Genetic characterization of a French cohort of GNE-mutation negative inclusion body myopathy patients using exome sequencing

Mathieu CERINO, PharmD,1,2 Svetlana GOROKHOVA, MD, PhD,1 Pascal LAFORET, MD, PhD,3 Rabah BEN YAOU, PhD 3,4 Emmanuelle SALORT-CAMPANA, MD, PhD,1,5 Jean POUGET, MD, PhD,1,5 Shahram ATTARIAN, MD, PhD,1,5 Bruno EYMARD, MD, PhD,3 Jean-François DELEUZE, PhD 6 Anne BOLAND, PhD 6

Anthony BEHIN, MD, PhD, 3 Tanya STOJKOVIC, MD, PhD,3 Gisele BONNE, PhD,4 Nicolas LEVY, MD, PhD,1,2 Marc BARTOLI*, PhD1,2 Martin KRAHN*, MD, PhD,1,2

1 Aix Marseille Univ, Inserm, GMGF, Marseille, France
2 APHM, Hôpital Timone Enfants, Département de Génétique Médicale, Marseille, France
4 Sorbonne Universités, UPMC Univ Paris 06, Inserm UMRS974, CNRS FRE3617, Center for Research in Myology, Institut de Myologie, G.H. Pitié Salpêtrière, Paris, France
5 APHM, Hôpital La Timone, Centre de référence des maladies neuromusculaires et de la SLA, Marseille, France
6 Centre National de Génotypage, Institut de Génomique, CEA, Evry, France.

* Authors with equal contribution.

Corresponding Authors:
Mathieu Cerino: mathieu.cerino@ap-hm.fr and Martin Krahn: martin.krahn@univ-amu.fr

INSERM AMU UMR_S910, Faculté de Médecine de Marseille, 4e étage Aile Verte,
27 boulevard Jean Moulin, 13385 Marseille Cedex 05, FRANCE
TEL: 0033-(0)4 91 32 49 40
FAX: 0033-(0)4 91 80 43 19

Abstract word count: 150 ; Manuscript word count: 1335

We confirm that we have read the Journal’s position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

The authors have no conflict of interest to report and no relevant financial relationships to disclose.

This study was supported by the France Génomique infrastructure (grant no. ANR-10-INBS-09) managed by the National Research Agency (ANR) part of the Investment for the Future program, Fondation Maladies Rares, FHU-MARCHÉ, the APHM, Inserm, the “Bureau des PU-PH de l’Assistance Publique – Hôpitaux de Marseille (AP-HM)”, and by a Grant FP7/2007–2013 from the European Community Seventh Framework Programme (Grant Agreement No. 2012–305121) “Integrated European omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases” (NEUROMICS).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an ‘Accepted Article’, doi: 10.1002/mus.25638

This article is protected by copyright. All rights reserved.
Genetic characterization of a French cohort of GNE-mutation negative inclusion body myopathy patients using exome sequencing

(Abstract)

INTRODUCTION: Hereditary inclusion body myopathy (hIBM) refers to a group of clinically and genetically heterogeneous diseases. The overlapping histochemical features of hIBM with other genetic disorders lead to low diagnostic rates with targeted single-gene sequencing. This is true for the most prevalent form of hIBM, GNEpathy. Thus, we used whole exome sequencing (WES) to evaluate whether a cohort of clinically suspected GNEpathy patients undiagnosed by targeted GNE analysis could be genetically characterized.

METHODS: 20 patients with hIBM but undiagnosed by targeted GNE sequencing were analyzed using WES before data filtering on 306 genes associated with neuromuscular disorders.

RESULTS: 7 patients out of 20 were found to have disease-causing mutations in genes associated with hIBM, or genes allowing for hIBM in the differential diagnosis, or associated with unexpected diagnosis.

DISCUSSION: NGS is an efficient strategy in the context of hIBM, resulting in a molecular diagnosis for 35% of the patients initially undiagnosed by targeted GNE analysis.

Keywords: exome, hIBM, GNE, NGS, diagnosis, myopathy
INTRODUCTION

Hereditary inclusion body myopathies (hIBM) represent a heterogeneous group of muscular disorders defined by the relatively nonspecific criterion of rimmed vacuoles on muscle biopsy\(^1\).

GNEpathy\(^2\), caused by mutations in \(GNE\) (UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase, MIM*603824)\(^3\) is the most common form of hIBM, with many clinical features overlapping with other forms of hIBM, implicating other genes or forms with yet unknown underlying genetic defects. Targeted analysis of \(GNE\) in a large recently-described French cohort with suspected GNEpathy provides only a 20% diagnostic yield (32 of 164 patients)\(^4\).

In the present study, we evaluated the extent to which a cohort of clinically suspected GNEpathy patients undiagnosed by \(GNE\) targeted analysis may be genetically characterized, by implicating other genes previously known to cause neuromuscular disorders using whole exome sequencing (WES) associated with data filtering for 306 genes of interest.

METHODS

We selected 20 unrelated index cases (IC) with clinically suspected GNEpathy associated with rimmed vacuoles on muscle biopsy samples, but for which no \(GNE\) disease-causing mutation had been identified by direct targeted sequencing. Samples had been prepared and stored by the Center of Biological Resources, Department of Medical Genetics, La Timone Hospital, Marseille, and were used following the ethical recommendations of our institution and according to the Declaration of Helsinki. All included patients gave their written consent prior to the genetic study, in accordance with French law.

WES was performed using the SureSelect Human All Exon Kit version 5 (Agilent Technologies, Santa Clara, California) and the HiSeq 2000 (Illumina, San Diego, California).
Sequencing data were processed on the Illumina pipeline (CASAVA1.8.2) before using GATK variant calling and ANNOVAR annotation using the GRCh37/hg19 Human genome version, coverage statistics were computed using VarAFT (Variant Analysis and Filtration Tool; http://varaft.eu, 2016), which uses BedTools. VarAFT was also used to sort and filter the obtained variants.

Our initial analysis strategy focused on 306 genes previously reported to cause neuromuscular disorders, and selected from the Gene Table of Neuromuscular Disorders (including groups 1 to 5 and the main differential diagnosis genes) as previously described. A mean overall sequencing depth of 106X and a mean coverage of the coding exons of 95% (at 20X depth) and 91% (at 30X depth) was obtained for these 306 genes. Predicted pathogenicity of identified variants was determined using UMD-predictor, SIFT (Sort Intolerant From Tolerant human Protein), PolyPhen-2 (Polymorphism Phenotyping v2) and HSF (Human Splicing Finder) softwares.

Regarding HSF in silico results, we defined four types of predicted splicing effects: 1) Probably damaging: associated with predicted strong splicing effect due to broken donor site (DS) or acceptor site (AS) and/or new DS/AS creation and/or strong possibility of broken Exonic Splicing Enhancer (ESE) site; 2) Possibly damaging: associated with predicted medium splicing effect relating to newly created DS/AS and/or medium possibility of broken ESE site; 3) Uncertain: associated with predicted mild splicing effect due to newly created DS/AS and/or low possibility of broken ESE site; and 4) Not affected: predicted weak or no splicing effect.
The overall pathogenicity score for each variant was determined according to the American College of Medical Genetics (ACMG) guidelines\textsuperscript{15}. We established four groups of patients based on the degree of certainty of molecular diagnosis using the ACMG guidelines. The group with “definite diagnosis” consisted of the following patients: 1) Those carrying a homozygous variant classified as “pathogenic” using ACMG guidelines in a gene known to cause an autosomal recessive form of disease; 2) Compound heterozygotes carrying two variants classified as pathogenic; 3) Patients carrying one variant classified as pathogenic in a gene known to cause an autosomal dominant form of disease. The group with “probable diagnosis” was composed of patients carrying variants that were classified as “likely pathogenic” by ACMG guidelines. Patients carrying variants found to be pathogenic by certain prediction tools, but classified as “variants of uncertain significance” by ACMG guidelines were placed in the group with “possible diagnosis”. For those patients in the “no established diagnosis” group, no variant compatible with the patient’s phenotype was found.

All disease-causing variants identified by WES were confirmed using direct targeted sequencing (Genetic analyzer 3500XL; Thermo Fisher Scientific, Waltham, Massachusetts) and the following gene sequence references: \textit{ACTA1} (NM_001100), \textit{CAPN3} (NM_000070), \textit{DES} (NM_001927), \textit{FLNC} (NM_001458), \textit{GYG1} (NM_004130), \textit{MYH2} (NM_017534), \textit{TARDBP} (NM_007375), \textit{TTN} (NM_001267550) and \textit{VCP} (NM_007126).

RESULTS

All phenotypic and mutational data are detailed in Table 1. A definite diagnosis was obtained for seven index cases (ICs). Patient P1 harbored a previously reported mutation in \textit{TTN} (Titin, MIM\textsuperscript{*}188840) associated with hereditary myopathy with early respiratory failure (HMERF)\textsuperscript{16}. The homozygous status of this mutation is consistent with the parental
consanguinity. For Patients P2 and P5, the same heterozygous mutation in \textit{VCP} (Valosin-Containing Protein, MIM *601023), previously described in the literature\textsuperscript{17}, was discovered and associated with similar onset and clinical features (distal myopathy of upper and lower limbs). Compound heterozygous known mutations in \textit{TTN}\textsuperscript{18,19} were found in patient P3, whereas patient P4 harbored a previously described heterozygous variant in \textit{DES} (Desmin, MIM*125660)\textsuperscript{20} leading to cardiomyopathy and myofibrillar abnormalities on the muscle biopsy, features that were retrieved in patient P4. For patient P6, a known \textit{FLNC} (Filamin C, MIM *102565) mutation was found\textsuperscript{21}. Surprisingly, we identified compound heterozygous mutations for the \textit{GYG1} (Glycogenin 1, MIM*603942) gene in patient P7, associated with polyglucosan body myopathy type 2. In this patient, we found a previously described \textit{GYG1} variant with a proven deleterious effect on splicing\textsuperscript{22}, associated with a novel \textit{GYG1} mutation leading to a frameshift of the reading frame and the introduction of a premature translation termination codon. Further investigations allowed additional clinical and histo-immunological features thus suggesting a polyglucosan body myopathy (data not shown). A probable diagnosis was obtained for patients P8 and P9, with novel compound heterozygous \textit{TTN} mutations and a heterozygous \textit{TARDBP} (Tar DNA-Binding Protein, MIM *605078) variant respectively while two novel heterozygous variants in the \textit{FLNC} and the \textit{ACTA1} (Actin, Alpha, skeletal muscle 1, MIM *102610) genes fulfilled the possible diagnosis overall pathogenicity score in patients P10 and P11 respectively.

Finally, 9 ICs remained without a molecular diagnosis following mutational analysis of the 306 genes of interest.

Considering only the first (definite) group of patients, the yield of diagnosed patients was 35% in this cohort (7/20).
DISCUSSION

Next-Generation Sequencing (NGS) is already used by many genetics laboratories and is being used with increasing frequency as the standard initial analysis for myopathies and other heterogeneous genetic disorders. The molecular diagnosis yield of 35% obtained in this study is consistent with other reports showing a range of 25 to 50 percent for rare genetic disorders diagnosis by WES\textsuperscript{23,24}.

Our study illustrates that a NGS approach is more efficient than the gene-by-gene strategy for several reasons. First, it allowed us to explore genes responsible for disorders within the differential diagnosis of hIBM, including \textit{VCP} and \textit{DES}. Second, our strategy permitted sequencing of large-sized genes, such as \textit{TTN} and \textit{FLNC}, which is not routinely performed, leading to the identification of variants in five index cases. Third, this approach allowed us to modify the incorrect diagnosis of hIBM in one patient, initially based on the presence of rimmed vacuoles on muscle biopsy, to a different muscle disorder caused by variants in the gene \textit{GYG1}. Thus, NGS has the potential to alter a misdiagnosis due to a misleading muscle biopsy. Although using WES to explore a subset of genes might not provide as much target sequence coverage as a sequencing strategy specifically designed for these genes\textsuperscript{10}, there are several advantages of using this approach. The sequencing results for samples where no pathogenic variants were identified in the initially explored genes can be reanalyzed to explore additional genes or all genes in the whole exome. In this way, further analyses are ongoing for the cases among our cohort that remain without genetic characterization. Another advantage of WES over targeted exome sequencing is its versatility and ability to be applied to many different diseases, as different sets of genes can be assessed without the need to develop and test a specific sequencing strategy\textsuperscript{25,26}. 
In conclusion, the exome-based sequencing strategy described here is an efficient way to diagnose such genetically heterogeneous disorders as hIBMs.
ABBREVIATIONS

ACMG: American College of Medical Genetics
AS: Acceptor Site
DS: Donor Site
ESE: Exonic Splicing Enhancer
hIBM: hereditary Inclusion Body Myopathies
HSF: Human Splicing Finder
IC: Index Case
NGS: Next-Generation Sequencing
PolyPhen-2: Polymorphism Phenotyping v2
SIFT: Sort Intolerant From Tolerant (amino acid substitutions)
VarAFT: Variant Analysis and Filtration Tool
WES: Whole Exome Sequencing

The authors sincerely thank Christel Castro, Jean-Pierre Desvignes, David Salgado, Christophe Béroud, Eric Salvo, Raafaelle Bernard, Jocelyn Laporte, Johann Bohm and Mark Lathrop for their contributions to this work. We also thank the patients and their referring physicians for their participation.
REFERENCES


### Table 1: Pathogenicity assessment for the identified variants in patients with definite, probable and possible diagnoses

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Phenotype / Genetic inheritance</th>
<th>Muscle biopsy</th>
<th>Genes/Variants</th>
<th>Status</th>
<th>UMD-predictor&lt;sup&gt;1&lt;/sup&gt;</th>
<th>SIFT&lt;sup&gt;13&lt;/sup&gt;</th>
<th>PolyPhen-2&lt;sup&gt;14&lt;/sup&gt;</th>
<th>ACMG Guidelines&lt;sup&gt;15&lt;/sup&gt;</th>
<th>Splicing prediction (HSF&lt;sup&gt;16&lt;/sup&gt;)</th>
<th>Pathogenic variants described in literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>F</td>
<td>DM of lower limbs at onset (40yo) evolving towards LMREF / AR</td>
<td>Rimmed vacuoles Cytoplasmic inclusions Disruption of the intermyofibrillar network</td>
<td>TTN: c.95195C&gt;T (p.Pro31732Leu)</td>
<td>HOZ</td>
<td>Pathogenic</td>
<td>Damaging</td>
<td>Probably damaging</td>
<td>Pathogenic</td>
<td>NP</td>
<td>YES&lt;sup&gt;16&lt;/sup&gt;</td>
</tr>
<tr>
<td>P2</td>
<td>F</td>
<td>DM of upper and lower limbs (tibialis anterior muscle) No axial muscle weakness / AD</td>
<td>Rimmed vacuoles Disruption of the intermyofibrillar network</td>
<td>VCP: c.1403C&gt;T (p.Pro137Leu)</td>
<td>HTZ</td>
<td>Pathogenic</td>
<td>Damaging</td>
<td>Probably damaging</td>
<td>Pathogenic</td>
<td>NP</td>
<td>YES&lt;sup&gt;17&lt;/sup&gt;</td>
</tr>
<tr>
<td>P3</td>
<td>F</td>
<td>Early onset (childhood) DM of lower limbs (tibial muscular dystrophy) slowly evolving / AR</td>
<td>Rimmed vacuoles Dystrophic muscle biopsy</td>
<td>TTN: c.102271C&gt;T (p.Arg34091Trp)</td>
<td>HTZ</td>
<td>Pathogenic</td>
<td>Damaging</td>
<td>Probably damaging</td>
<td>Pathogenic</td>
<td>NP</td>
<td>YES&lt;sup&gt;18&lt;/sup&gt;</td>
</tr>
<tr>
<td>P4</td>
<td>M</td>
<td>Late onset (45yo) DM of upper and lower limbs with cardiac involvement / AD</td>
<td>Rimmed vacuoles</td>
<td>DES: c.1369C&gt;T (p.Arg454Trp)</td>
<td>HTZ</td>
<td>Pathogenic</td>
<td>Damaging</td>
<td>Probably damaging</td>
<td>Pathogenic</td>
<td>NO</td>
<td>YES&lt;sup&gt;19&lt;/sup&gt;</td>
</tr>
<tr>
<td>P5</td>
<td>M</td>
<td>DM of upper and lower limbs; Onset at 30yo / AD</td>
<td>Rare rimmed vacuoles (-5)</td>
<td>VCP: c.410C&gt;T (p.Pro137Leu)</td>
<td>HTZ</td>
<td>Pathogenic</td>
<td>Damaging</td>
<td>Probably damaging</td>
<td>Pathogenic</td>
<td>NP</td>
<td>YES&lt;sup&gt;20&lt;/sup&gt;</td>
</tr>
<tr>
<td>P6</td>
<td>F</td>
<td>Late onset (45yo) DM of lower limbs and pelvic girdle myopathy / AD</td>
<td>Rimmed vacuoles</td>
<td>FLNC: c.8130G&gt;A (p.Arg2710*)</td>
<td>HTZ</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>Pathogenic</td>
<td>NP</td>
<td>YES&lt;sup&gt;21&lt;/sup&gt;</td>
</tr>
<tr>
<td>P7</td>
<td>F</td>
<td>Late onset (45yo) DM of upper and lower limbs with slow evolution / AR</td>
<td>Rimmed vacuoles (on the initial biopsy) recharacterized as polyglucosan bodies (on the second biopsy)</td>
<td>GYG1: c.1433G&gt;C (p.Arg478Arg) GYG1: c.996_1005del10 (p.Tyr332fs*1)</td>
<td>Comp. HTZ</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>Pathogenic</td>
<td>NO</td>
<td>YES&lt;sup&gt;22&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

### Patients with probable diagnosis

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Phenotype / Genetic inheritance</th>
<th>Muscle biopsy</th>
<th>Genes/Variants</th>
<th>Status</th>
<th>UMD-predictor&lt;sup&gt;1&lt;/sup&gt;</th>
<th>SIFT&lt;sup&gt;13&lt;/sup&gt;</th>
<th>PolyPhen-2&lt;sup&gt;14&lt;/sup&gt;</th>
<th>ACMG Guidelines&lt;sup&gt;15&lt;/sup&gt;</th>
<th>Splicing prediction (HSF&lt;sup&gt;16&lt;/sup&gt;)</th>
<th>Pathogenic variants described in literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>P8</td>
<td>M</td>
<td>Early onset (14yo) DM of lower limbs (tibial muscular dystrophy) evolving towards hamstring muscle with quadriceps sparing / AR</td>
<td>Rimmed vacuoles</td>
<td>TTN: c.15346C&gt;T (p.Arg5116*)</td>
<td>HTZ</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>Pathogenic</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>P9</td>
<td>M</td>
<td>Late onset (50yo) DM of upper and lower limbs / AD</td>
<td>Rare rimmed vacuoles (-5)</td>
<td>FLNC: c.6526C&gt;T (p.Arg2176Cys)</td>
<td>HTZ</td>
<td>Pathogenic</td>
<td>Tolerated</td>
<td>Benign</td>
<td>Likely pathogenic</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

### Patients with possible diagnosis

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Phenotype / Genetic inheritance</th>
<th>Muscle biopsy</th>
<th>Genes/Variants</th>
<th>Status</th>
<th>UMD-predictor&lt;sup&gt;1&lt;/sup&gt;</th>
<th>SIFT&lt;sup&gt;13&lt;/sup&gt;</th>
<th>PolyPhen-2&lt;sup&gt;14&lt;/sup&gt;</th>
<th>ACMG Guidelines&lt;sup&gt;15&lt;/sup&gt;</th>
<th>Splicing prediction (HSF&lt;sup&gt;16&lt;/sup&gt;)</th>
<th>Pathogenic variants described in literature</th>
</tr>
</thead>
</table>

In bold: pathogenicity prediction strong and moderate; NP: not performed (UMD-predictor, SIFT and PolyPhen-2 algorithms do not provide a pathogenicity score for variants creating a stop or a frameshift); UMD-predictor<sup>11</sup>; Universal Mutation Database predictor; SIFT<sup>12</sup>; Sort Intolerant From Tolerant; PolyPhen-2<sup>13</sup>; Polymorphism Phenotyping v2; HSF<sup>16</sup>; Human Splicing Finder; ACMG<sup>15</sup>; American College of Medical Genetics; AD: Autosomal Dominant; AR: Autosomal Recessive; HMREF: Hereditary Myopathy with early Respiratory Failure; DM: Distal Myopathy; PM: Proximal Myopathy; yr: years old. HOZ: homozygous; HTZ: heterozygous; Comp. HTZ: compound heterozygous (with confirmed segregation analysis).

Frequency in 1000G, ESP and ExAC databases of all the variants described in Table 1 is lower than 0.2%.


† Variant affecting the same nucleic and amino acid positions as another variant, c.1127G>A (p.Gly376Asp), previously described in the literature<sup>27,28,29</sup> in two different familial amyotrophic lateral sclerosis cases with a similar phenotype presentation as patient P9 (upper and lower limb weakness with no cognitive impairment).