22	22	22
Position	51 018 153	51 018 153	51 018 153
Gene name	*CHKB*	*CHKB*	*CHKB*
Depth	47	39	34
Reference allele	C	C	C
Number of reads with reference allele (frequency)	2 (0.04)	16 (0.41)	14 (0.41)
Alternate allele	G	G	G
Number of reads with alternate allele (frequency)	45 (0.96)	23 (0.59)	20 (0.49)
Mutation type	Splice site	Splice site	Splice site
Refseq accession number	NM_005198.4	NM_005198.4	NM_005198.4
Mutation DNA	c.1031+3G>C	c.1031+3G>C	c.1031+3G>C
Confirmed by Sanger sequencing	Yes	Yes	Yes

Abbreviation: BAM, Binary Alignment Map.
Data derived from manual inspection of the BAM file and correction of artifacts for the candidate position in each of the parents and the patient.

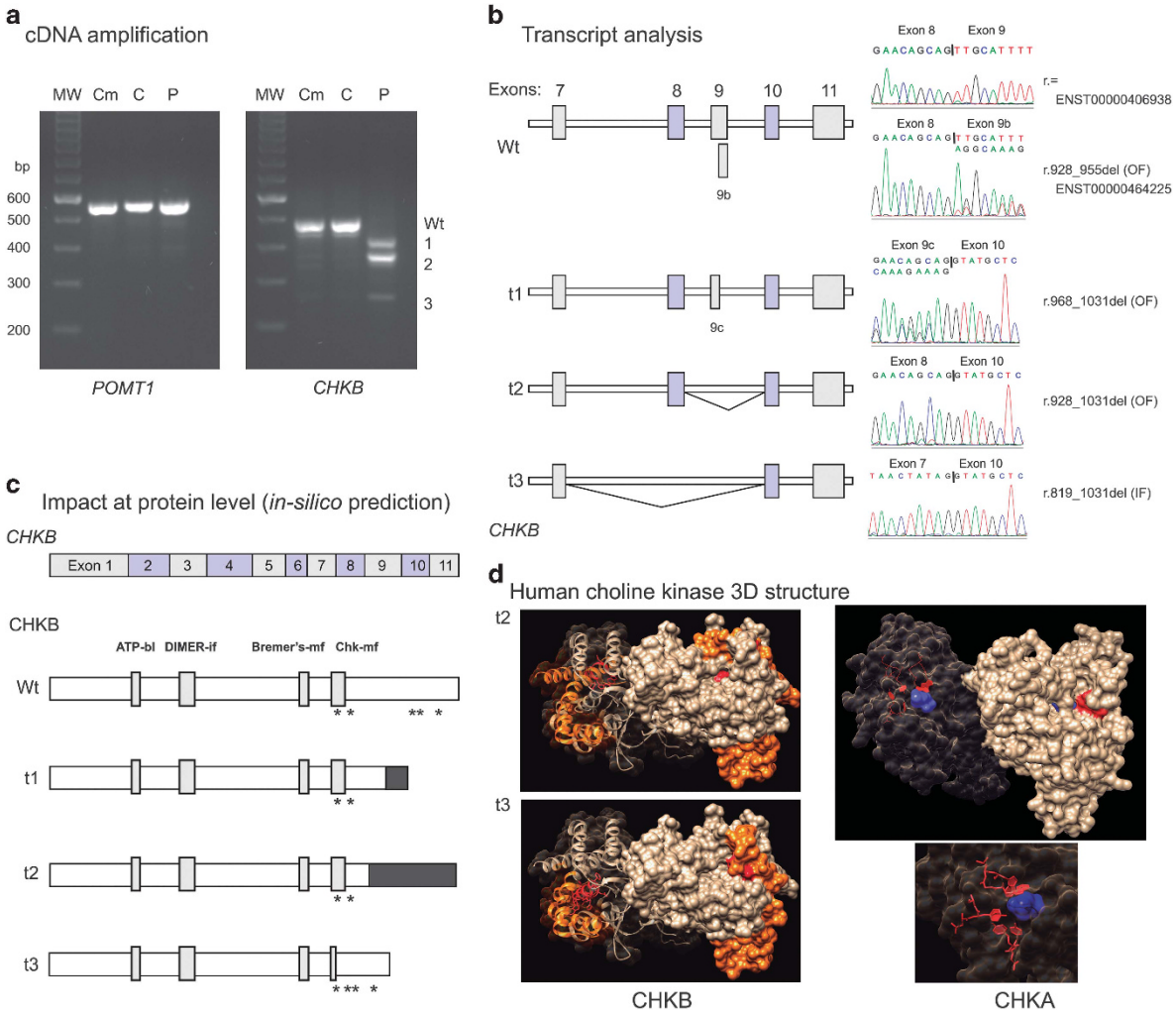

Figure 2 *CHKB* gene expression analysis. (**a**) Reverse transcriptase-polymerase chain reaction (RT-PCR) experiments using total RNA obtained from the peripheral blood (C, control; P, patient) and from the muscle (available only for the control sample, Cm). Amplification of *POMT1* transcripts was performed to show successful amplification in the different samples. Normal *CHKB* transcripts (Wt) were seen in the control sample, whereas in the patient only three aberrant transcripts (1–3) were present. (**b**) Resulting PCR products were sequenced for further transcript characterization. Besides the normal full-length transcript, there is also an alternatively spliced isoform (9 b). Three aberrant transcripts were identified in the patient: t1, resulting from the use of a cryptic splice site in exon 9, causing an out-of-frame (OF) deletion; t2, full exon 9 skipping causing a frameshift; and t3, skipping of exons 8 and 9 leading to corresponds to an in-frame (IF) deletion. (**c**) Bioinformatic analysis of aberrant transcripts. Gray rectangles draw attention to relevant CHKB domains: (i) ATP-binding loop (bl); (ii) dimer interface (if); (iii) Bremer's motif (mf); and (iv) Chk-mf.[26] Asterisks highlight important aromatic residues involved in stabilization of the positively charged amine of choline. Black regions represent residues predictably altered as a result of the frameshift. (**d**) Crystallographic three dimensional (3D) structures of the human choline kinase beta (CHKB) with Protein Data Bank (PDB): 2IG7 and also the α isoform (choline kinase alfa (CHKA)) with PDB: 2CKQ, showing its interaction with the subtract (choline). Both structures were visualized using the Chimera software V1.9.[27] In one of the chains of the CHK homodimers, transparency was increased to show relevant secondary structures and/or residues. The regions affected by the deletions identified in transcripts t2 and t3 are depicted in orange. Red residues indicate aromatic amino acids known to interact with choline, which is shown in blue.

sequence change located in the *CHKB* gene (NM_005198.4: c.843T>C).

Variant data interpretation

The novel homozygous variant c.843T>C located in exon 8 of the *CHKB* gene (Supplementary Data III) was evaluated by bioinformatic tools and prompted a detailed analysis of the alignments encompassing this gene. An additional novel variant was detected (NM_005198.4: c.1031+3G>C), located in the donor splice site of intron 9 (Figure 1a and Table 3). In this region, the read depth was ~40-fold and only reverse sequences were successfully aligned (Supplementary Data IV). It should be noted that the c.1031+3G>C variant was listed as heterozygous in the VCF files of all three exomes, thereby explaining why it was filtered-out with GEMINI's 'autosomal_recessive' function. Conventional (Sanger) sequencing confirmed it to be homozygous in the patient and heterozygous in both parents (Figure 1b). In addition, this variant was not found in homozygosity in the patient's clinically unaffected sister (data not shown). Taking into account the possibility of other false-negatives in the remaining candidate loci, all heterozygous variants were reassessed. The genotype quality score of 90 was determined as a suitable cutoff value for this reassessment (Supplementary Data V). Eleven sequence variants (including the c.1031+3G>C change in *CHKB*) were retrieved that had been called as heterozygous in all three samples and having a genotype quality score <90 in the patient. With the exception of the *CHKB* variant, all were polymorphisms and thus not considered as possible candidates.

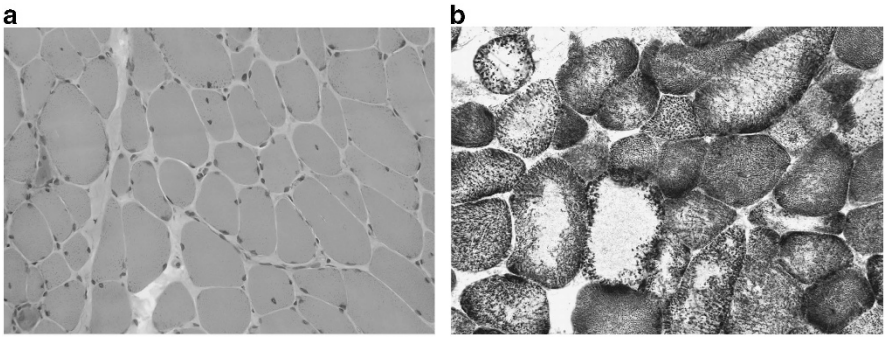

Figure 3 Muscle biopsy showing. (**a**) variation in fiber size, basophilic fibers and endomysial fibrosis (HE, magnification x200) and (**b**) large pale central areas with large mitochondria toward the periphery, seen with oxidative enzyme stain (succinate dehydrogenase (SDH), magnification x200). A full color version of this figure is available at the *Journal of Human Genetics* journal online.

The location of the c.1031+3G>C variant, at the exon–intron junction, leads us to suspect a possible effect on the splicing process; ~95% of mammalian transcripts have an A or a G nucleotide in the +3 position of the donor splice site.[20] This, together with the clinical presentation, led us to consider c.1031+3G>C as a candidate variant accountable for the disease.

CHKB gene expression analysis

Further evidence supporting the pathogenicity of the c.1031+3G>C variant was obtained through bioinformatic analysis. Five different algorithms were used to evaluate splicing: SpliceSiteFinder-Like, MaxEntScan,[21] NNSPLICE,[22] GeneSplicer[23] and Human Splicing Finder,[24] incorporated in the Alamut Visual V2.4 software (Interactive Biosoftware, Rouen, France). In all algorithms there was a reduction of probability scores, and particularly in two (NNSPLICE and GeneSplicer), the results were below the value set as a minimum threshold for splicing (Supplementary Data VI). This effect was experimentally determined by the analysis of *CHKB* transcripts. cDNA derived from the peripheral blood was used, given that no muscle specimens of the patient were available for study. CHKB is ubiquitously expressed and enzymatic activity for this kinase was previously reported in different tissues.[25] Moreover, in control individuals there were no differences in *CHKB* expression between cDNA samples derived from the blood and those derived from muscle (Figure 2a). In the patient there was a drastic effect on splicing; no normal expression for *CHKB* was detected as the expected normal transcript (with 468 bp, as seen to be present in control samples) was not observed. Instead, she presented at least three aberrant transcripts not detected in the controls. These were further characterized by sequencing (Figure 2b). All aberrant transcripts were due to splicing defects related with the mutation-derived weakening of the donor splice site. The first transcript (r.968_1031del) corresponds to a 64-bp out-of-frame deletion, where a cryptic splice site located in the middle of exon 9 is used as an alternative. The other two transcripts originate by exon 9 skipping (r.928_1031del) leading to a 104-bp out-of-frame deletion and by exons 8 and 9 skipping (r.819_1031del) leading to a 213-bp in-frame deletion. Should these mutated transcripts be translated, they predictably result in the loss of critical portions of the enzyme (Figure 2c and d).

The crystallographic structures are available in the Protein Data Bank (PDB) for the human CHKB (PDB: 2IG7) and the α isoform CHKA (PDB: 2CKQ).[26] On the basis of these structures it is predictable that in the mutated polypeptides derived from the transcripts 1 and 2, there is loss of some aromatic residues known to stabilize the positively charged amine of choline; the interaction with this substrate would thus be affected (Figure 2c and d). In addition, the formation of the choline-binding pocket site itself is probably also compromised. The mutated polypeptide with an in-frame deletion derived from transcript 3 (Figure 2c and d) is predicted to abolish part of the choline kinase motif located in the C-terminal region of CHKB, also involved in the binding of the choline moiety.[26]

Muscle histology

The results of WES analysis prompted the revision of the patient's muscle biopsy. Oxidative enzyme staining (succinate dehydrogenase) revealed not only large pale central areas but also abnormal oxidative activity, particularly toward the periphery of the myofiber (Figure 3). Electron microscopy shows the presence of slightly enlarged mitochondria (Supplementary data VII).

DISCUSSION

We describe the identification of a novel intronic splicing mutation in the *CHKB* gene, resorting to the innovative analysis of an exome trio (patient and respective parents). Mutations in this locus were previously reported as associated with an extremely rare form of CMD, with only 21 patients accounted for in the literature[7–11] and at least three additional cases presented in conference proceedings.[28,29] Defects in CHKB have been shown to cause phosphatidylcholine deficiency in muscle cells.[13] The c.1031+3G>C variant located in the donor splice site near exon 9 was considered pathogenic upon *CHKB* expression analysis. It was experimentally demonstrated that in the patient reported herein, there are no normal full-length *CHKB* transcripts, and only aberrant transcripts were identified. If these transcripts escape nonsense-mediated decay, they will probably fail to code for a functional CHKB enzyme. Our experimental findings are similar to those observed in a Turkish CMD patient with a different mutation (c.1031+1G>A) affecting the same donor splice site.[7]

The major clinical features found in the patient (progressive muscle weakness, mental retardation and dilated cardiomyopathy) are in agreement with the phenotype previously described for this entity (OMIM no. 602541). In the first clinical report of 15 patients,[7] findings included mildly to moderately elevated CK levels, floppiness during the neonatal period (in 9/14 patients), mental retardation (15/15), cardiomyopathy (6/14) and independent walking (11/15). Muscle histology was systematically dystrophic and mitochondrial enlargement was observed in all patients. An isolated case with a homozygous nonsense mutation in *CHKB* was identified in a child with a similar clinical and muscle phenotype.[8] The phenotypic spectrum linked to *CHKB* deficiency has been recently expanded with additional reports, including a female patient with a milder phenotype

resembling LGMD[9] and the oldest known patient (50 years old) having a progressive 'limb-girdle' myopathy.[29] Both patients presented missense mutations (two compound heterozygotes and one homozygote in the second report). Muscle histology showed large mitochondria located at the periphery and central areas in the myofiber, devoid of activity. In fact, it should be emphasized that the majority of patients described in the literature harboring *CHKB* mutations were initially identified by the existence of this specific histological hallmark: the presence of enlarged mitochondria in the periphery of the myofibers and the absence of this organelle in the center of the cells.[5] These enlarged mitochondria were correlated with the reduced levels of phosphatidylcholine in the mitochondrial membrane. This seems to be the consequence of a compensatory mechanism resulting from the depletion of functionally compromised mitochondria by autophagy.[14,30] A common aspect, especially in the atypical cases, is that several muscle biopsies needed to be performed. In the oldest patient reported to date, these mitochondrial changes were more evident in the last biopsy (from a total of three), which was performed at the age of 49 years.[29] This pathognomonic finding (megaconial) is highly specific for the defective gene; however, their detection requires awareness and high-quality histological preparations. Therefore, when MD (congenital or progressive) presents in association with mental retardation and cardiac involvement, the muscle biopsy should be carefully reassessed and the *CHKB* gene considered a possible candidate.

As previously referred, NGS technology is improving the capabilities to perform molecular genetic diagnostics. The most rapidly growing application of this technology is the development of NGS gene panels (targeted resequencing), which can help to deal with high genetic and clinical heterogeneity found in several diseases, such as neuromuscular disorders.[31] By performing targeted resequencing it is possible to obtain higher redundancy and thus more reliable variant data. This is achievable because the extension of these assays (covered regions) is usually quite smaller when compared with other (broader) NGS approaches such as WES or whole-genome sequencing. Accordingly, NGS-targeted resequencing will rapidly replace conventional sequencing for routine genetic diagnosis. However, there is still lack of consensus as to which genes should be included (or not) in such disease-specific gene panels and each laboratory implementing these techniques has been defining their own sets of genes. Another important aspect is that the application of NGS gene panels in some diseases such as hereditary myopathies will perpetuate the long-established clinical classification, which is currently being challenged by the increasing reports of the same gene being involved in several different subtypes. The ryanodine receptor 1 (*RYR1*) gene is an excellent example of this paradigm, considering its association with different diseases, ranging from congenital myopathies and MDs to malignant hyperthermia susceptibility, as well as by presenting dissimilar muscle phenotypes: myopathic with or without structural changes (such as central cores, minicores or centrally placed nuclei), dystrophic signs or apparently normal histology but with compromised function (defective intracellular calcium homeostasis causing hypermetabolic response when exposed to anesthetic agents).[32]

In contrast to NGS-targeted resequencing, WES has been proposed as an important research tool for gene discovery or further expanding the genetic causes of a particular disease.[33] Probably in the near future WES will be used as a first-tier diagnostic tool for several medical specialties, allowing the identification of genetic defects linked to monogenic or even complex multigenic disorders. Our initial analysis strategy was orientated toward the genes known to be implicated in hereditary myopathies and, among the attempted approaches, an autosomal recessive model was assumed. It is interesting to note that, although the mutation in the *CHKB* gene was successfully identified by WES, its zygosity was not called correctly in the VCF file used for automated analysis. A different unreported variant in the *CHKB* gene drew our attention to this locus and enabled us to identify the causative mutation. At this stage, considering the novelty aspect of the work, there is no indication whether this is a common pitfall of WES data analysis, and the overall impact on filtering strategies is as yet unknown. To ensure that no additional false-negative results were obtained upon filtering, a strategy based on the genotype quality score (threshold <90) was used. Presently, WES should be seen as a screening technique, and it does not provide the answer to all aspects of clinical molecular genetic testing. As shown here, several other techniques and biological samples are very often required to clarify the impact on the phenotype of novel sequence variants. Moreover, different molecular techniques are required for the detection of more complex mutations such as repeat expansions or structural (copy number) variants that are not readily identified by NGS.[34] Another WES analysis constraint is the occurrence of false-positives. Some of these sequence artifacts are not easily flagged; therefore, they may be interpreted by the software as true variants. Even though they are later excluded by conventional sequencing, they can certainly create extra burden on the analysis.

In conclusion, we present a case resolved by WES that enabled improvement of patient follow-up and management and genetic counseling of family members. A novel *CHKB* gene mutation linked to a rare form of MD was identified by WES, in which previously the conventional diagnostic approaches had failed to provide a definitive diagnosis.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

UMIB is funded by the National Funds through FCT (Foundation for Science and Technology) under the Pest-OE/SAU/U10215/2014.

1 Shieh, P. B. Muscular dystrophies and other genetic myopathies. *Neurol. Clin.* **31**, 1009–1029 (2013).
2 Nigro, V. & Savarese, M. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myol.* **33**, 1–12 (2014).
3 Bönnemann, C. G., Wang, C. H., Quijano-Roy, S., Deconinck, N., Bertini, E. & Ferreiro, A. *et al.* Diagnostic approach to the congenital muscular dystrophies. *Neuromuscul. Disord.* **24**, 289–311 (2014).
4 Oliveira, J., Gonçalves, A., Oliveira, M. E., Fineza, I., Pavanello, R. C. M. & Vainzof, M. *et al.* Reviewing large LAMA2 deletions and duplications in congenital muscular dystrophy patients. *J. Neuromuscul. Dis.* **1**, 169–179 (2014).
5 Nishino, I., Kobayashi, O., Goto, Y., Kurihara, M., Kumagai, K. & Fujita, T. *et al.* A new congenital muscular dystrophy with mitochondrial structural abnormalities. *Muscle Nerve* **21**, 40–47 (1998).
6 Sher, R. B., Aoyama, C., Huebsch, K. A., Ji, S., Kerner, J. & Yang, Y. *et al.* A rostrocaudal muscular dystrophy caused by a defect in choline kinase beta, the first enzyme in phosphatidylcholine biosynthesis. *J. Biol. Chem.* **281**, 4938–4948 (2006).
7 Mitsuhashi, S., Ohkuma, A., Talim, B., Karahashi, M., Koumura, T. & Aoyama, C. *et al.* A congenital muscular dystrophy with mitochondrial structural abnormalities caused by defective de novo phosphatidylcholine biosynthesis. *Am. J. Hum. Genet.* **88**, 845–851 (2011).
8 Gutiérrez Ríos, P., Kalra, A. A., Wilson, J. D., Tanji, K., Akman, H. O. & Area Gómez, E. *et al.* Congenital megaconial myopathy due to a novel defect in the choline kinase Beta gene. *Arch. Neurol.* **69**, 657–661 (2012).
9 Quinlivan, R., Mitsuahashi, S., Sewry, C., Cirak, S., Aoyama, C. & Mooore, D. *et al.* Muscular dystrophy with large mitochondria associated with mutations in the CHKB gene in three British patients: extending the clinical and pathological phenotype. *Neuromuscul. Disord.* **23**, 549–556 (2013).
10 Castro-Gago, M., Dacruz-Alvarez, D., Pintos-Martínez, E., Beiras-Iglesias, A., Delmiro, A. & Arenas, J. *et al.* Exome sequencing identifies a CHKB mutation in Spanish patient with megaconial congenital muscular dystrophy and mtDNA depletion. *Eur. J. Paediatr. Neurol.* **18**, 796–800 (2014).

11 Cabrera-Serrano, M., Junckerstorff, R. C., Atkinson, V., Sivadorai, P., Allcock, R. J. & Lamont, P. *et al.* Novel CHKB mutation expands the megaconial muscular dystrophy phenotype. *Muscle Nerve* **51**, 140–143 (2014).
12 Aoyama, C., Liao, H. & Ishidate, K. Structure and function of choline kinase isoforms in mammalian cells. *Prog. Lipid Res.* **43**, 266–281 (2004).
13 Mitsuhashi, S., Hatakeyama, H., Karahashi, M., Koumura, T., Nonaka, I. & Hayashi, Y. K. *et al.* Muscle choline kinase beta defect causes mitochondrial dysfunction and increased mitophagy. *Hum. Mol. Genet.* **20**, 3841–3851 (2011).
14 Mitsuhashi, S. & Nishino, I. Phospholipid synthetic defect and mitophagy in muscle disease. *Autophagy* **7**, 1559–1561 (2011).
15 Buermans, H. P. & den Dunnen, J. T. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* **1842**, 1932–1941 (2014).
16 den Dunnen, J. T. & Antonarakis, S. E. Nomenclature for the description of human sequence variations. *Hum. Genet.* **109**, 121–124 (2001).
17 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E. & DePristo, M. A. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
18 Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.* **9**, e1003153 (2013).
19 Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
20 Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
21 Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
22 Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).
23 Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).
24 Desmet, F. O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. & Béroud, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
25 Aoyama, C., Ohtani, A. & Ishidate, K. Expression and characterization of the active molecular forms of choline/ethanolamine kinase-alpha and -beta in mouse tissues, including carbon tetrachloride-induced liver. *Biochem. J.* **363**, 777–784 (2002).
26 Malito, E., Sekulic, N., Too, W. C., Konrad, M. & Lavie, A. Elucidation of human choline kinase crystal structures in complex with the products ADP or phosphocholine. *J. Mol. Biol.* **364**, 136–151 (2006).
27 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S, Greenblatt, D. M & Meng, E. C. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
28 Nascimento, A., Jou, C., Ortez, C., Hayashi, Y.K., Nishino, I. & Olivé, M. *et al.* P.15.11 Megaconial congenital muscular dystrophy in two children with mutations in the CHKB Gene. *Neuromuscul. Disord.* **23**, 822 (2013).
29 Behin, A., Laforêt, P., Malfatti, E., Pellegrini, N., Hayashi, Y. & Carlier, R. Y. *et al.* P.15.10 Megaconial myopathy presenting as a progressive limb-girdle myopathy. *Neuromuscul. Disord.* **23**, 821(2013).
30 Mitsuhashi, S. & Nishino, I. Megaconial congenital muscular dystrophy due to loss-of-function mutations in choline kinase β. *Curr. Opin. Neurol.* **26**, 536–543 (2013).
31 Valencia, C.A., Ankala, A., Rhodenizer, D., Bhide, S., Littlejohn, M. R. & Keong, L. M. *et al.* Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS ONE* **8**, e53083 (2013).
32 Brislin, R. P. & Theroux, M. C. Core myopathies and malignant hyperthermia susceptibility: a review. *Paediatr. Anaesth.* **23**, 834–841 (2013).
33 Ku, C. S., Cooper, D. N., Polychronakos, C., Naidoo, N., Wu, M. & Soong, R. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann. Neurol.* **71**, 5–14 (2012).
34 Vasli, N. & Laporte, J. Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders. *Acta Neuropathol.* **125**, 173–185 (2013).

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