A Reliable Targeted Next-Generation Sequencing Strategy for Diagnosis of Myopathies and Muscular Dystrophies, Especially for the Giant Titin and Nebulin Genes

Reda Zenagui, Delphine Lacourt, Henri Pégeot, Kevin Yauy, Raul Juntas Morales, Corine Thèze, Francois Rivier, Claude Cances, Guilhem Solé, Dimitri Renard, et al.

▶ **To cite this version:**

Reda Zenagui, Delphine Lacourt, Henri Pégeot, Kevin Yauy, Raul Juntas Morales, et al.. A Reliable Targeted Next-Generation Sequencing Strategy for Diagnosis of Myopathies and Muscular Dystrophies, Especially for the Giant Titin and Nebulin Genes. Journal of Molecular Diagnostics, 2018, 20 (4), pp.533-549. 10.1016/j.jmoldx.2018.04.001 . hal-01800470

HAL Id: hal-01800470
https://hal.umontpellier.fr/hal-01800470

Submitted on 24 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Reliable Targeted Next-Generation Sequencing Strategy for Diagnosis of Myopathies and Muscular Dystrophies, Especially for the Giant Titin and Nebulin Genes

Reda Zenagui,[*] Delphine Lacourt,[*] Henri Pegeot,[*] Kevin Yauy,[*] Raul Juntas Morales,[†‡] Corine Theze,[*] François Rivier,[‡§¶] Claude Cances,[‡‖] Guilhem Sole,[‡**] Dimitri Renard,[‡††] Ulrike Walther-Louvier,[‡§] Xavier Ferrer-Monasterio,[‡**] Caroline Espil,[‡,‡‡] Marie-Christine Arné-Bes,[‡§§] Pascal Cintas,[‡§§] Emmanuelle Uro-Coste,[‡¶¶] Marie-Laure Martin Negrier,[‡‖‖] Valérie Rigau,[‡***] Eric Bieth,[‡†††] Cyril Goizet,[‡,‡‡‡] Mireille Claustres,[§§§] Michel Koenig,[*§§§] and Mireille Cossée[*‡§§§]

From the Molecular Diagnostic Laboratory and the Departments of Neurology,[†] Pathology,*** and Neuropaediatrics,[§] Centre Hospitalier Universitaire Montpellier, Montpellier; AOC (Atlantique-Occitanie-Caraïbe) Reference Center for Neuromuscular Disorders,[‡] Aquitaine; PhyMedExp,[¶] INSERM, CNRS, and the Rare Diseases Genetics Laboratory,[§§§] Equipe Accueil EA7402, Université de Montpellier, Montpellier; the Departments of Neuropaediatrics,[‖] Neurology,[§§] Pathology,[¶¶] and Genetics,[†††] Centre Hospitalier Universitaire, Toulouse; the Services of Neurology** and Pathology[‖‖] and the Departments of Neuropaediatrics[‡‡] and Genetics,[‡‡‡] Centre Hospitalier Universitaire, Bordeaux; and the Department of Neurology,[††] Centre Hospitalier Universitaire, Nîmes, France*

CME Accreditation Statement: This activity ("JMD 2018 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint providership of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2018 CME Program in Molecular Diagnostics") for a maximum of 18.0 AMA PRA Category 1 Credit(s)™. Physicians should claim only credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Myopathies and muscular dystrophies (M-MDs) are genetically heterogeneous diseases, with >100 identified genes, including the giant and complex titin (*TTN*) and nebulin (*NEB*) genes. Next-generation sequencing technology revolutionized M-MD diagnosis and revealed high frequency of *TTN* and *NEB* variants. We developed a next-generation sequencing diagnostic strategy targeted to the coding sequences of 135 M-MD genes. Comparison of two targeted capture technologies (SeqCap EZ Choice library capture kit and Nextera Rapid Capture Custom Enrichment kit) and of two whole-exome sequencing kits (SureSelect V5 and TruSeq RapidExome capture) revealed best coverage with the SeqCap EZ Choice protocol. A marked decrease in coverage was observed with the other kits, affecting mostly the first exons of genes and the repeated regions of *TTN* and *NEB*. Bioinformatics analysis strategy was fine-tuned to achieve optimal detection of variants, including small insertions/deletions (INDELs) and copy number variants (CNVs). Analysis of a cohort of 128 patients allowed the detection of 52 substitutions, 13 INDELs (including a trinucleotide repeat expansion), and 3 CNVs. Two INDELs were localized in the repeated regions of *NEB*, suggesting that these mutations may be frequent but underestimated. A large deletion was also identified in *TTN* that is, to our knowledge, the first published CNV in this gene.

Myopathies and muscular dystrophies (M-MDs) are a set of phenotypically and genetically heterogeneous diseases, with >100 genes identified so far.[1] Massively parallel sequencing or next-generation sequencing (NGS)

Supported by Agence de la Biomédecine Appel d'Offre Recherche Assistance Médicale à la Procréation, Diagnostic Prénatal et Diagnostic Génétique (M.C.).

Disclosures: None declared.

technologies revolutionized the field of molecular characterization of these heterogeneous disorders, because they allowed sequencing, in a single run, of all currently known genes.[2] In particular, several genes that were not currently analyzed because of their large size, such as the titin gene *TTN* (Online Mendelian Inheritance in Man no. 188840) with 363 exons and the nebulin gene *NEB* (Online Mendelian Inheritance in Man no. 161650) with 183 exons, can now be analyzed exhaustively by NGS.

Compared with whole-exome sequencing (WES), NGS strategies targeted to specific genes (exons and exon-intron junctions) require sequencers and computing resources of lower throughput and generate fewer variants that are restricted to phenotype-related genes and for which laboratories have expertise. Moreover, targeted strategies demonstrate better performance in terms of coverage (depth and uniformity), highlighted by the discovery of 20% to 30% additional variants in comparison with WES in a study of 177 unsolved cases of myopathies.[3] For these reasons, although targeted NGS does not allow detection of mutations in genes not previously implicated in M-MDs, this approach represents a first choice strategy in diagnostic practice.[4] Because there is no commercial kit dedicated for each specific syndrome, implementation of targeted NGS strategy in a diagnostic context requires an in-depth development step to ensure capture of regions of interest in an exhaustive manner. This can be particularly tricky in GC-rich sequences located in first exons of most genes and in repeated regions, such as the tandem repeat regions of *TTN* and *NEB*.[5,6] In *NEB*, the triplicated region of 24 exons is arranged as three sets of eight nearly perfectly repeated exons (82 to 89, 90 to 97, and 98 to 105). The *TTN* gene has three tandem repeats (exons 172 to 180, 181 to 189, and 190 to 198) sharing 99% DNA sequence homology among them.

In contrast to their high sensitivity for detection of substitutions, published bioinformatics pipelines are not fully reliable for insertion/deletion (INDEL) detection.[7] This is detrimental because INDELs are the second most common type of genomic variants after single-nucleotide substitutions,[8] in particular in nebulinopathies.[9] This warrants the use of customized tools associating several alignment and variant calling algorithms to study this class of variations. However, guidelines for optimal detection of biologically significant INDELs are limited.[7,9]

In a previous report of amplicon-based targeted NGS of the *DMD* gene (Multiplicom, Niel, Belgium) on MiSeq (Illumina, San Diego, CA), we reported that the MiSeq Reporter (MSR) software version 2.6.2 provided by Illumina can fail for INDEL calling.[10] SeqNext software (JSI Medical Systems, Ettenheim, Germany) is a private software that performs alignment, variant calling, and variant annotation. The combination of MSR/Variant Studio software version 2.2 (an Illumina variant annotation software) with the SeqNext software version 3.5.0 was a reliable diagnostic strategy for the identification of single-nucleotide variations (SNVs), including INDELs, in the *DMD* gene.[10] With respect to copy number variants (CNVs), most NGS studies do not include CNV detection in their analyses, because of technological limitations, such as capture and PCR bias and reads mappability, which together drastically influence final read coverage for a particular region.[5]

The aim of the present study was to implement a comprehensive and reliable targeted NGS strategy for M-MD diagnosis, with a specific focus on the complex *NEB* and *TTN* genes. Because the clinical presentation does not always point to a specific entity (eg, congenital nemaline myopathy or congenital muscle dystrophy), an NGS strategy analyzing a large panel of genes, including not only the genes reported in the Gene Table of Neuromuscular Disorders (*http://www.musclegenetable.fr*, accessed 2013), but also genes that can lead to atypical phenotypes mimicking M-MDs, was implemented. Two technologies that require low amount of template DNA and that present innovative aspects were chosen to compare: the capture kit SeqCap EZ Choice library, provided by Roche-NimbleGen (Madison, WI), which already demonstrated effectiveness in producing capture probes for comparative genomic hybridization[11]; and the Nextera Rapid Capture Custom Enrichment (NRCCE), from Illumina. An attractive aspect of the Illumina NRCCE kit is the use of a transposase catalyzing DNA fragmentation and ligation of adapter sequences in a single reaction step, referred to as tagmentation.[12,13] This technology allows a technical simplification for library preparation compared with Roche-NimbleGen technology, in which an initial step of DNA sonication is required. Because of the decrease in costs, development of new versions of WES capture kit providing better sequencing coverage,[14,15] acquisition by most molecular diagnostic laboratories of higher-throughput sequencers,[15] and publications of recommendations from international expert committees on the return to patients of incidental finding,[16,17] the relevance of WES compared with custom capture for diagnosis is still an issue. It was evaluated whether two recent WES technologies, SureSelect V5 (Agilent Technologies, Santa Clara, CA) and Illumina TruSeq RapidExome capture, could now be implemented for M-MD diagnosis.

On the basis of our experience and published data, in addition to the MSR/Variant Studio pipeline provided by Illumina, we decided to use the SeqNext software for independent alignment and calling of SNVs, to which we associated the SOPHiA Data Driven Medicine (DDM) software version 4.7.5, recently developed by Sophia Genetics (Ecublens, Switzerland). This software, already reported for molecular diagnosis,[18] has the advantage of detecting and annotating not only SNVs (including INDELs) but also CNVs. It also allows customizable filtering, the generation of a private database of variants, and comparison with variants generated by all SOPHiA DDM users. In addition to the CNV calling algorithm Muskat of the private SOPHiA DDM pipeline, an in-house read

Table 1 Control DNAs with Known Variants Used to Evaluate the Capture Kits and the Bioinformatic Pipeline

Control DNA	Gene	Variant	Type	Status	Confirmatory methods
DNAs sequenced after NRCCE and SeqCap EZ Choice capture kits					
D2712	*DMD*	c.10141C>T	Substitution	Ht	Sanger
D2187	*DMD*	c.3603+2dupT	INDEL	He	Sanger
SR61	*TRIM32*	c.1602delC	INDEL	Ht	Sanger
SR61	*TRIM32*	chr9.hg19:g.119447866_119572263del	CNV (entire gene deletion)	Ht	CGH array
DNAs sequenced only with SeqCap EZ Choice evaluation					
D2867	*DMD*	c.2169-19_2169-3del	INDEL	Ht	Sanger
D2925	*DMD*	Duplication of exon 44	CNV	Ht	MLPA

The nomenclature of variants is indicated in agreement with the following isoforms: NM_004006.2 for *DMD* and NM_12210.3 for *TRIM32* (National Center for Biotechnology Information Nucleotide database; *https://www.ncbi.nlm.nih.gov/nuccore*).

CGH, comparative genomic hybridization; CNV, copy number variant; He, hemizygous; Ht, heterozygous; INDEL, insertion/deletion; MLPA, multiple ligation probe amplification; NRCCE, Nextera Rapid Capture Custom Enrichment.

depth–based approach for CNV detection[19] previously validated in our laboratory on another NGS panel was used.[20] This approach allowed detection of 68 pathogenic or likely pathogenic mutations, including 13 INDELs and 3 CNVs.

Materials and Methods

Study Design and Patients

A specific customized panel of 135 genes (Supplemental Table S1) was designed, including 76 genes associated

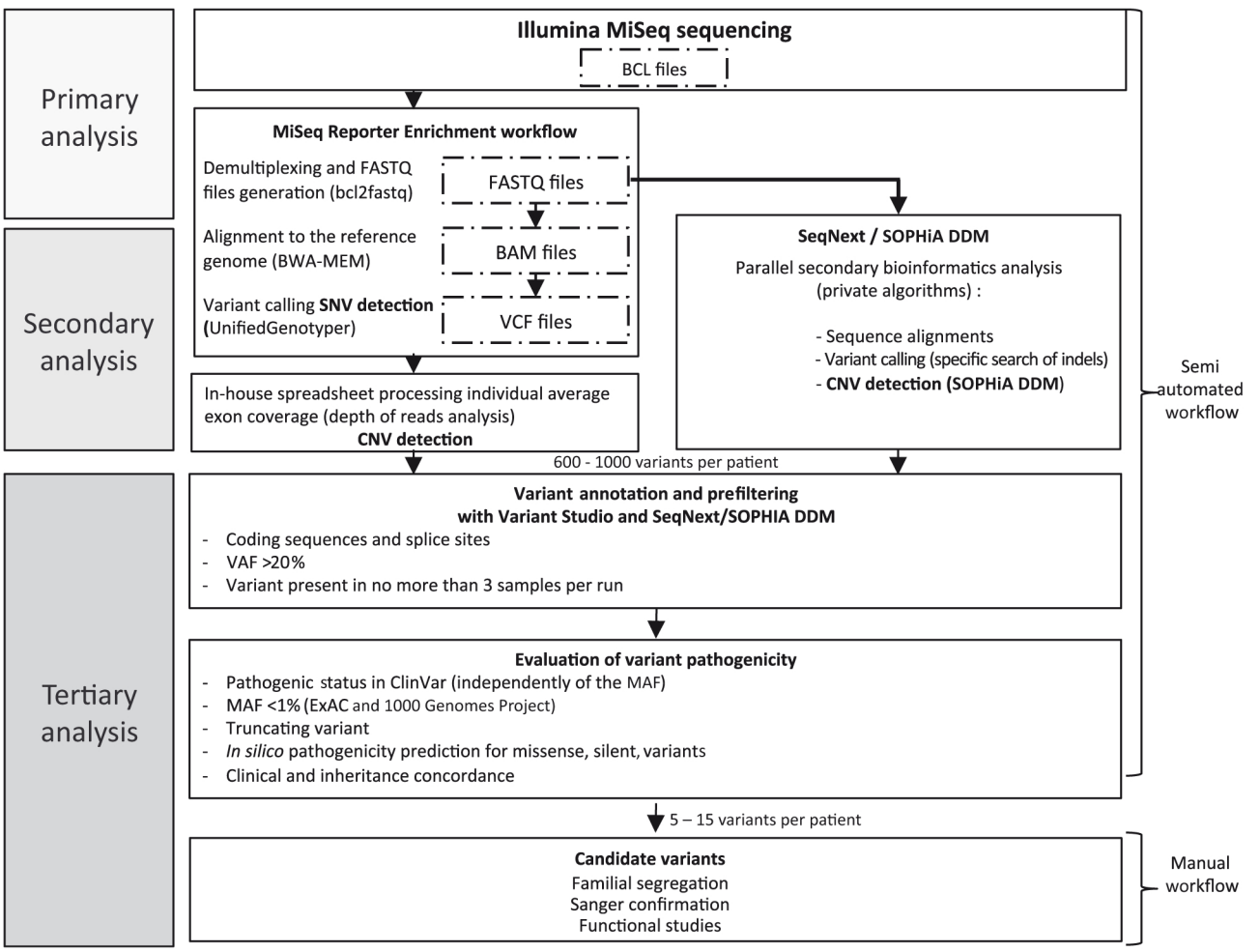


Figure 1 Pipeline for bioinformatics analyses of data generated by targeted next-generation sequencing. BAM, Binary Alignment Map; BCL, Illumina Base Call File; BWA-MEM, Burrows-Wheeler Aligner-Maximal Exact Match; CNV, copy number variant; ExAC, Exome Aggregation Consortium; MAF, minor allele frequency; SNV, single-nucleotide variation; VAF, variant allele frequency; VCF, variant call file.

with congenital myopathies, congenital muscular dystrophies, limb-girdle muscular dystrophies, distal myopathies, and myofibrillar myopathies reported in the Gene Table of Neuromuscular Disorders (*http://www.musclegenetable.fr*) and 59 genes implicated in other neuromuscular diseases, such as congenital myasthenic syndromes, metabolic myopathies, and periodic paralysis, for potential atypical clinical presentations mimicking M-MD. The RefSeq coding sequences NM_accession numbers from University of California, Santa Cruz (*http://hgdownload.cse.ucsc.edu*), reported in the Leiden Open Variation Database (*http://www.lovd.nl/3.0/home*) and determined as consensual for genetic diagnosis within a French nationwide working group (M. Krahn, V. Biancalana, M. Cerino, A. Perrin, L. Michel-Calemard, J. Nectoux, F. Leturcq, C. Bouchet-Séraphin, C. Acquaviva-Bourdain, E. Campana-Salort, A. Molon, J. A. Urtizberea, F. Audic, B. Chabrol, J. Pouget, R. Froissart, J. Melki, J. Rendu, F. Petit, C. Metay, N. Seta, D. Sternberg, J. Fauré, and M. Cossée, unpublished data), were used.

The list curation was performed by the physicians of the French South-West Reference Center for Neuromuscular Diseases. The exons and exon-intron junctions of these 135 genes, corresponding to 1014 kb of sequences of interest (3072 exons and ≈50 bp of exon-intron junctions), were included. Designs were made by taking into consideration size and organization of each gene on the basis of the data from the human genome reference (*http://genome.ucsc.edu/cgi-bin/hgGateway*; GRCh37/hg19). The total size of the regions of interest was too high for PCR-based enrichment, explaining the choice for capture-based enrichment of the regions of interest.

For all patients analyzed in the study (5 controls with known mutations and 128 patients without known mutations), a complete clinical, biochemical, radiological, and histopathological (from muscle biopsy specimen) evaluation was initially performed by the clinicians from the French South-West Reference Center for Neuromuscular Diseases. Informed consents, signed by adult patients or parents (or their legal representatives) for children, were obtained for all samples.

The study was designed in two steps (Supplemental Figure S1). The effectiveness of two targeted capture designs [SeqCap EZ Choice library capture kit (Roche-NimbleGen) and NRCCE kit (Illumina)] and two WES kits [SureSelect V5 Exome (Agilent Technologies) and TruSeq RapidExome (Illumina)] was first evaluated to identify the most efficient one. Each targeted capture kit was evaluated using the quality metrics of NGS data of 12 DNA samples from a MiSeq sequencing run. Each series included the same three control DNAs with known mutations previously identified by classic techniques for diagnostic purposes, for comparison with the identified variants, and nine DNAs from patients suspected of having M-MD without identified genetic etiology. The three control DNAs were as follows (Table 1): D2712 with a heterozygous substitution in the *DMD* gene (NM_004006.2: c.10141C>T), D2187 with a hemizygous INDEL in the *DMD* gene (NM_004006.2: c.3603+2dupT), and SR61 with a heterozygous INDEL

Table 2 Primers Used for cDNA PCR and Sequencing

Gene mutation	Primer PCR1	Primer PCR2	Primers used for sequencing
CAPN3 c.1536+3A>G	F: 5′-TCCTACGAAGCTCTGAAAGG-3′	F: 5′-ACAAGCTTCAGACCTGGACA-3′	F: 5′-ACAAGCTTCAGACCTG-GACA-3′
	R: 5′-CTCCATGTCATCTCCTGCTA-3′	R: 5′-TTCCTCAGAGAGGTTCCTCT-3′	R: 5′-TTCCTCAGAGAGGTTC-CTCT-3′
NEB c.6075+5 G>A	F: 5′-GAACAAGCCAAAGGGAAACACA-3′	F: 5′-TCCCTGGAAGCAGAGAAAAACA-3′	F: 5′-TCCCTGGAAGCAGAGAA-AAACA-3′
	R: 5′-AGCCAAGCCCTTTGTACCAT-3′	R: 5′-TGCGGAAACCAACCATTTTCC-3′	R: 5′-TGCGGAAACCAACCAT-TTTCC-3′
TTN c.51437-4 _51444del	F: 5′-ATAAGCTCGGCTCAGCAACA-3′	F: 5′-CCAGTCATGTCTGGTGAGAACA-3′	F: 5′-ATAGAGAAGATTGCTAA-GGGTGA-3′
	R: 5′-CATTGGCTTTCAAGGCA-TTCAGA-3′	R: 5′-GCTTCCATTATCTTTT-GGTTCATTCCA-3′	R: 5′-CTAAGTGTTGAAGTGACA-ACAATCCA-3′
TTN c.65575+2T>G	F: 5′-GAATGCACGAGTCACCAAAG-3′	F: 5′-CAGTTTGCTGTGGGTGAAAG-3′	F: 5′-CAGTTTGCTGTGGGTG-AAAG-3′
	R: 5′-AGCCAAGCCCTTTGTACCAT-3′	R: 5′-TCTGAGCCTCCATCTTCAAG-3′	F: 5′-CAAAGCATGATGGAGG-CAGT-3′
			R: 5′-CTGAATAGCTCCAAGG-TGCA-3′
TTN c.106531G>C	F: 5′-ACAGATGACGGAGACAAGGG-3′	F: 5′-GAACCAAAAGCTCCTGAACCA-3′	F: 5′-GAACCAAAAGCTCCTGA-ACCA-3′
	R: 5′-GAACACTGGCCACGGAAATT-3′	R: 5′-GGGAGGTTGCTGCTGATTTC-3′	R: 5′-GGGAGGTTGCTGCTGA-TTTC-3′

F, forward; PCR1, first PCR; PCR2, second PCR; R, reverse.

Table 3 Evaluation of Capture Technologies

	Targeted NGS		WES	
Variable	NRCCE (Illumina)	SeqCap EZ Choice (Roche-NimbleGen)	SureSelect version 5 (Agilent Technologies)	TruSeq RapidExome (Illumina)
General characteristics				
Nature of probes	DNA	DNA	RNA	DNA
Probe spacing	Nonoverlapping probes (spacing, 120 bp)	Overlapping probes	Adjacent probes	Nonoverlapping probes (spacing, 120 bp)
DNA fragmentation	Enzymatic (tagmentation)	Mechanical (sonication)	Mechanical (sonication)	Enzymatic (tagmentation)
Optimal size of fragmented DNA	250–350 bp	180–220 bp	150–200 bp	180–200 bp
Bioinformatic development of the probe design	Design Studio version March 2014	NimbleDesign version 3.0 + Bioinformatics team	NA	NA
Theoretical specificity	100% Unique hits	98% Unique hits 0.9% Two hits 1.1% Three hits		
Theoretical mean deep sequencing of ROIs	390×	400×	NA	NA
Quality metrics of the experimental runs				
ON target/total target, %	82	80	66.40	51.90
Duplicate reads, %	49	5	18.40	40.65
Observed mean deep sequencing of ROIs	198×	380×		
Coverage ≥50×, %	88.70	98.52		
Coverage ≥20×, %	95.2	99.03		

NA, not available; NGS, next-generation sequencing; NRCCE, Nextera Rapid Capture Custom Enrichment; ROI, region of interest; WES, whole-exome sequencing.

(NM_12210.3: c.1602delC) and an entire deletion of the *TRIM32* gene (chr9.hg19:g.119447866_119572263del) previously validated by comparative genomic hybridization array.[21] The nine patients without known mutations were different for each capture kit. The NGS data of the 12 samples in each kit allowed the evaluation of the capture metrics, independently of the identified variants. WES capture kits were evaluated on NGS data from a NextSeq 500 (Illumina) sequencing run on nine DNA samples without known mutation for the TruSeq RapidExome kit and on nine different DNA samples without known mutation from a HiSeq (Illumina) sequencing run for the SureSelect V5 Exome kit (Agilent Technologies)

For all these experiments, only DNA samples of high quality without any sign of either degradation or protein contamination were used.

The efficiency of the bioinformatics pipeline (detailed below) was then assessed for detection of SNVs and CNVs. The NGS data from the two runs of targeted NGS described above (SeqCap EZ Choice library and NRCCE kits), including the 3 control DNAs with known mutations and 18 DNA samples without known mutations, were used. After selecting SeqCap EZ Choice library capture kit on the basis of quality metrics, NGS was performed on additional samples and the data were analyzed. Two additional control DNAs with known mutations were sequenced to assess the ability of the general workflow to identify mutations that are classically considered difficult to detect: a heterozygous deletion of 17 bp in the DMD gene (D2867; NM_004006.2: c.2169-19_2169-3del), previously identified by Sanger sequencing; and a heterozygous duplication of a single exon in the *DMD* gene (D2925; exon 44 duplication), previously detected by multiple ligation probe amplification (Table 1). A total of 110 additional DNAs were also tested from patients suspected of having M-MDs without identified mutations after analysis of a small number of genes or without prior genetic analysis

DNA Extraction

Genomic DNA was extracted from blood samples following the manufacturer's standard procedure of the FlexiGene DNA kit (Qiagen, Courtaboeuf, France). DNA concentrations were measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE), and DNA quality was assessed by agarose/ethidium bromure 1% electrophoresis.

NGS

Each capture library was constructed following the manufacturer's instructions. For SeqCap EZ Choice capture protocol, optimal DNA fragmentation (average fragment size of 180 to 220 bp) was achieved by using the Bioruptor Pico (Diagenode, Liège, Belgium) sonication instrument with the following program: 30 seconds ON/30 seconds

OFF during 13 cycles of sonication. Before sequencing, the quality of the final libraries was assessed on a Bioanalyzer High Sensitivity DNA chip (Agilent Technologies) and quantified with a Qubit High Sensitivity kit (Invitrogen by Life Technologies, Carlsbad, CA). The purified and individually tagged libraries were pooled equimolarly following the manufacturer's procedure, along with an internal quality control (PhiX; Illumina).

For targeted NGS, paired-end sequencing (2 × 150 bp) was performed on a V3 flowcell using a MiSeq sequencer (Illumina), according to manufacturer's instructions. Twelve libraries were multiplexed per run.

The nine DNA samples enriched using Illumina TruSeq RapidExome Library Kit were sequenced by 2 × 150 bp paired-end runs on a NextSeq 500 sequencer (Illumina), according to manufacturer's instructions. The nine libraries obtained from SureSelect Human All Exon V5 Kit (Agilent Technologies) were sequenced on 96 samples run on HiSeq 2000 sequencer (Illumina).

Bioinformatics

In the case of targeted NGS, primary bioinformatics analyses were performed using MSR version 2.6.2 (Illumina). In this pipeline, samples demultiplexing and generation of FASTQ files were performed by the bcl2fastq Conversion Software version 2.18, whereas sequence alignment against the human reference sequences was performed with the Burrows-Wheeler Aligner-Maximal Exact Match (BWA-MEM) algorithm.[22] Variant calling was done using the Genome Analysis Tool Kit UnifiedGenotyper[23] (Figure 1). Base target coverage was reported by Picard (*http://broadinstitute.github.io/picard*, last accessed January 2017). FASTQ files were also processed using the patented algorithms of the SeqNext software version 3.5.0 (JSI Medical Systems) and the SOPHiA DDM software version 4.7.5 (Sophia Genetics, Saint-Sulpice, Switzerland) for independent read alignment and variant calling. The generated variant call files were analyzed by VariantStudio software version 2.2 (Illumina), SeqNext, and SOPHiA DDM software for annotation and first-step filtering of the variants.

For CNV analyses, an in-house bioinformatics spreadsheet that computes the intersample normalized depth of coverage per exon in a given run was implemented, as previously reported.[5] First, intrasample normalization was performed by dividing the number of reads covering each exon by the mean of the read number of all other exons of all genes of the sample to obtain the ratio of patient. The same calculation was made for the other samples of the sequencing run (means) to obtain the ratio of controls. An intersample normalization was then performed, by calculation of the ratio of patient/ratio of controls, to obtain the ratio relative coverage (RRC) for the analyzed exon. An exon was considered deleted/duplicated when the RRC fell within predefined intervals (homozygous deletion RRC < 0.1, heterozygous deletion $0.3 <$ RRC < 0.7, heterozygous duplication $1.3 <$ RRC ≤ 1.75, and homozygous duplication $1.75 <$ RRC < 2.4). Deletion or duplication of consecutive exons was highly suggestive of a CNV. To exclude possible artifacts in case of abnormal ratio of a single exon, interpretation of pathogenicity was considered only under specific settings, which included total read number per exon >300, absence of abnormal ratio in the other DNA samples, and identification of an SNV on the other allele in the case of an autosomal recessive disorder. An independent search for CNVs was also performed from the FASTQ files using SOPHiA DDM software version 5.3.1 that includes the CNV calling algorithm Muskat.

Table 4 Exons with Coverage <10× after Targeted NGS with NRCCE or SeqCap EZ Capture Kits

	Mean deep sequencing (number of reads) for the control DNAs D2712, D2187, and SR61	
Gene-exon	NRCCE library (Illumina)	SeqCap EZ Choice library (Roche-NimbleGen)
HSPG2—exon 1	0	3
SEPN1—exon 1	0	8.5
MTM1—exon 1	0	44
MYBPC3—exon 11	0	761.9
COL6A2—exon 16	0	852.4
SCGB—exon 1	0	134.2
ASTN2/TRIM32—exon 1	0	221
ITGA7—exon 5	0.4	1260.6
TNNT3—exon 7	1.4	1582.9
FHL1—exon 1	1.8	186.4
ABHD5—exon 1	3	509.2
Bin1—exon 11	6	845.5
DLG4—exon 1	7	1206.1
LARGE—exon 9	8	1847.9

NGS, next-generation sequencing; NRCCE, Nextera Rapid Capture Custom Enrichment.

For WES, data were analyzed using a custom informatics pipeline on the basis of Broad Institute Genome Analysis Tool Kit Best Practices.[24] Design analysis and exon coverage analysis were made using BEDtools (Quinlan Lab, Salt Lake City, UT; *http://bedtools.readthedocs.io*), with intersect and multicov (minimum of 30 reads for mapping quality) function. Quality analysis and base target coverage were reported by CollectHSmetrics on Picard tools.

Evaluation of Variant Pathogenicity

First, all variants showing >1% of minor allele frequency in any population [on the basis of the 1000 Genomes Project (*http://www.1000genomes.org*, last accessed 2017), the Exome Variant Server from National Heart, Lung, and Blood Institute Gene Ontology Exome Sequencing Project (*http://evs.gs.washington.edu/EVS*, last accessed 2017), and Exome Aggregation Consortium (*http://exac.broadinstitute.org*, last

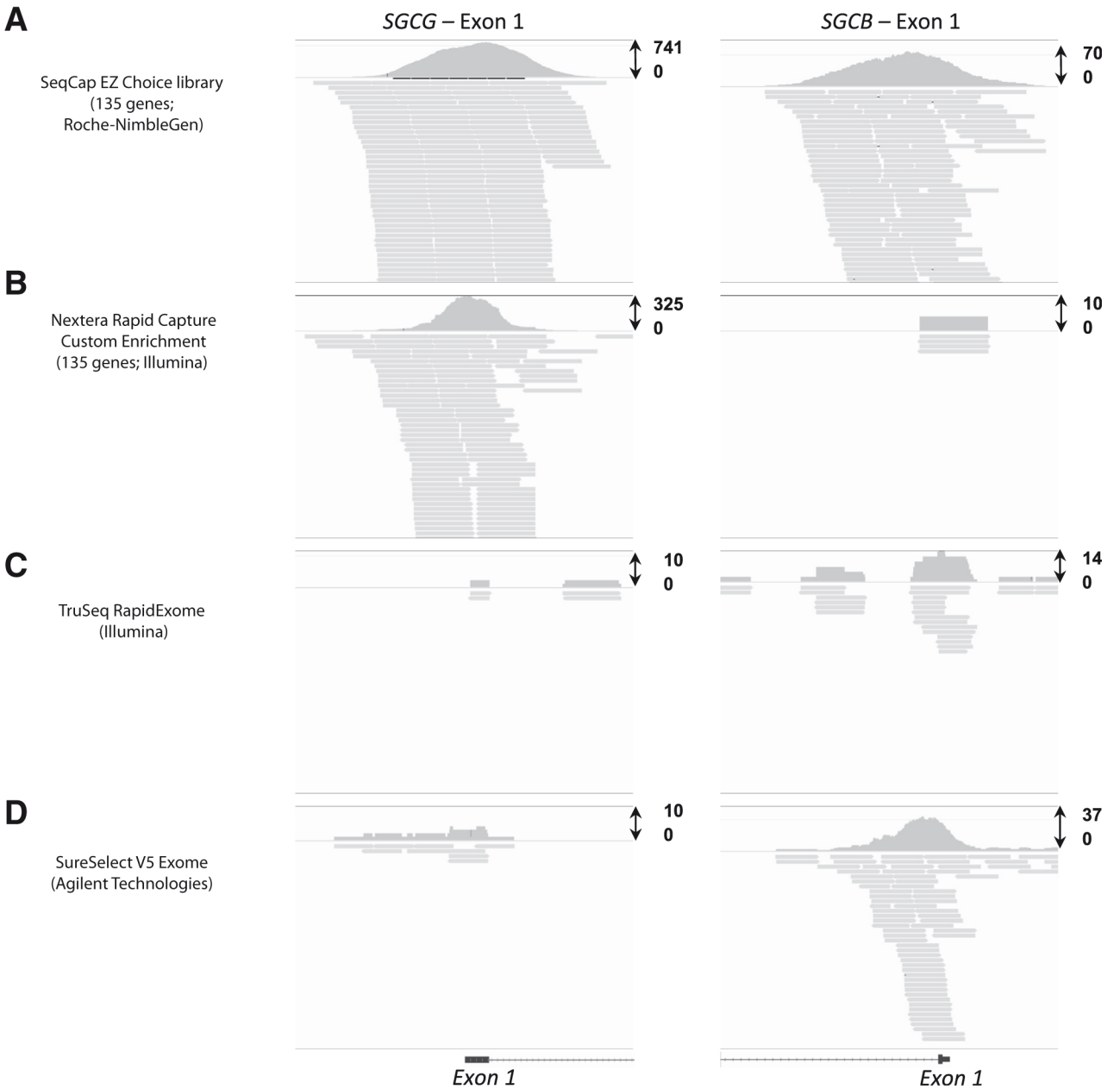


Figure 2 Integrative Genomic Viewer visualization of *SGCG* and *SGCB*—exon 1 reads after targeted next-generation sequencing using targeted (**A** and **B**) and whole-exome (**C** and **D**) sequencing kits. There is no coverage of the exon 1 of *SGCB* with Nextera Rapid Capture Custom Enrichment design (**B**) and low coverage (<10×) for Illumina TruSeq RapidExome (**C**). *SGCG*—exon 1 is almost not covered after both whole-exome sequencing kits [TruSeq RapidExome (Illumina) and SureSelect V5 Exome (Agilent Technologies)].

accessed 2017)] were excluded. Then, only variants within coding and splice site regions, with a variant allele frequency (VAF) >20% of total reads and present in less than three patients for each run, were considered (Figure 1). Independently, to detect possible pathogenic variants even with an allelic frequency >1%, a selection of variants already described as pathogenic in ClinVar, regardless of their allelic frequency, was set up.

The pathogenicity of the candidate variants was then assessed through a set of criteria, according to the American College of Medical Genetics and Genomics guidelines[16]: variant known as pathogenic [Human Gene Mutations database (*http://www.hgmd.cf.ac.uk*), ClinVar (*http://www.ncbi.nlm.nih.gov/clinvar*), Leiden Muscular Dystrophy database (*http://www.dmd.nl*), Online Mendelian Inheritance in Man (*http://www.ncbi.nlm.nih.gov/omim*), and PubMed (*http://www.ncbi.nlm.nih.gov*)]; *in silico* prediction tools [Sift version 1.03 (*http://sift.jcvi.org*), PolyPhen-2 version 2.2.2 (r394) (*http://genetics.bwh.harvard.edu/pph2*), Align Grantham Variation Grantham Deviation (*http://agvgd.iarc.fr*), and Mutation Taster (*http://www.mutationtaster.org*)]; and functional domain of

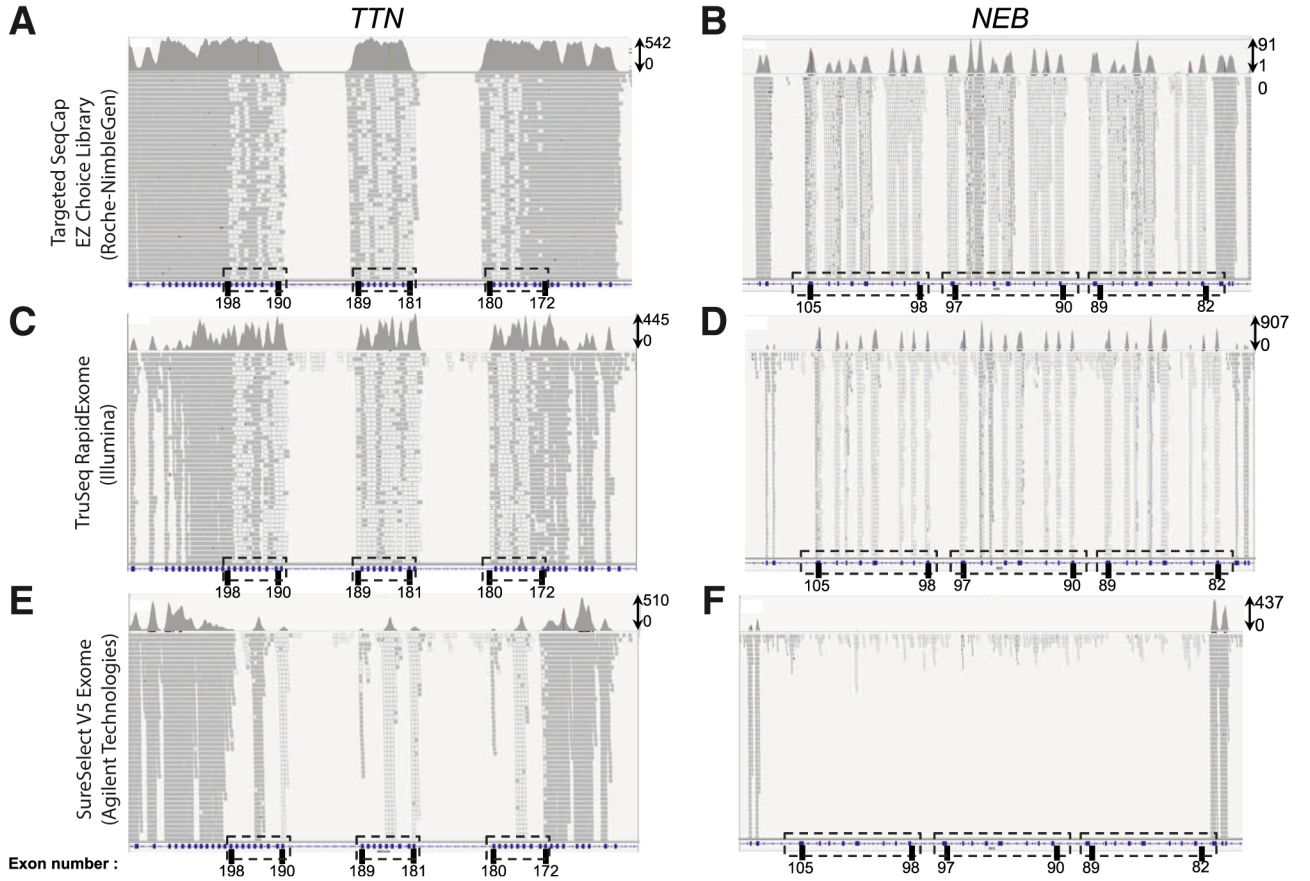


Figure 3 Integrative Genomic Viewer (IGV) visualization of *NEB* and *TTN* repeated region reads after targeted next-generation sequencing and whole-exome sequencing using different capture kits. IGV view of read alignment of *TTN* (NM_001267550; **A**, **C**, and **E**) and *NEB* (NM_001271208; **B**, **D**, and **F**) repeated regions, according to three capture methods: Targeted SeqCap EZ Choice library (Roche-NimbleGen), TruSeq RapidExome (Illumina), and SureSelect V5 Exome (Agilent Technologies). Groups of repeated exons are framed in **dashed boxes** in the reference gene track. White reads indicate reads with a mapping quality of 0, meaning that they could be located in more than one genomic location. Reads correctly mapped are represented in gray. In each case, depth of coverage is shown in the upper right corner.

the protein [UniProt release 2016_11 (*http://www.uniprot.org*)]. The Alamut Visual software version 2.9 (Interactive Biosoftware, Rouen, France), which integrates several of these *in silico* prediction criteria (*http://www.interactivebiosoftware.com*), was used. Because *in silico* prediction tools are unreliable for the interpretation of the functional impact of variants in the *TTN* gene, only truncating variants (nonsense variants, frameshifts, variants affecting splicing, and CNVs) that were absent or with a frequency $<1\%$ in the general population were considered as pathogenic or likely pathogenic.

Concordance with the patient's phenotype and with the suspected mode of inheritance was also considered. Best candidate variants were then discussed in multidisciplinary meetings with clinicians and other specialists involved in this project (pathologists, radiologists, and biochemists). Clinical reassessment of the patient and/or of relatives, familial segregation studies, and additional analyses, such as specific immunolabeling on muscle biopsy, were performed if needed, to support or invalidate these variants. If NGS technology revealed a candidate variant in a gene involved in a recessive disease and that is relevant with the patient phenotype, the sequencing coverage of this region, and the immunolabeling data of the corresponding protein were evaluated more precisely. Poorly covered regions (<20 reads) in candidate genes were then analyzed by Sanger sequencing. If protein immunolabeling was abnormal, molecular RNA analyses from a muscle biopsy specimen were offered to search for a deep intronic mutation affecting splicing of the mRNA.

Confirmatory Methods

Sanger sequencing confirmation of candidate variants and familial segregation analyses were performed by classic Sanger sequencing protocols. Bidirectional sequencing of the purified PCR products was performed on an Applied Biosystem 3130XL automated capillary sequencer (Applied Biosystems, Foster City, CA) (primers and protocols available on request).

For CNVs, confirmation was performed using multiple ligation probe amplification (MRC Holland, Amsterdam,

the Netherlands; *https://www.mlpa.com*), semiquantitative fluorescent PCR, and/or sequencing of the junction fragment.

For variants predicted to affect splicing, cDNA analyses of the specific genes were performed on mRNAs extracted from muscle biopsy specimens. Muscle mRNA was extracted using 30 mg of muscle with the Kit SV Total RNA Isolation System (Promega, Madison, WI), and cDNA synthesis was performed using SuperScript II First-Strand Synthesis System for RT-PCR (Invitrogen by Life Technologies) and random hexamer primers (Invitrogen by Life Technologies). The targeted transcripts were amplified and sequenced by Sanger sequencing: the first PCR was performed in a reaction volume of 25 μL containing 12.5 μL of 2× Master Mix buffer containing Taq DNA Polymerase, dNTPs, and Reaction Buffer (Promega), 2 μL of forward primer (5 mmol/L), 2 μL of reverse primer (5 mmol/L), 1 μL of cDNA template, and 7.5 μL of H_2O. After a 2-minute denaturation at 94°C, PCR amplification was performed using the following cycle profile: denaturation at 94°C for 30 seconds, 35 cycles of annealing at 59°C for 30 seconds, extension at 72°C for 30 seconds, and extension at 72°C for 8 minutes. The second PCR used the first PCR products as template and was performed in the same way as the first PCR. The second PCR products were purified using a QIAquick PCR Purification Kit (Qiagen, Courtaboeuf, France). The second PCR purified products were used for sequencing reactions, according to standard protocols (BigDye Terminator version 1.1 Cycle Sequencing Kit; Applied Biosystems). Primers used for cDNA PCR and sequencing are reported in Table 2.

Table 5 CNV Detection with the In-House Bioinformatic Script in Control DNAs after SeqCap EZ Choice Library Capture Kit and MiSeq Sequencing

Variable	Value
Detection of a heterozygous deletion of the entire *TRIM32* (SR61 patient DNA)	
Exon 1	
Reads in the patient DNA, *n*	51.7
ROP	0.094
Mean of reads in control DNAs, *n*	95.1
ROC	0.223
RRC	0.42
Exon 2	
Reads in the patient DNA, *n*	326.9
ROP	0.594
Mean of reads in control DNAs, *n*	492.8
ROC	1.156
RRC	0.51
Detection of a heterozygous duplication of *DMD* exon 44 (D2925 patient DNA)	
Exon 44	
No. of reads in the patient DNA	188.6
ROP	0.731
Mean of reads in control DNAs, *n*	402.4
ROC	1.070
RRC	1.54
Exon 45	
Reads in the patient DNA, *n*	168.1
ROP	0.652
Mean of reads in control DNAs, *n*	420.2
ROC	1.117
RRC	1.04

CNV, copy number variant; ROC, ratio of controls; ROP, ratio of patient; RRC, ratio relative coverage.

Results

Capture Kit Evaluations

The general characteristics of the capture kits are detailed in Table 3. Close interactions with bioinformatics support of Roche-NimbleGen allowed an optimization of the SeqCap EZ Choice probe design. In the initial design, only the probes that did not have homology with other genomic regions were retained. This high stringency resulted in 45 exons theoretically not properly covered, corresponding mainly to the repeated exons in the *TTN* and *NEB* genes. Therefore, a more relaxed design was chosen in which the probes with two other homologous genomic regions were tolerated. The theoretical coverage of regions at 50× was thus improved from 98.04% to 99.1%.

The effectiveness of targeted capture SeqCap EZ Choice (Roche-NimbleGen) and NRCCE (Illumina) kits was evaluated after NGS analysis of 12 DNA samples on a MiSeq sequencing run for each kit, including three identical control samples with known mutations and nine different samples without known mutations (*Materials and Methods*). To evaluate the WES capture kits, nine DNA samples without known mutations were analyzed using SureSelect 5 (Agilent Technologies) and nine other ones using TruSeq RapidExome (Illumina) capture kits. To reduce experimental costs, most DNA samples used for capture metric analyses, which is independent of the identified variants, were different between each kit (*Materials and Methods*). To overcome the risk of DNA-dependent differences, only DNA samples of high quality were used.

After sequencing, a high score of quality (Phred score ≥30 = 90%) was obtained for libraries from both capture kits. The proportion of sequenced bases within targeted regions (on target) was similar for SeqCap EZ Choice and NRCCE (80% and 82%, respectively). The two methods appeared to have a similar mean depth in the target regions of approximately 400×. However, a significant difference with respect to the proportion of PCR duplicates was detected. For SeqCap EZ Choice design, only 5.2% of reads were identified as PCR duplicate, whereas with NRCCE design, the percentage of PCR duplicates was higher, amounting to 49.3% of reads. When duplicates were removed from each result, the effective coverage decreased. A major reduction was observed for the NRCCE design,

Table 6 Pathogenic or Likely Pathogenic Mutations Identified by Targeted NGS in M-MD Patients

Gene[†]	Mutation	Type of mutation	Effect of mutation
ACTA1	c.493 G>C; p.(Val165Leu)	Substitution	Missense
NM_001100.3	c.1057A>G; p.(Thr353Ala)	Substitution	Missense
	c.553C>T; p.(Arg185Cys)	Substitution	Missense
	c.889G>A; p.(Ala297Thr)	Substitution	Missense
ANO5	c.656A>G; p.(Tyr219Cys)	Substitution	Missense
NM_213599.2	c.1991T>C; p.(Phe664ser)	Substitution	Missense
	c.1627dup; p.(Met543AsnFs*)	INDEL	Frameshift
CACNA1S	c.5104C>T; p.(Arg1702*)	Substitution	Nonsense
NM_000069	c.2970G>A; p.(Trp990*)	Substitution	Nonsense
CAPN3	c.1250C>T; p.(Thr417Met)	Substitution	Missense
NM_000070.2	c.1536+3A>G	Substitution	Splicing[‡]
COL6A1	c.299_428+152delinsGACAGGA	CNV	Splicing
NM_001848.2	c.G806A; p.(Gly269Glu)	Substitution	Missense
COL6A3	c.5035G>T; p.(Gly1679Trp)	Substitution	Missense
NM_004369			
DMD	c.2665C>T; p.(Arg889*)	Substitution	Nonsense
NM_004006.2			
DYSF	c.1168G>A; p.(Asp390Asn)	Substitution	Missense
NM_003494.3	c.5302C>T; p.(Arg1768Trp)	Substitution	Missense
	c.755C>T; p.(Thr252Met)	Substitution	Missense
	c.3118C>T; p.(Arg1040Trp)	Substitution	Missense
	c.2665C>T; p.(Arg889*)	Substitution	Nonsense
FLNC	c.4927+2T>A	Substitution	Splicing
NM_001458.4			
FKRP	c.1364C>A; p.(Ala455Asp)	Substitution	Missense
NM_024301.4	c.1384C>T; p.(Pro462Ser)	Substitution	Missense
GMPPB	c.902C>G; p.(Ser301Cys)	Substitution	Missense
NM_013334.2	c.1069G>A; p.(Val357Ile)	Substitution	Missense
LAMA2	c.752T>C; p.(Leu251Pro)	Substitution	Missense
NM_000426.3	c.1884+1_1884+6del	INDEL	Splicing
	c.2461A>C; p.(Thr821Pro)	Substitution	Missense
	c.3976C>T; p.(Arg1326*)	Substitution	Nonsense
LMNA	c.1357C>T; p.(Arg453Trp)	Substitution	Missense
NM_170707.3			
MTM1	c.519C>G; p.(Tyr173*)	Substitution	Nonsense
NM_000252.2	c.290G>A; p.(Gly97Glu)	Substitution	Missense
	c.1261-10A>G	Substitution	Splicing
MYH7	c.5791-1G>T (*LP*)	Substitution	Splicing
NM_000257			
NEB	c.6075+5G>A	Substitution	Splicing[‡]
NM_001271208.1	c.13661_13666delinsA; p.(Ser4554Asnfs*10) [or c.15119_15124delinsA; p.(Ser5040Asnfs*10)]	INDEL	Frameshift
	c.23998dupG; p.(Glu8000Glyfs*11)	INDEL	Frameshift
	c.518delA; p.(Lys173Serfs*55)	INDEL	Frameshift
	c.21928T>C p.(Ser7310Pro)	Substitution	Missense
	c.8860del; p.(Ala2954Profs*8)	INDEL	Frameshift
	c.5039G>A; p.(Tyr1680Cys)	Substitution	Missense
	c.17541dupA; p.(Tyr5848Ilefs*15)	INDEL	Frameshift
	c.13720delC; p.(His4574Thrfs*11) [or c.15178 delC; p.(His5060Thrfs*11)]	INDEL	Frameshift
	c.2835+5G>A	Substitution	Splicing
	c.21790G>C; p.(Asp7264His) (*LP*)	Substitution	Missense
	c.194C>T; p.(Pro65Leu) (*LP*)	Substitution	Missense
PABPN1	c.30_32dupAGC; p.(Ala11dup) (*LP*)	INDEL (trinucleotide repeat expansion)	Alanine repeat expansion
NM_004643.3			
RYR1	c.11186T>C; p.(Met3729Thr)	Substitution	Missense

(*table continues*)

Table 6 (*continued*)

Gene[†]	Mutation	Type of mutation	Effect of mutation
NM_000540.2	c.11122A>C; p.(Thr3708Pro)	Substitution	Missense
	c.38T>C; p.(Leu13Pro)	Substitution	Missense
	c.7628C>T; p.(Thr2543Ile)	Substitution	Missense
	c.13690C>T; p.(Arg4564Trp)	Substitution	Missense
	c.14762T>C; p.(Phe4921Ser)	Substitution	Missense
	c.12727G>A; p.(Glu4243Lys)	Substitution	Missense
	c.5926C>T; p.(Arg1976Cys)	Substitution	Missense
	c.10649G>C; p.(Arg3550Pro)	Substitution	Missense
	c. 880 G>A; p.(Glu294Lys)	Substitution	Missense
SGCA NM_000023.2	c.850C>T; p.(Arg284Cys)	Substitution	Missense
SGCG NM_000231.2	c.800_801del; p.(Cys267Serfs*51)	INDEL	Frameshift
TNNT1 NM_003283.5	c.192+244_388-1191del	CNV	Dele8-9
TTN NM_001267550.1	c.51437-4_51444del	INDEL	Splicing[‡]
	c.26503A>T; p.(Lys8835*)	Substitution	Nonsense
	c.65575+2T>G	Substitution	Splicing[‡]
	c.105036C>A; p.(Tyr35012*)	Substitution	Nonsense
	c.106531G>C; p.(Ala35511Pro)	Substitution	Splicing[‡]
	c.1662+15_3101-3del	CNV	Dele11-18
	c.89900_89903delATTA; p.(Asn29967Metfs*27)	INDEL	Frameshift
	c.6379_6380delTA; p.(Tyr2127Leufs*8)	INDEL	Frameshift

CNV, copy number variant; Dele, deletion; INDEL, insertion/deletion; LP, likely pathogenic; M-MD, myopathy and muscular dystrophy; NGS, next-generation sequencing.

[†]National Center for Biotechnology Information Nucleotide database (*https://www.ncbi.nlm.nih.gov/nuccore*).

[‡]Confirmed by cDNA studies (data not shown).

with a loss of 150× (390× to 240×), whereas there was a moderate loss of 22× for the NimbleGen design, leading to an effective mean depth of sequencing of 380×. Thus, 98.5% of the target regions were covered by ≥50× after SeqCap EZ Choice Roche capture (99% for 20×), whereas only 88.7% were covered by ≥50× with the NRCCE Illumina kit (95% for 20×) (Table 3 and). Several regions covered with the SeqCap EZ Choice library capture kit (Figure 2A) were not covered with the NRCCE targeted design (Table 4 and Figure 2B). Most of these noncovered regions are located in the first exons of the genes, known to be rich in GC nucleotides. All pathogenic mutations of the three control DNA samples were detected after SeqCap EZ Choice kit capture. On the other hand, the *TRIM32* gene deletion was not detected after capture with the NRCCE kit, because of the absence of exon 1 capture (this gene being composed of two exons). Comparison of the number of all variants detected by NGS in the three control DNAs after each capture kit revealed that 1075 substitutions and 71 INDELs were detected after capture with both kits. However, 116 substitutions and 18 INDELs were detected only with SeqCap EZ Choice kit capture, probably because of a better coverage of regions of interest

Similarly to the targeted designs, a higher rate of duplicate reads was observed for WES with Illumina TruSeq RapidExome kit (40.65%) compared with the SureSelect V5 kit (Agilent Technologies) (18.40%) (Table 3). Because sequencing was not performed with the same sequencers for the two kits, the reading depths were not compared, but only the homogeneity of coverage of the target sequences was studied. With both WES kits, most of the poorly captured regions in the M-MD genes also belong to the first exon of genes (Figure 2, C and D).

Concerning the repeated regions of *TTN* (exons 172 to 180, 181 to 189, and 190 to 198) and *NEB* (exons 82 to 89, 90 to 97, and 98 to 105), both targeted capture kits allowed correct coverage, although mapping of sequences in one of the three repeated segments was not optimal. On the other hand, these regions were poorly or not covered with the WES Agilent SureSelect V5 kit (Figure 3).

The better coverage obtained with the SeqCap EZ Choice library capture kit (Roche-NimbleGen) warrants the choice of targeted NGS with this technology for diagnostic applications.

Bioinformatic Pipeline Evaluation

Three bioinformatics software programs, MSR version 2.2.31 (Illumina), SeqNext version 3.5.0 (JSI Medical Systems), and SOPHiA DDM version 4.7.5 (Sophia Genetics), were used for independent alignment and variant calling of data generated by targeted MiSeq sequencing of M-MD genes. For CNV analyses, an in-house bioinformatics script, based on the depth of sequencing for each

exon and the specific algorithm Muskat contained within the SOPHiA DDM software, was used (*Materials and Methods*).

To evaluate the effectiveness of the bioinformatics pipeline for detection of SNVs and CNVs, the NGS data from five control DNAs with known mutations were first analyzed: the three DNAs analyzed for comparison of the targeted kits (D2712 with a heterozygous substitution in *DMD*, D2187 with a 1-bp duplication in *DMD*, and SR61 with compound heterozygosity for a 1-bp deletion and an entire deletion of *TRIM32*) and two additional DNAs with known mutations that are classically considered of difficult detection [D2867 with a heterozygous deletion of 17 bp in the *DMD* gene and D2925 with a heterozygous duplication of a single exon (exon 44) of the *DMD* gene] (Table 1). All SNVs, including the 17-bp deletion, were correctly identified with MSR, SeqNext, and SOPHiA DDM. Reliable CNV detection required a coverage depth >300×. *TRIM32* exon 1 was not covered by NGS after capture by the NRCCE kit (Table 4) and, consequently, the deletion of this exon (SR61 sample) was not detected. It was covered with low read depth by NGS after capture by the SeqCap EZ Choice library kit (52×). The in-house bioinformatics script detected the deletion (Table 5), whereas SOPHiA DDM rejected the SR61 sample from CNV analyses, because of the presence of numerous artifacts. The duplication of *DMD* exon 44 in D2925 sample was correctly detected on data from NGS after capture by both targeted kits, with the in-house software (Table 5) and SOPHiA DDM.

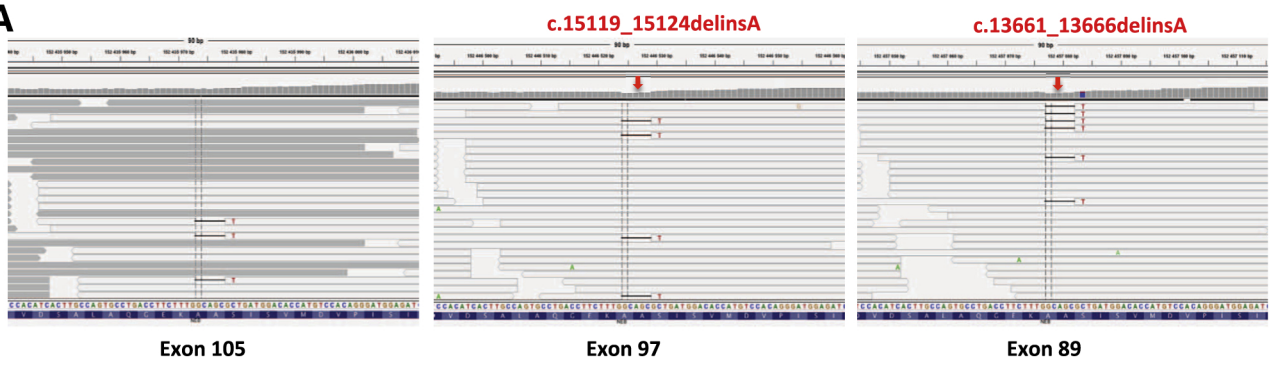
NGS data from 128 patients affected by M-MDs without identified genetic etiology were also analyzed . On average, 900 variants were detected for each patient. After semiautomated filtering, the average number of candidate variants was approximately 5 to 15 per patient. The pathogenicity of these candidate variants was assessed through a set of criteria, according to American College of Medical Genetics and Genomics guidelines[16] (*Materials and Methods*) (Figure 1).

Overall, 68 pathogenic or likely pathogenic mutations were identified: 52 substitutions, 13 INDELs (including a trinucleotide repeat expansion in *PABPN1*), and 3 CNVs, all confirmed by an independent method (*Materials and Methods*) (Table 6). All 52 substitutions were detected with MSR, SeqNext, and SOPHiA DDM. With respect to the INDELs, 10 of 13 were identified with the three software programs. Two INDELs located within the repeated regions of *NEB* were not detected by MSR. In the first sample, the INDEL was a deletion of 6 bp associated with a one-base insertion. Because exons 89 and 97 are identical, it was not possible to localize this variant in exon 89 [c.13661_13666delinsA; p.(Ser4554Asnfs*10)] or exon 97 [c.15119_15124delinsA; p.(Ser5040Asnfs*10)] (Figure 4). Exon 105, within the third repeated region, contains one nucleotide that differs with sequences from exons 89 and 97, allowing correct mapping of most reads of exon 105. The INDEL was detected by SeqNext and SOPHiA DDM in exons 89 and 97, but with a VAF of approximately 25% in each of the two exons, instead of 50% for a single exon (Figure 4), probably because of random alignment on either copy of the duplication. It was checked whether the absence of INDEL detection by MSR was because of misalignment or variant calling failure. An independent analysis of the aligned Binary Alignment Map (BAM) files generated by MSR on SeqNext software detected the INDEL, suggesting that the absence of INDEL detection was attributable to failure of MSR variant calling. Similar to this first case of *NEB* INDEL not detected by MSR, it was not possible to locate the second *NEB* INDEL in exon 89 [c.13720del; p.(His4574Thrfs*11)] or exon 97 [c.15178 del; p.(His5060Thrfs*11)], and the corresponding VAF was also approximately 25% in each exon with both SeqNext and SOPHiA DDM analyses. MSR also failed to detect a trinucleotide repeat expansion of one repeat in *PABPN1* [c.30_32dupAGC; p.(Ala11dup)]. This expansion was detected by SeqNext and SOPHiA DDM with a VAF of 25% of the reads, probably because of misalignment within the 11 repeated triplets. Sanger sequencing confirmed that the variant was heterozygous in the patient (Corinne Métay and Pascale Richard, personal communication, 2017). This expansion, (GCN)11, has been reported in late-onset dominant or in recessive forms of oculopharyngeal muscular dystrophy.[21,25]

Variant analysis of an NGS run (12 libraries) by SOPHiA DDM detected five false-positive SNVs (ie, not confirmed by Sanger sequencing) among 1845 variants. All of these false positives had a VAF of <30% and were located within homopolymeric or dinucleotide repeat sequences.

Three CNVs were identified among the 128 samples with the in-house bioinformatics spreadsheet and SOPHiA DDM, and all were further confirmed by sequencing of the junction fragment. A heterozygous deletion of exon 3 of *COL6A1* was detected by the in-house bioinformatics spreadsheet, indicating an RRC of 0.53 (Figure 5A). Specific search for CNVs in *COL6* genes was performed in this case because of phenotypic and familial data highly suggestive of collagenopathy and the absence of identified SNVs in these genes. Interestingly, Integrative Genomic Viewer version 2.3.97 (University of California San Diego, La Jolla, CA; *http://software.broadinstitute.org/software/igv*) visualization and SOPHiA DDM indicated that the proximal breakpoint was located within exon 3 (Figure 5B). Sanger sequencing confirmed that the deletion, extending over 282 bp, involved only the 130 bp distal of exon 3 (NM_001848.2: c.299_428+152delinsGACAGGA) (Figure 5C). Familial segregation of the INDEL was consistent with the autosomal dominant transmission of the disease (Figure 5D). A heterozygous deletion of exons 11 to 18 of the *TTN* gene was also identified in a mother and her daughter with distal myopathy (Figure 6). Because the breakpoints were close to exon-intron junctions, SOPHiA DDM

A



B

	Reads with the variation/all reads (%)		
	Exon 105	Exon 97	Exon 89
MSR	Not detected	Not detected	Not detected
IGV	3.7	21.7	26.6
SeqNext	0	27	26
SOPHIA DDM	0	24.2	23.5

Figure 4 Results of an insertion/deletion (INDEL) detection in *NEB* triplicated exons. **A:** Integrative Genomic Viewer (IGV) visualization. White reads indicate reads that have mapping quality = 0. Reads correctly mapped are represented in gray. **Arrows** indicate the DELINS location. Because exons 89 and 97 are identical, it was not possible to localize this variant in exon 89 [c.13661_13666delinsA; p.(Ser4554Asnfs*10)] or exon 97 [c.15119_15124delinsA; p.(Ser5040Asnfs*10)]. Exon 105 contains one different nucleotide, allowing correct mapping of reads. **B:** Percentage of reads with the INDEL detected by the different software programs. The INDEL is not detected by MiSeq Reporter (MSR) and is reported in exons 89 and 97 with a variant allele frequency of approximately 25% by SeqNext and SOPHiA DDM.

reported the precise breakpoints (NM_001267550.1: c.1662+15_3101-3del), that were confirmed by Sanger sequencing of the junction fragment. A homozygous deletion of exons 8 and 9 was identified in the *TNNT1* gene of a patient with congenital myopathy. The deletion was inherited from both parents, who were consanguineous. Sanger sequencing of the junction fragment confirmed the deletion as homozygous in the patient and heterozygous in the parents, and allowed precise localization of the breakpoints (NM_003283.5: c.192+244_388-1191del).

SOPHiA DDM analyses, performed on four NGS runs for software validation, indicated that CNVs restricted to first exons of genes had a higher probability of being artifacts. As for SNVs, CNVs within regions harboring ambiguous read mapping were hard to interpret. A false-positive CNV in the *NEB* repeated regions, a heterozygous deletion that was not confirmed by custom comparative genomic hybridization array analysis (Vilma Lehtokari, personal communication, 2017), was identified.

Discussion

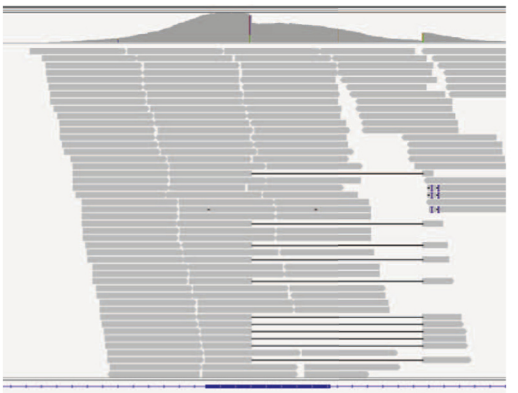
We report the development of a suitable targeted NGS strategy for detection of mutations in M-MD genes, including the complex *NEB* and *TTN* genes, and the availability of CNV detection. The possibility to customize the capture design was of high interest for improving capture efficiency, in particular for the repeated regions of *TTN* and *NEB*. A relaxed probe set design containing probes with up to two close matches in the genome was chosen. Only then, genomic regions from the repeated exons of *TTN* and *NEB* were captured and sequenced.

Comparison of capture technologies showed a higher rate of duplicate reads with both Illumina capture kits (NRCCE targeted capture kit and TruSeq RapidExome capture kit). The high fraction of duplicate reads sometimes reflects a DNA quality problem.[26] However, this was not the case in this study because only DNA samples of high quality (without any signs of degradation or protein contamination) were used. The best coverage was achieved with SeqCap EZ Choice technology (Roche-NimbleGen), with 98.5% coverage at a depth of 50× and 99% coverage at a depth of 20×. Another strength of the capture solution proposed by Roche-NimbleGen is the use of the Kapa Biosystems technology. A recent study[27] comparing nine commercial kits for the preparation of DNA libraries showed that Kapa Biosystems kits achieved better performances in terms of library diversity, duplicate rates, and regularity of coverage. Sequencing coverage was lower after capture with NRCCE targeted capture kit and with the two tested WES kits [TruSeq RapidExome

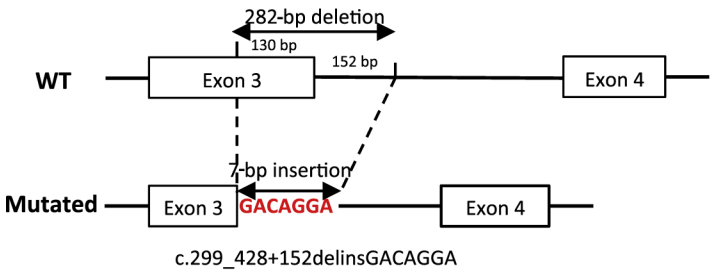
A

				RRC			
Chr	Start	Stop	Localization	Sample 1	Sample 2	Sample 3	...
...	...	...	...	...	...	...	...
21	47401613	47401911	COL6A1:NM_001848_exon_1	1.13	0.91	1.09	...
21	47402498	47402727	COL6A1:NM_001848_exon_2	0.99	0.98	1.01	...
21	47404133	47404433	**COL6A1:NM_001848_exon_3**	0.90	**0.53**	1.17	...
21	47406390	47406649	COL6A1:NM_001848_exon_4	0.98	0.92	1.01	...
21	47406808	47407027	COL6A1:NM_001848_exon_5	0.97	0.97	1.07	...
21	47407028	47407139	COL6A1:NM_001848_exon_6	0.91	1.02	0.96	...
...	...	...	...	...	...	...	...

B



C



D

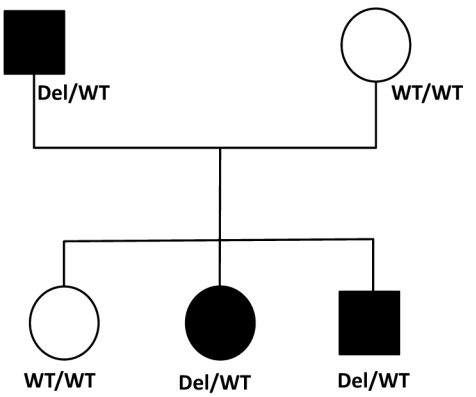


Figure 5 Heterozygous deletion of part of exon 3 in the *COL6A1* gene. **A:** The in-house bioinformatics script for copy number variant analysis: The shaded bold cells show a ratio relative coverage of 0.53 for the *COL6A1* exon 3 in patient I144 (sample 2), suggesting a heterozygous deletion of this exon. **B:** Integrative Genomic Viewer visualization of the deletion indicates that the proximal breakpoint is located within exon 3 and allows breakpoint targeting. **C:** Confirmation and genomic characterization of the *COL6A1* deletion. Precise breakpoint determination and junction fragment sequences were obtained by Sanger sequencing. The genomic alteration was a deletion of 282 bp, including 130 bp in exon 3, and an insertion of 7 bp at the fragment junction (NM_001848.2: c.299_428+152delinsGACAGGA). **D:** Familial segregation of the deletion. The *c.299_428+152delinsGACAGGA* allele is represented as Del, and the wild-type allele is represented as WT. The mutated allele was also identified in the affected brother and father, whereas the healthy sister and mother were not carriers. Chr, chromosome.

(Illumina) and SureSelect V5 WES (Agilent Technologies)], with several regions having complete absence of coverage. These regions were located mainly in the first exons of genes, known to be rich in GC. Among them, some have been described as the site of pathogenic CNVs, such as deletions of the entire *TRIM32* gene,[21] of *MTM1* exon 1,[28] of *SGCG* exon 1,[29] and of partial duplications of *SGCB* exon 1.[29] Although some of the noncovered exons are noncoding, mutations in these exons can lead to lack of transcription.[27,30] The repeated regions of *TTN* and *NEB* were also less or not covered with SureSelect V5 WES kit. All of these results were detracting for the choice of the NRCCE (Illumina) targeted kit and for the first-line use of the TruSeq RapidExome (Illumina) and SureSelect V5 WES commercial WES enrichment kits for diagnosis, because they may lead to loss of a subsequent amount of important coding regions for M-MD diagnosis.

However, the requirement for regular updating of gene panels after the identification of novel genes responsible for M-MDs represents a limitation of the targeted approach. Since the implementation of our targeted NGS strategy in 2013, 25 additional genes have been implicated in M-MDs.[31] Even genes that were reported only in single cases or in small cohorts will be added to our next custom probe design, to be as comprehensive as possible.

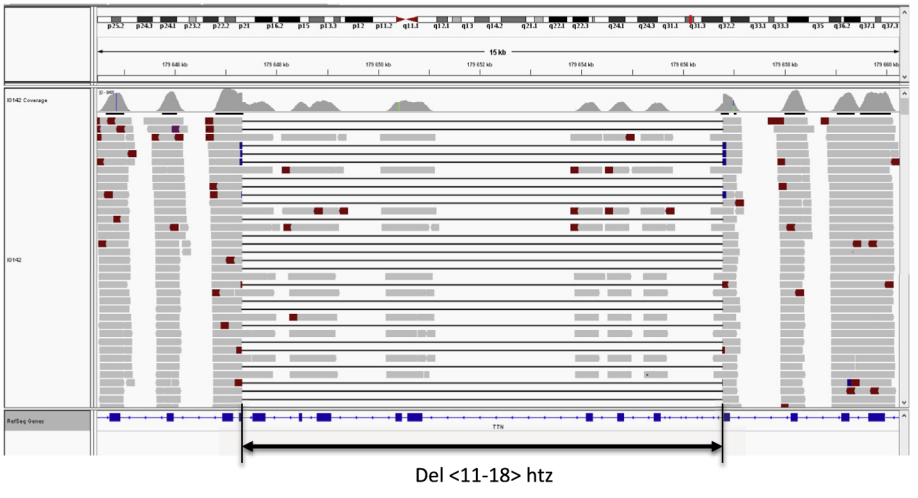
Deep intronic sequences were not included in the custom targeted NGS design. A whole genome sequencing strategy should allow additional detection of deep intronic pathogenic variant(s) affecting splicing. However, it would require higher-performance sequencers, higher computational power for bioinformatics analysis, and high-throughput mRNA sequencing technologies to assess the impact of intronic variants on splicing. Our strategy is to perform a first-line muscle biopsy with immunolabeling of specific proteins, depending on the phenotype. If NGS technology identifies a candidate variant in a gene involved in a recessive disease that is relevant with patient phenotype, potential intronic mutations affecting splicing could be

A

Reads number | RRC

Chr	Start	Stop	RegionID	Labels	Mean	I40	I119	I120	I123	I129	I142	I162	I220	I270	I288	I338	I40	I119	I120	I123	I129	I142	I162	I220	I270	I288	I338	I339
2	179644677	179644982	42611387	TTNNM_001267550-22	**618.4**	443.1	672.7	645.5	688	543.8	601.7	602.2	564.5	690.7	639.9	574.8	1.04	1.04	0.99	0.95	1.00	0.88	1.08	0.92	1.09	0.99	1.04	1.01
2	179645798	179646040	42611388	TTNNM_001267550-21	**623.2**	439.6	676.1	647.3	768.8	551	639.6	578.7	579.1	579.6	648.8	600.4	1.03	1.03	0.99	1.06	1.01	0.93	1.03	0.94	0.89	1.00	1.08	1.02
2	179646889	179647204	42611389	TTNNM_001267550-20	**676.7**	405.7	678.3	731.6	755.7	569.9	782.1	599.1	695	713.3	700	650	0.87	0.95	1.03	0.95	0.96	1.06	0.98	1.05	1.02	0.99	1.08	1.03
2	179647216	179647379	42611390	TTNNM_001267550-19	**606.3**	413.1	701.9	564	738.3	537.5	538.6	482.4	673	622.5	689.1	558	0.99	1.11	0.87	1.05	1.01	0.80	0.87	1.14	0.99	1.10	1.03	1.04
2	179647483	179647841	42611391	TTNNM_001267550-18	**652.8**	395.9	665.8	649.6	737.5	538.4	314.7	284.6	596	646.8	672.9	542.9	1.03	1.14	1.11	1.14	1.10	0.49	0.54	1.08	1.12	1.16	1.09	1.04
2	179648397	179648562	42611392	TTNNM_001267550-17	**571.3**	460.4	679.8	627.3	721.7	503.2	311.6	280.5	598.7	576.7	667.7	592	1.19	1.14	1.05	1.09	1.01	0.47	0.52	1.07	0.98	1.13	1.17	1.24
2	179648747	179649128	42611393	TTNNM_001267550-16	**553.3**	408.1	629.8	576.2	690.6	522	346.3	281.4	578.2	618.5	638.6	583.8	1.08	1.09	0.99	1.08	1.08	0.55	0.54	1.07	1.09	1.12	1.20	1.16
2	179650297	179650519	42611394	TTNNM_001267550-15	**594.7**	447.9	669.5	685.6	754.8	576.5	348.5	293.1	658.5	666.7	684.6	580	1.10	1.08	1.11	1.10	1.12	0.51	0.52	1.14	1.09	1.12	1.10	1.08
2	179650525	179650918	42611395	TTNNM_001267550-14	**567.0**	406	620.1	656.2	724.9	536.5	338.1	281.9	624.3	649.4	656.3	554.3	1.05	1.04	1.11	1.11	1.09	0.52	0.53	1.13	1.12	1.12	1.10	1.11
2	179654037	179654274	42611396	TTNNM_001267550-13	**524.8**	431.2	618.4	643.6	648.7	534.3	306.4	247.4	530.3	612.8	532.8	514.6	1.21	1.13	1.18	1.07	1.18	0.51	0.50	1.03	1.14	0.97	1.11	1.07
2	179654655	179654892	42611397	TTNNM_001267550-12	**520.5**	393.7	615	590.9	661.8	458.7	306.9	240.6	548.7	627.2	588.5	527	1.11	1.14	1.09	1.10	1.01	0.51	0.49	1.08	1.19	1.09	1.15	1.10
2	179655385	179655622	42611398	TTNNM_001267550-11	**534.9**	388.9	626.2	604.3	716.8	517.3	290.9	248.1	529.1	611.9	624.1	550.2	1.06	1.12	1.09	1.17	1.11	0.47	0.49	1.01	1.12	1.13	1.17	1.11
2	179656749	179656974	42611399	TTNNM_001267550-10	**598.7**	383.5	675.5	616.6	725.9	539	503.5	498.3	620.8	607.8	658.3	560	0.93	1.08	0.98	1.04	1.03	0.75	0.91	1.06	0.98	1.06	1.05	1.11
2	179658081	179658318	42611400	TTNNM_001267550-9	**634.2**	473.9	716.8	664.6	672.9	568.2	704.4	555.3	611.4	671.4	630.9	564.1	1.09	1.09	1.00	0.90	1.03	1.02	0.96	0.98	1.03	0.95	0.99	1.01
2	179659076	179659328	42611401	TTNNM_001267550-8	**598.7**	413.8	604.7	604.7	653.7	494.4	612.4	550.6	523.1	621.2	613.4	532.5	1.04	0.99	0.99	0.96	0.97	0.96	1.05	0.91	1.04	1.01	1.03	1.06

B



C

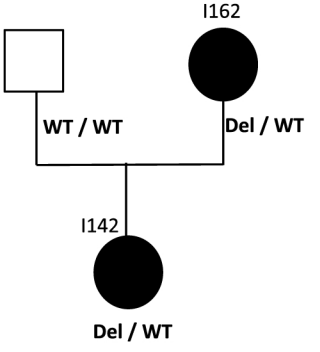


Figure 6 Heterozygous deletion of exons 11 to 18 of the *TTN* gene. **A:** The in-house bioinformatics script for copy number variant (CNV) analysis shows a ratio relative coverage (RRC) of 0.47 to 0.55 for the *TTN* exons 11 to 18 in I162 (affected mother sample) and I142 (affected daughter; RRC highlighted in yellow) compared with the other DNA samples (I40, I119, I120, I123, I129, I220, I270, I288, I338, and I339), suggesting a heterozygous deletion of these exons. The read numbers of these samples are indicated. **B:** Integrative Genomic Viewer visualization of the deletion shows that the distal breakpoint is located close to exon 19 and allows breakpoint targeting. Precise breakpoint determination was obtained by Sanger sequencing (NM_001267550: c.1662+15_3101-3del). The distal one is located in the acceptor splice site of exon 19, suggesting a pathogenic effect on exon 19 splicing. **C:** Sanger sequencing confirms the cosegregation of the CNV with the disease. The CNV allele is represented as Del, and the wild-type allele is represented as WT. Chr, chromosome.

searched by analyses of the specific transcript from a muscle biopsy specimen.

Diagnosis guidelines recommend the use of two independent bioinformatics pipelines with different alignment and variant calling algorithms as a safety measure to ensure the most comprehensive identification of variants. This is particularly relevant for INDEL detection. Although no false-negative result was observed for substitutions, MSR failed to detect some INDELs, as reported previously.[10,32,33] This can be explained by the used variant caller, UnifiedGenotyper of Genome Analysis Tool Kit.[34] Indeed, UnifiedGenotyper was shown to call only a small number of INDELs from a gold standard benchmark data set.[34] Another study reported that UnifiedGenotyper does not perform well for large (>5-bp) INDEL calling, because it could only call 52% of them on a benchmark data set.[7,35] However, in this study, the lack of detection was restricted to the two INDELs located in the repeated regions of *NEB* (one of which being a deletion of only 1 bp) and to a trinucleotide repeat expansion in *PABPN1* gene. Incorrect read mapping in the *NEB* repeated regions, and misalignment within the 11 repeated triplets in *PABPN1*, leading in all cases to aberrantly low VAF, may also contribute to failure of INDEL detection. Interestingly, only two mutations within the *NEB* repeated exons are reported in the literature, of >150 reported patients,[9,36] whereas two such mutations were identified among the 12 *NEB* mutations of our cohort. The limited number of reported cases with mutation within the repeated exons may reflect underdetection because of defective capture and/or variant calling. Our results, along with the high frequency of INDELs in nebulinopathies,[9] suggest that the *NEB* triplicated region could be a hot spot for INDELs that are currently underdetected.

With respect to CNV detection, this strategy was capable of detection of a heterozygous single-exon duplication (control sample) that represents the most difficult type of CNV to identify, and even of detection of a partial deletion of exon 3 of the *COL6A1* gene in one patient with highly suggestive phenotypic and familial data. However, because of the high level of false positives for single-exon CNVs, particularly for first exons, phenotypic and genetic data are

important to consider for pointing toward a specific gene. A large heterozygous deletion in *TTN* was also found in a mother and her daughter with distal myopathy. The deletion of exons 11 to 18 (c.1662+15_3101-3del) is predicted to lead to out-of-frame transcripts. However, the distal breakpoint is located in the acceptor splice site of exon 19, suggesting a pathogenic effect on exon 19 splicing. If the implication of this deletion in the phenotype of the patient is confirmed (functional analyses are in progress), it will be the first CNV reported as pathogenic in the *TTN* gene. The lack of previously reported deletions in *TTN* is surprising, given the large size of the gene. This may be explained by the incomplete sequencing of *TTN* coding sequence and by the difficulty for detection of CNVs on NGS data. The third CNV identified in our series was a homozygous deletion of exons 8 and 9 in the *TNNT1* gene encoding for the slow skeletal muscle isoform of troponin T. This is a predicted in-frame deletion, reminiscent of a previously reported mutation at the splice donor site of exon 8, leading to an in-frame skipping of exon 8 in *TNNT1* transcripts.[37]

Some studies reported difficulties for detection of mutation in the repeated exons of *TTN* and *NEB*.[5,6] The difficulties probably result from the combination of capture failure, incorrect mapping of sequences, and INDEL underdetection within triplicated exons. We showed that the use of a customized capture kit and of different alignment and variant calling software programs allowed the detection of variants located in the repeated exons, but without precise localization in one of the three repeats. Recent long-read sequencing technologies could be interesting in this respect because, most of the time, they allow unambiguous mapping on repeated regions.[38]

Although NGS targeted on large panel of genes does not allow identification of novel genes or detection of deep-intronic mutations, it represents an efficient first-step approach for genetic diagnosis of patients with M-MDs. Molecular diagnostic laboratories are now able to offer a growing catalog of analyzed M-MD genes, therefore improving the diagnostic efficiency for these genetically heterogeneous diseases, in particular for incomplete or atypical phenotypes.[39] This is of high importance for patient care, genetic counseling, and M-MD classification.

Acknowledgments

We thank John Rendu (Centre Hospitalier Universitaire, Grenoble, France), Pascale Richard and Corinne Métay (Hôpital Pitié-Salpétrière, Assistance Publique, Hôpitaux de Paris, Paris, France), Valérie Biancalana (Centre Hospitalier Universitaire, Strasbourg, France), Nathalie Seta and Céline Bouchet-Séraphin (Hôpital Bichat, Assistance Publique, Hôpitaux de Paris, Paris, France), France Leturcq and Juliette Nectoux (Hôpital Cochin, Assistance Publique, Hôpitaux de Paris, Paris, France), and Vilma Lehtokari and Bjarne Udd (University of Helsinki, Helsinki, Finland) for expert advice on the interpretation of variant pathogenicity, Pascal Richard and Corinne Métay for confirmation of *PABPN1* repeat expansion by Sanger analysis, and Vilma Lehtokari for confirmation study of *NEB* copy number variant by custom comparative genomic hybridization array.

References

1. Nigro V, Savarese M: Genetic basis of limb-girdle muscular dystrophies: the 2014 update. Acta Myol 2014, 33:1−12
2. Vasli N, Böhm J, Le Gras S, Muller J, Pizot C, Jost B, Echaniz-Laguna A, Laugel V, Tranchant C, Bernard R, Plewniak F, Vicaire S, Levy N, Chelly J, Mandel J-L, Biancalana V, Laporte J: Next generation sequencing for molecular diagnosis of neuromuscular diseases. Acta Neuropathol 2012, 124:273−283
3. Savarese M, Di Fruscio G, Mutarelli M, Torella A, Magri F, Santorelli FM, Comi GP, Bruno C, Nigro V: MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples. Acta Neuropathol Commun 2014, 2:100
4. Weiss MM, Van der Zwaag B, Jongbloed JDH, Vogel MJ, Brüggenwirth HT, Lekanne Deprez RH, Mook O, Ruivenkamp CAL, van Slegtenhorst MA, van den Wijngaard A, Waisfisz Q, Nelen MR, van der Stoep N: Best practice guidelines for the use of next-generation sequencing applications in genome diagnostics: a National Collaborative Study of Dutch Genome Diagnostic Laboratories. Hum Mutat 2013, 34:1313−1321
5. Evilä A, Arumilli M, Udd B, Hackman P: Targeted next-generation sequencing assay for detection of mutations in primary myopathies. Neuromuscul Disord 2016, 26:7−15
6. Fernández-Marmiesse A, Carrascosa-Romero MC, Alfaro Ponce B, Nascimento A, Ortez C, Romero N, Palacios L, Jimenez-Mallebrera C, Jou C, Gouveia S, Couce ML: Homozygous truncating mutation in prenatally expressed skeletal isoform of TTN gene results in arthrogryposis multiplex congenita and myopathy without cardiac involvement. Neuromuscul Disord 2017, 27:188−192
7. Ghoneim DH, Myers JR, Tuttle E, Paciorkowski AR: Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. BMC Res Notes 2014, 7:864
8. Mullaney JM, Mills RE, Pittard WS, Devine SE: Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet 2010, 19: R131−R136
9. Lehtokari V-L, Kiiski K, Sandaradura SA, Laporte J, Repo P, Frey JA, Donner K, Marttila M, Saunders C, Barth PG, den Dunnen JT, Beggs AH, Clarke NF, North KN, Laing NG, Romero NB, Winder TL, Pelin K, Wallgren-Pettersson C: Mutation update: the spectra of nebulin variants and associated myopathies. Hum Mutat 2014, 35: 1418−1426
10. Alame M, Lacourt D, Zenagui R, Mechin D, Danton F, Koenig M, Claustres M, Cossée M: Implementation of a reliable next-generation sequencing strategy for molecular diagnosis of dystrophinopathies. J Mol Diagn 2016, 18:731−740
11. Vasson A, Leroux C, Orhant L, Boimard M, Toussaint A, Leroy C, Commere V, Ghiotti T, Deburgrave N, Saillour Y, Atlan I, Fouveaut C, Beldjord C, Valleix S, Leturcq F, Dodé C, Bienvenu T, Chelly J, Cossée M: Custom oligonucleotide array-based CGH: a reliable diagnostic tool for detection of exonic copy-number changes in multiple targeted genes. Eur J Hum Genet 2013, 21:977−987

12. Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD: Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. Genome Res 2012, 22:125−133
13. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P: Improved workflows for high throughput library preparation using the transposome-based Nextera system. BMC Biotechnol 2013, 13:104
14. Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, Xu Z, Steinmann B, Carrel T, Röthlisberger B, Schlapbach R, Bruggmann R, Matyas G: New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res 2015, 43:e76
15. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C: Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. Hum Mutat 2015, 36:815−822
16. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee: Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015, 17:405−424
17. Hehir-Kwa JY, Claustres M, Hastings RJ, van Ravenswaaij-Arts C, Christenhusz G, Genuardi M, Melegh B, Cambon-Thomsen A, Patsalis P, Vermeesch J, Cornel MC, Searle B, Palotie A, Capoluongo E, Peterlin B, Estivill X, Robinson PN: Towards a European consensus for reporting incidental findings during clinical NGS testing. Eur J Hum Genet 2015, 23:1601−1606
18. Lucarelli M, Porcaro L, Biffignandi A, Costantino L, Giannone V, Alberti L, Bruno SM, Corbetta C, Torresani E, Colombo C, Seia M: A new targeted CFTR mutation panel based on next-generation sequencing technology. J Mol Diagn 2017, 19:788−800
19. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z: Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013, (14 Suppl 11):S1
20. Baux D, Vaché C, Blanchet C, Willems M, Baudoin C, Moclyn M, Faugère V, Touraine R, Isidor B, Dupin-Deguine D, Nizon M, Vincent M, Mercier S, Calais C, García-García G, Azher Z, Lambert L, Perdomo-Trujillo Y, Giuliano F, Claustres M, Koenig M, Mondain M, Roux AF: Combined genetic approaches yield a 48% diagnostic rate in a large cohort of French hearing-impaired patients. Sci Rep 2017, 7:16783
21. Nectoux J, de Cid R, Baulande S, Leturcq F, Urtizberea JA, Penisson-Besnier I, Nadaj-Pakleza A, Roudaut C, Criqui A, Orhant L, Peyroulan D, Ben Yaou R, Nelson I, Cobo AM, Arné-Bes M-C, Uro-Coste E, Nitschke P, Claustres M, Bonne G, Lévy N, Chelly J, Richard I, Cossée M: Detection of TRIM32 deletions in LGMD patients analyzed by a combined strategy of CGH array and massively parallel sequencing. Eur J Hum Genet 2015, 23:929−934
22. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009, 25:1754−1760
23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010, 20:1297−1303
24. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA: From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics 2013, 43:11.10.1−11.10.33
25. Richard P, Trollet C, Stojkovic T, de Becdelievre A, Perie S, Pouget J, Eymard B; Neurologists of French Neuromuscular Reference Centers CORNEMUS and FILNEMUS: Correlation between PABPN1 genotype and disease severity in oculopharyngeal muscular dystrophy. Neurology 2017, 88:359−365
26. Shigemizu D, Momozawa Y, Abe T, Morizono T, Boroevich KA, Takata S, Ashikawa K, Kubo M, Tsunoda T: Performance comparison of four commercial human whole-exome capture platforms. Sci Rep 2015, 5:12742
27. Aigrain L, Gu Y, Quail MA: Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays: a systematic comparison of DNA library preparation kits for Illumina sequencing. BMC Genomics 2016, 17:458
28. Herman GE, Kopacz K, Zhao W, Mills PL, Metzenberg A, Das S: Characterization of mutations in fifty North American patients with X-linked myotubular myopathy. Hum Mutat 2002, 19:114−121
29. Trabelsi M, Kavian N, Daoud F, Commere V, Deburgrave N, Beugnet C, Llense S, Barbot JC, Vasson A, Kaplan JC, Leturcq F, Chelly J: Revised spectrum of mutations in sarcoglycanopathies. Eur J Hum Genet 2008, 16:793−803
30. Rouzier C, Le Guédard-Méreuze S, Fragaki K, Serre V, Miro J, Tuffery-Giraud S, Chaussenot A, Bannwarth S, Caruba C, Ostergaard E, Pellissier J-F, Richelme C, Espil C, Chabrol B, Paquis-Flucklinger V: The severity of phenotype linked to SUCLG1 mutations could be correlated with residual amount of SUCLG1 protein. J Med Genet 2010, 47:670−676
31. Kaplan J-C, Hamroun D: The 2015 version of the gene table of monogenic neuromuscular disorders (nuclear genome). Neuromuscul Disord 2014, 24:1123−1153
32. Strom CM, Rivera S, Elzinga C, Angeloni T, Rosenthal SH, Goos-Root D, Siaw M, Platt J, Braastadt C, Cheng L, Ross D, Sun W: Development and validation of a next-generation sequencing assay for BRCA1 and BRCA2 variants for the clinical laboratory. PLoS One 2015, 10:e0136419
33. García-García G, Baux D, Faugère V, Moclyn M, Koenig M, Claustres M, Roux A-F: Assessment of the latest NGS enrichment capture methods in clinical context. Sci Rep 2016, 6:20948
34. Hasan MS, Wu X, Zhang L: Performance evaluation of indel calling tools using real short-read data. Hum Genomics 2015, 9:20
35. Fang H, Wu Y, Narzisi G, O'Rawe JA, Jimenez Barrón LT, Rosenbaum J, Ronemus M, Iossifov I, Schatz MC, Lyon GJ: Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med 2014, 6:89
36. Malfatti E, Lehtokari V-L, Böhm J, De Winter JM, Schäffer U, Estournet B, Quijano-Roy S, Monges S, Lubieniecki F, Bellance R, Viou MT, Madelaine A, Wu B, Taratuto AL, Eymard B, Pelin K, Fardeau M, Ottenheijm CAC, Wallgren-Pettersson C, Laporte J, Romero NB: Muscle histopathology in nebulin-related nemaline myopathy: ultrastrastructural findings correlated to disease severity and genotype. Acta Neuropathol Commun 2014, 2:44
37. van der Pol WL, Leijenaar JF, Spliet WGM, Lavrijsen SW, Jansen NJG, Braun KPJ, Mulder M, Timmers-Raaijmakers B, Ratsma K, Dooijes D, van Haelst MM: Nemaline myopathy caused by TNNT1 mutations in a Dutch pedigree. Mol Genet Genomic Med 2014, 2:134−137
38. Guo X, Zheng S, Dang H, Pace RG, Stonebraker JR, Jones CD, Boellmann F, Yuan G, Haridass P, Fedrigo O, Corcoran DL, Seibold MA, Ranade SS, Knowles MR, O'Neal WK, Voynow JA: Genome reference and sequence variation in the large repetitive central exon of human MUC5AC. Am J Respir Cell Mol Biol 2014, 50: 223−232
39. Savarese M, Di Fruscio G, Tasca G, Ruggiero L, Janssens S, De Bleecker J, Delpech M, Musumeci O, Toscano A, Angelini C, Sacconi S, Santoro L, Ricci E, Claes K, Politano L, Nigro V: Next generation sequencing on patients with LGMD and nonspecific myopathies: findings associated with ANO5 mutations. Neuromuscul Disord 2015, 25:533−541