Copy number variation analysis increases the diagnostic yield in muscle diseases

Salla Välipakka, MSc
Marco Savarese, PhD
Mridul Johari, MSc
Lydia Sagath, MSc
Meharji Arumilli, MSc
Kirsu Kiiski, PhD
Amets Sáenz, PhD
Adolfo Lopez de Munain, MD, PhD
Ana-Maria Cobo, PhD
Katarina Pelin, PhD
Bjarne Udd, MD, PhD
Peter Hackman, PhD

Correspondence to
S. Välipakka:
salla.valipakka@helsinki.fi

ABSTRACT

Objective: Copy number variants (CNVs) were analyzed from next-generation sequencing data, with the aim of improving diagnostic yield in skeletal muscle disorder cases.

Methods: Four publicly available bioinformatic analytic tools were used to analyze CNVs from sequencing data from patients with muscle diseases. The patients were previously analyzed with a targeted gene panel for single nucleotide variants and small insertions and deletions, without achieving final diagnosis. Variants detected by multiple CNV analysis tools were verified with either array comparative genomic hybridization or PCR. The clinical significance of the verified CNVs was interpreted, considering previously identified variants, segregation studies, and clinical information of the patient cases.

Results: Combining analysis of all different mutation types enabled integration of results and identified the final cause of the disease in 9 myopathy cases. Complex effects like compound heterozygosity of different mutation types and compound disease arising from variants of different genes were unraveled. We identified the first large intragenic deletion of the titin (TTN) gene implicated in the pathogenesis of a severe form of myopathy. Our work also revealed a “double-trouble” effect in a patient carrying a single heterozygous insertion/deletion mutation in the TTN gene and a Becker muscular dystrophy causing deletion in the dystrophin gene.

Conclusions: Causative CNVs were identified proving that analysis of CNVs is essential for increasing the diagnostic yield in muscle diseases. Complex severe muscular dystrophy phenotypes can be the result of different mutation types but also of the compound effect of 2 different genetic diseases. Neurol Genet 2017;3:e204; doi: 10.1212/NXG.0000000000000204

Glossary

aCGH = array comparative genomic hybridization; BMD = Becker muscular dystrophy; CK = creatine kinase; CNV = copy number variant; indels = insertions and deletions; LGMD = limb-girdle muscular dystrophy; NGS = next-generation sequencing; SNV = single nucleotide variant.

Next-generation sequencing (NGS) methods have become the most common method for the genetic diagnosis of genetically heterogeneous disorders.1,2 We have previously developed a targeted NGS gene panel, MyoCap.1 An updated version used here includes probes for the exons of nearly 300 myopathy genes and candidate genes. Similar platforms are currently in use in many laboratories.2 The reported diagnostic success rates are significantly higher than those obtained by traditional gene-by-gene sequencing.2 However, over 50% patients remain undiagnosed when only concentrating on single nucleotide variants (SNVs) and small insertions and deletions (indels).2

Copy number variants (CNVs) are defined as genomic deletions or duplications greater than 1 kb in size.3 CNVs cause microdeletion and microduplication syndromes, and they have also

From the Folkhålsan Institute of Genetics (S.V., M.S., M.J., L.S., M.A., K.K., K.P., B.U., P.H.), Medicum, Faculty of Biological and Environmental Sciences (K.P.), University of Helsinki, Finland; Neuromuscular Research Center (B.U.), Vaasa Central Hospital, Finland; Biodonostia Health Research Institute (A.S., A.L.D.M), Neurosciences Area, CIBERNED, University of the Basque Country, San Sebastián, Spain; and Centre de Référence Maladies Neuromusculaires (GNMH) (A.-M.C.), Hôpital Marin APHP, Hendaye, France.

Funding information and disclosures are provided at the end of the article. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Neurology.org/ng Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.
been associated with several complex diseases.³⁴ Generally, studies aiming for the identification of causative disease variants in skeletal muscle disorders have not systematically used CNV screening. Multiplex ligation-dependent probe amplification has a lower throughput when the amount of investigated genes increases.⁵ Array comparative genomic hybridization (aCGH) has long been considered the only reliable and robust platform for CNV discovery.⁴ However, in NGS studies, the diagnostic evaluation may end with a discovery of a single pathogenic or likely pathogenic mutation before the utilization of complementary methods, which may lead to an underestimation of CNV contribution to diseases. Recently, several CNV analysis tools for NGS data have been developed and are in use for routine diagnosis.³⁴ Here, we describe the detection of CNVs from NGS data with a combination of already available bioinformatic tools.

METHODS  Standard protocol approvals, registrations, and patient consents. DNA samples of muscle disease patients and healthy family members were obtained from clinicians in different countries. The study was approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa. The samples were obtained according to the Helsinki declaration. Written informed consent was obtained from all patients.

RESULTS Table 1 shows the clinical features of patients in this study, genetic findings, amount of programs/verified with genetic findings, and detected by n programs/verified with CNV assessment from NGS data. CNVs were analyzed in smaller batches from NGS data alignment files (.bam) obtained by analyzing DNA from 791 myopathy patients with MyoCap.¹ We used 4 CNV analysis programs: Copy Number Inference From Exome Reads (CoNIFER) v0.2.2,⁶ eXome-Hidden Markov Model (XHMM) v1.1,⁷ ExomeDepth v1.1.10,⁸ and CODEX v1.4.0,⁹ with the recommended default settings of each program. A minimum of 1-bp overlap was used to determine whether calls intersecting between different programs originated from the same CNV. In this article, we prioritize the specific cases with a very high clinical interest, focusing on 7 variants detected by multiple programs and verified by using independent tools.

Table 1  Patient and variant features

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Onset</th>
<th>First symptom</th>
<th>Muscle weakness</th>
<th>CK</th>
<th>Previous genetic tests</th>
<th>Genetic findings</th>
<th>Detected by n programs/verified with</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>M</td>
<td>Early adult</td>
<td>Weakness</td>
<td>Proximal LL</td>
<td>9x</td>
<td>CAPN3; FKRIP; LIMMA; RIM32⁴</td>
<td>TTN: c.107780-107790 delinsTGAAAAGAAAAA; DMD: c.86438+1_6439-1</td>
<td>3/PCR</td>
</tr>
<tr>
<td>IIa</td>
<td>M</td>
<td>Infancy</td>
<td>Weakness, exercise intolerance</td>
<td>Proximal and distal LL</td>
<td>11x</td>
<td>CAPN3; DYSF⁴</td>
<td>TTN: c.107889delAp, (Lys35963Asnfs¹) c. (7855-1,7856-1), (9703-1,9704-1)del</td>
<td>4/aCGH</td>
</tr>
<tr>
<td>IIIa</td>
<td>F</td>
<td>Infancy</td>
<td>Weakness, walking difficulties</td>
<td>Proximal LL, Gowers +</td>
<td>&gt;10x</td>
<td>CAPN3; ANO5; WES</td>
<td>SGCD: c.?._1 (505,510-1) del hom</td>
<td>4/PCR</td>
</tr>
<tr>
<td>IV</td>
<td>M</td>
<td>Infancy</td>
<td>Walking difficulties</td>
<td>Proximal &gt; distal; Gowers +</td>
<td>20x</td>
<td>DMD⁴</td>
<td>CAPN3: c.309 + 3_1310-1, (1,7) del hom</td>
<td>2/PCR</td>
</tr>
<tr>
<td>V</td>
<td>M</td>
<td>Early adult</td>
<td>Walking difficulties</td>
<td>Proximal LL</td>
<td>5-6x</td>
<td>None</td>
<td>DMD: c.86438+1_6439-1, (86437+1,86438-1)del</td>
<td>3/PCR</td>
</tr>
<tr>
<td>VI</td>
<td>M</td>
<td>Early adult</td>
<td>Walking difficulties</td>
<td>Muscle atrophy; calf hypertrophy</td>
<td>10x</td>
<td>None</td>
<td>DMD: c.86438+1_6439-1, (7099+1,7099-1)del</td>
<td>2/PCR</td>
</tr>
<tr>
<td>VII</td>
<td>M</td>
<td>Infancy</td>
<td>Walking difficulties</td>
<td>Proximal LL, Gowers +</td>
<td>&gt;10x</td>
<td>None</td>
<td>DMD: c.822+1_5923-1, (6290+1,6291-1)del</td>
<td>2/PCR</td>
</tr>
</tbody>
</table>

Abbreviations: aCGH = array comparative genomic hybridization; CK = creatine kinase; hom = homozygous; LL = lower limbs; NA = not available. Genetic findings are described using the reference transcripts: NM_001267550.1 for TTN, NM_004006.1 for DMD, NM_000231.2 for SGCD, and NM_000070.2 for CAPN3.

¹Sanger sequencing.
²Multiplex ligation-dependent probe amplification.
CK level. We excluded the presence of other deleterious variants in \textit{TTN}. Surprisingly, the patient was found to have a previously reported dystrophin \textit{(DMD)} deletion (exons 45–55) known to cause Becker muscular dystrophy (BMD).\textsuperscript{11}

Patient IIa has a previously reported Iberian frameshift mutation, p.(Lys35963Asnfs*), in the last exon of \textit{TTN} usually determining 1 component of the recessive distal titinopathy phenotype.\textsuperscript{10} Compared with the other carriers of this variant, this patient has a more severe disease progression with proximal weakness, loss of ambulation before the age of 40 years, and marked hyperCKemia. The patient had no further causative SNVs or indels in the \textit{TTN} gene. We identified a large deletion in \textit{TTN} (exons 34–41; figure 1A) in trans with the Iberian frameshift variant in the proband IIa as well as in patient IIb, a similarly affected brother. Their healthy relatives are heterozygous for only one of the aforementioned \textit{TTN} mutations demonstrating the recessive effect of the detected deletion. The severe distal and proximal titinopathies were thus caused by the compound heterozygosity of the Iberian frameshift and the deletion (figure 2).

**Figure 1** Visualizations of copy number variants of patients IIa and IIIa

CoNIFER visualizations of a heterozygous deletion in the \textit{TTN} in patient IIa (A) and a homozygous deletion in the SGCD in patient IIIa (B) accompanied with corresponding regions visualized with Integrative Genomics Viewer (IGV). In CoNIFER visualizations, the red line corresponds to the read depths for sample with deletion and black lines correspond to control samples. In the IGV visualization, the first row shows the sample with deletion and the second row shows a control sample without deletion.
Segregation of the TTN Iberian tibial muscular dystrophy variant, p.(Lys35963Asnfs*41), and deletion in the TTN (exons 34–43) in the family of patients IIa and IIb. The affected brothers IIa and IIb (black squares) are compound heterozygous for the variants. Their healthy parents and siblings are heterozygous for only one of the variants. Genetic findings are described using the reference transcript NM_001267550.1 for TTN.

Patients IIIa and IIIb, with a severe limb-girdle muscular dystrophy (LGMD) phenotype, are the daughters of first-cousin parents. Sanger sequencing of candidate genes (CAPN3 and ANOS1), MyoCap, and whole-exome sequencing had been performed without identifying the causative variant. All the CNV detection programs identified a homozygous deletion in the SGCD gene (exons 1–5) (figure 1B).

Patient IV is the child of consanguineous parents, and he has been suffering from lower limb muscular weakness in the lower limbs since the age of 7. Deletions and duplications in the DMD gene had been excluded. Immunochemistry showed normal staining for dystrophin as well as for the sarcoglycans. A homozygous deletion in CAPN3 (exons 2–8) was detected.

In 2 males with an LGMD-like phenotype (patients V and VI), we found previously reported DMD deletions explaining the observed proximal muscular weakness. Both patients have in-frame deletions (exons 45–55 and exons 45–48) causing BMD.11

Patient VII with a Duchenne phenotype was also included in our screening. As expected, a previously reported out-of-frame DMD deletion (exons 42–43) was identified.11

**DISCUSSION** Genomes are usually analyzed for SNVs and indels, but studies of CNVs are often underrepresented. CNVs in muscle diseases have previously been studied with complementary methods like targeted aCGH, which still remains the gold standard technique for CNV detection.4,5 Here, we show that combining analysis of all different mutation types enables integration of results and identifies the final cause of the disease in several cases. Complex effects like compound heterozygosity in patients IIa and IIb and compound genetic disease arising from variants of different genes in patient I can be unraveled. It has been suggested that the number of patients with a combination of 2 or more genetic diseases is probably underestimated.12 The phenotypic complexity in these patients may erroneously be interpreted as a new genetic disease with unidentified genetic defect or as a phenotypic expansion for a single disease.12 The proximal phenotype seen in patient I is mainly due to BMD, as the typical anterior lower-leg muscle lesions of the FINmaj TTN mutation may develop only after age 60. However, the identification of multilocus genomic variants is crucial for a proper genetic counseling in the family.

The combination of 4 analysis tools aided us to identify already known and previously unknown CNVs. CNVs can explain some of the missing heritability and undiagnosed cases in skeletal muscle disorders. An NGS-based strategy for CNV detection will be of great value for increasing the diagnostic yield in patients affected by mendelian muscle diseases. Current technology does not adequately capture repeat expansion diseases such as myotonic dystrophy or diseases related to other repetitive elements. Long-read sequencing technologies may further help the identification and mapping of CNVs as well as of repeat expansions. However, the clinical interpretation of CNVs remains challenging, in particular for CNVs identified in genes without previously reported disease causing deletions or duplications. The inclusion of CNV data in public databases, e.g., ExAC, could help in pathogenicity assessment of CNVs.

**AUTHOR CONTRIBUTIONS**
Salla Välipakka: study concept and design, acquisition of data, analysis and interpretation of data, and drafting the manuscript for intellectual content. M. Savarese and Mridul Johari: study concept and design, acquisition of data, analysis and interpretation of data, and revising the manuscript for intellectual content. Lydia Sagath, Meharji Arumilli, and Kirsi Kiiski: acquisition of data, analysis and interpretation of data, and revising the manuscript for intellectual content. Amets Sáenz, Adolfo Lopez De Munain, and Ana-Maria Cobo: analysis and interpretation of data. Katarina Pelin, Bjarni Udd, and Peter Hackman: study concept and design and revising the manuscript for intellectual content.

**ACKNOWLEDGMENT** The authors thank Merja Soininen and Helena Luque for their technical help and Sini Penttilä, Tiina Suominen, and Sara Lehtinen for acquisition of samples. They also thank all the patients and family members as well as the clinicians who provided samples.

**STUDY FUNDING** This study was supported by the Folkhalsan Research Foundation, the Jane and Aatos Erkko Foundation, the Academy of Finland (no. 138491, B.U.), the Sigrid Jusélius Foundation, the Association Française contre les Myopathies, and the Orion Research Foundation sr.

**DISCLOSURE** S. Välipakka has received research support from the Folkhalsan Research Foundation. M. Savarese has received research support from the Association Française contre les Myopathies and Orion Research Foundation. M. Johari,
REFERENCES