



**HAL**  
open science

## Long-Reads Sequencing Strategy to Localize Variants in TTN Repeated Domains

Aurélien Perrin, Charles van Goethem, Corinne Thèze, Jacques Puechberty,  
Thomas Guignard, Bérénice Lecardonnell, Delphine Lacourt, Corinne Métay,  
Arnaud Isapof, Sandra Whalen, et al.

### ► To cite this version:

Aurélien Perrin, Charles van Goethem, Corinne Thèze, Jacques Puechberty, Thomas Guignard, et al.. Long-Reads Sequencing Strategy to Localize Variants in TTN Repeated Domains. *Journal of Molecular Diagnostics*, In press, 10.1016/j.jmoldx.2022.04.006 . hal-03672757

**HAL Id: hal-03672757**

**<https://hal.science/hal-03672757>**

Submitted on 19 May 2022

**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.

Copyright

# TECHNICAL ADVANCE

## Long-Reads Sequencing Strategy to Localize Variants in *TTN* Repeated Domains

Aurélien Perrin,<sup>\*†</sup> Charles Van Goethem,<sup>\*</sup> Corinne Thèze,<sup>\*</sup> Jacques Puechberty,<sup>‡</sup> Thomas Guignard,<sup>§</sup> Bérénice Lecardonnel,<sup>§</sup> Delphine Lacourt,<sup>\*</sup> Corinne Métay,<sup>¶</sup> Arnaud Isapof,<sup>||</sup> Sandra Whalen,<sup>\*\*</sup> Ana Ferreira,<sup>††‡‡</sup> Marie-Christine Arne-Bes,<sup>§§</sup> Susana Quijano-Roy,<sup>¶¶|||</sup> Juliette Nectoux,<sup>\*\*\*</sup> France Leturcq,<sup>†††</sup> Pascale Richard,<sup>¶</sup> Marion Larrieux,<sup>\*</sup> Anne Bergougou,† Franck Pellestor,<sup>§</sup> Michel Koenig,<sup>\*†</sup> and Mireille Cossée<sup>\*†</sup>

From the Molecular Diagnostic Laboratory,<sup>\*</sup> Montpellier University Hospital, Montpellier; the PhyMedExp,<sup>†</sup> University of Montpellier, INSERM, CNRS, Montpellier; the Department of Medical Genetics,<sup>‡</sup> Arnaud de Villeneuve Hospital, Montpellier; the Laboratoire de Génétique Chromosomique,<sup>§</sup> Plateforme ChromoStem, CHU de Montpellier, Université de Montpellier, Montpellier; the AP-HP,<sup>¶</sup> UF Molecular Cardiogenetics and Myogenetics, Sorbonne Université and Sorbonne Université UPMC Paris 06—Inserm UMR974, Research Center in Myology, Pitié-Salpêtrière Hospital, Paris; the Centre de Référence des Maladies Neuromusculaires Nord/Est/Ile de France,<sup>||</sup> Service de Neuropédiatrie, Hôpital Trousseau, Paris; the Genetics and Cytogenetics Department,<sup>\*\*</sup> Centre de Référence Déficiences Intellectuelles de Causes Rares, Pitié-Salpêtrière, Assistance Publique des Hôpitaux de Paris, Paris; the APHP,<sup>††</sup> Centre de Référence des Pathologies Neuromusculaires Nord-Est-Ile de France, Institut de Myologie, GHU Pitié-Salpêtrière, Paris; the Basic and Translational Myology Laboratory,<sup>‡‡</sup> Université de Paris BFA, UMR 8251, CNRS, Paris; the Department of Neurology,<sup>§§</sup> Purpan University Hospital, Toulouse; the APHP, GH Université Paris-Saclay,<sup>¶¶</sup> Neuromuscular Center, Child Neurology and ICU Department, Raymond Poincaré Hospital, Garches; the Université de Versailles,<sup>|||</sup> U1179 INSERM-UVSQ, Montigny; the Service de Génétique et Biologie Moléculaires,<sup>\*\*\*</sup> Hôpital Cochin, DMU BioPhyGen, Assistance Publique-Hôpitaux de Paris, AP-HP, Centre-Université de Paris, Paris; and the Department of Genetics and Molecular Biology,<sup>†††</sup> AP-HP, Cochin Hospital, Paris, France

Address correspondence to Mireille Cossée, Laboratoire de Génétique Moléculaire—CHU Montpellier/PhymedExp INSERM U1046—CNRS UMR9214—UM—IURC, Institut Universitaire de Recherche Clinique 641, Ave. du Doyen Gaston Giraud 34093, Montpellier Cedex 5, France. E-mail: [mireille.cossee@inserm.fr](mailto:mireille.cossee@inserm.fr).

Titin protein is responsible for muscle elasticity. The *TTN* gene, composed of 364 exons, is subjected to extensive alternative splicing and leads to different isoforms expressed in skeletal and cardiac muscle. Variants in *TTN* are responsible for myopathies with a wide phenotypic spectrum and autosomal dominant or recessive transmission. The I-band coding domain, highly subject to alternative splicing, contains a three-zone block of repeated sequences with 99% homology. Sequencing and localization of variants in these areas are complex when using short-reads sequencing, a second-generation sequencing technique. We have implemented a protocol based on the third-generation sequencing technology (long-reads sequencing). This new method allows us to localize variants in these repeated areas to improve the diagnosis of *TTN*-related myopathies and offer the analysis of relatives in postnatal or in prenatal screening.

Titinopathies are hereditary skeletal and cardiac myopathies due to alteration of titin, a giant elastic protein of the sarcomere encoded by the *TTN* gene. Titin, the largest known protein (3 to 3.8 MDa), is the third most abundant protein after myosin and actin in striated muscle.<sup>1</sup> Two titin filaments with opposite polarity span each half of muscle sarcomere from the Z-line to the M-line; as a result, it plays a crucial role in maintaining the integrity of the sarcomeric structural arrangement and in generating passive muscle stiffness.<sup>2</sup> Many structural domains compose titin structure, and some of them can interact with specific protein interaction partners.

The amino terminus of titin is embedded in the sarcomere Z-disk and participates in myofibril assembly, stabilization, and maintenance. The elastic I-band, containing repetitive Ig-type

Supported by AFM 21384 grant [The French Muscular Dystrophy Association (AFM-Téléthon)], the Délégation à la Recherche Clinique et à l'Innovation du Groupement de Coopération Sanitaire de la Mission d'Enseignement, de Recherche, de Référence et d'Innovation de Montpellier-Nîmes, and the research platform Chromostem.

C.T. and J.P. contributed equally to this work.

F.P. and M.K. contributed equally to this work.

Disclosures: None declared.

domains, behaves as a bidirectional spring, restoring sarcomeres to their resting length after muscle contraction. This domain is also subject to strong alternative splicing during development. The inextensible A-band binds myosin and myosin-binding protein and is thought to be critical for biomechanical sensing and signaling.<sup>3</sup>

The initial description of skeletal muscle disease due to *TTN* variants was tibial muscular dystrophy, an autosomal dominant late-onset tibial muscular dystrophy. Because of the widespread use of next-generation sequencing (NGS), *TTN* is emerging as a major causative gene in neuromuscular disorders, with >10 different disorders of dominant or recessive inheritance.<sup>4–6</sup> Phenotypic variability spreads from congenital myopathies to adult forms of skeletal defects associated or not with cardiomyopathies. Multiple domains constitute titin, which are encoded by the 364 exons that form the *TTN* gene (NM\_001267550.1, <https://www.ncbi.nlm.nih.gov/nucore/388998876>, last accessed April 20, 2022). Transcripts are subject to strong alternative splicing and more particularly the exons coding for I-band. This I-band also has a particular domain composed of three tandem repeats (exons 173 to 181, 182 to 190, and 191 to 199), sharing 99% DNA sequence homology among them.

NGS analyses are commonly performed using short-reads sequencers classically performed as a 2 \* 150 bp reading in a molecular diagnostic laboratory. This approach allows the detection of variants located in the repeated exons, but

without precise localization in one of the three repeats and with artifactual decrease of allelic balance.<sup>7</sup> The challenge of perfectly locating the variants identified in this area is multiple, ensuring the patient a precise diagnosis and allowing the analysis of relatives and prenatal diagnosis in a targeted way. We propose a reliable method targeting *TTN* repeated regions to confirm the localization of variants identified in high-throughput DNA sequencing. This method is based on a first step of amplification of an amplicon of the three repeated zones and a second step of long-reads sequencing using the MinION technology of Oxford Nanopore Technologies (Oxford, UK) to generate up to 9000-bp reads.

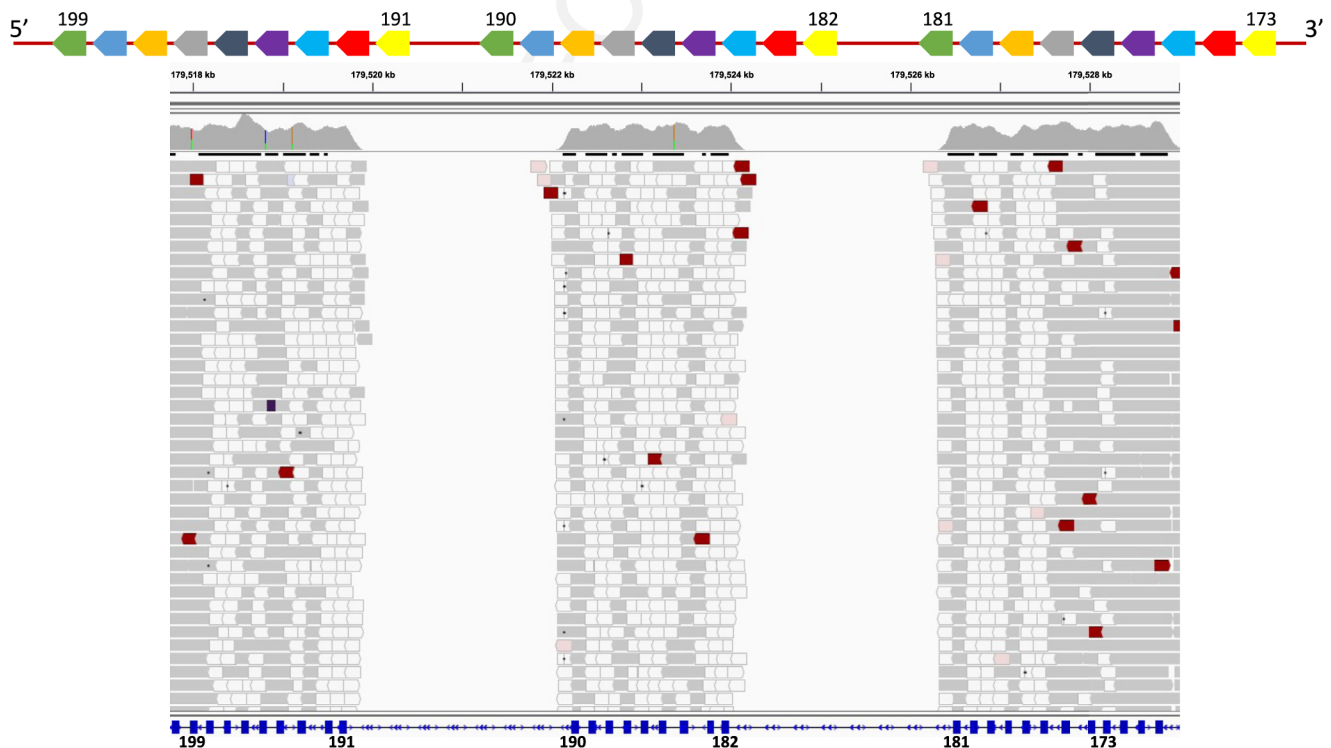
## Materials and Methods

### DNA Extraction

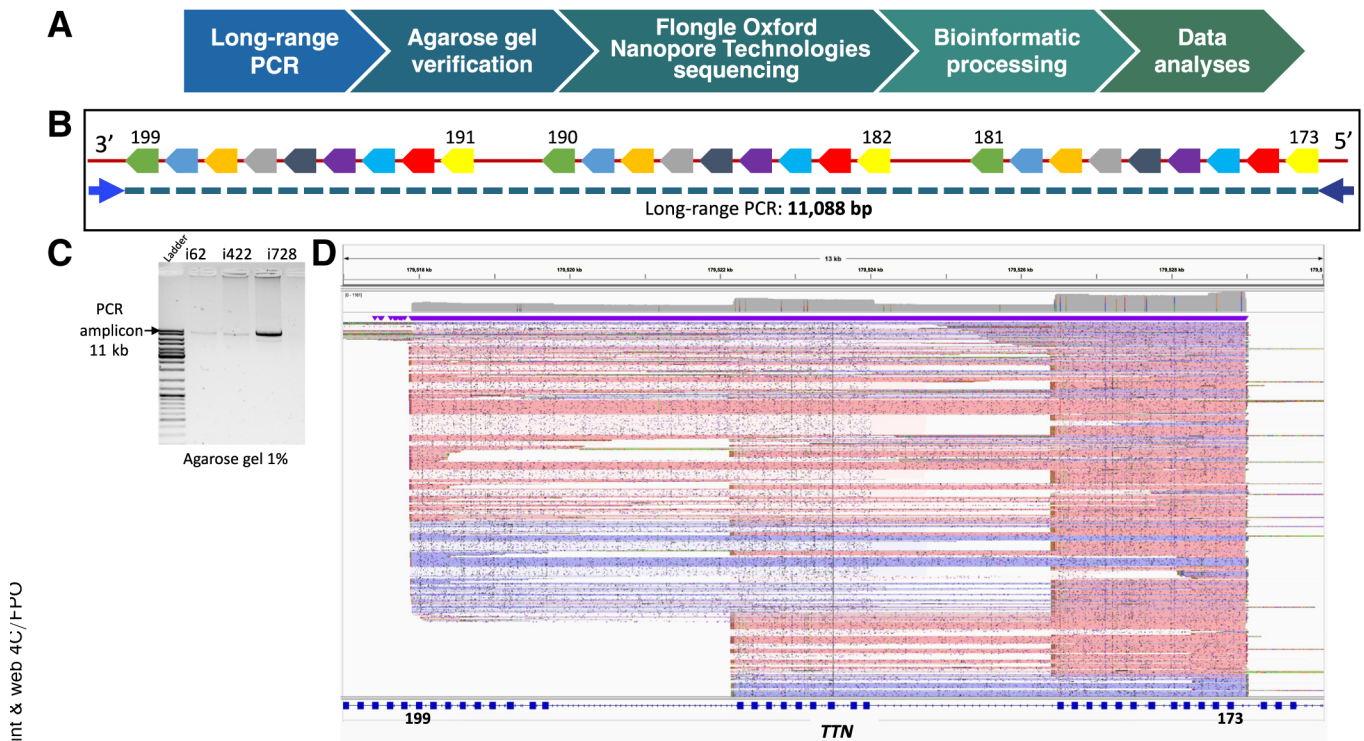
Genomic DNA was extracted from blood samples following the manufacturer's standard procedure of the FlexiGene DNA kit (Qiagen, France). DNA concentrations were measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE).

### Long-Range PCR Enrichment and Purification

We used Qiagen Long Range PCR kit to enrich DNA amplicon from intron 169 to intron 199 with PCR primers



**Figure 1** Integrative Genomics Viewer visualization of repeated regions. Each colored **arrow** represents an exon, and **arrows** of the same color are homologous exons. White reads indicate reads that have mapping quality of 0. Reads correctly mapped are represented in gray. Purple reads indicate insert size smaller than expected; dark red reads, insert size longer than expected; and pink reads, insert size longer than expected with mapping quality at 0. The ruler indicates the chromosomal position in kilobases (kb).



**Figure 2** Long-range PCR followed by long-reads sequencing. **A:** Workflow of sequencing strategy. **B:** Representation of the three repeated blocks and the primers represented by **light** and **dark blue arrows** used for long-range PCR. Each colored **arrow** represents an exon, and **arrows** of the same color are homologous exons. **C:** Verification of the presence of a specific long-range PCR product of approximately 11 kilobases (kb) in the three control DNAs (i62, i422, and i728) using electrophoretic agarose gel analysis. **D:** Integrative Genomics Viewer screenshot of sequencing raw data, showing three populations of reads.

(forward, 5'-GAAGTTCCTCCTCTAAAACCTGA-3'; and reverse, 5'-GAACAAAATGTCTTCAACTGCAAAA-3'). The conditions of PCR were 93°C, 3 minutes (1 cycle); 93°C, 15 seconds, 61°C, 30 seconds, and 68°C, 11 minutes (10 cycles); 93°C, 15 seconds, 62°C, 30 seconds, 68°C, 11 minutes plus 20 seconds/cycle (28 cycles); and 10°C, ∞. Long-range PCR amplification was assessed by agarose/ethidium bromure 1% electrophoresis. Long-range PCR products were purified with AMPure XP Beads (Agencourt, France).

### Long-Reads Sequencing with MinION Nanopore Technology and Bioinformatics

DNA-long-range PCR library preparation was performed using the SQK-LSK109 kit, according to the supplier's protocol, and sequenced on a Flongle flow cell in a MinION device and with MinKNOW software version 20.10.3 for 2 hours (Oxford Nanopore Technologies). Basecalling was performed with Guppy 4.2.2 using high-accuracy model (template\_r9.4.1\_450bps\_hac.jsn). Reads were aligned on hg19 genome assembly with minimap2 version 2.17.

## Results

As reported previously, short-reads sequencing is not reliable enough for determination of the localization of

variants in repeated exons of *TTN* gene.<sup>7</sup> Variants called in exon 191 are also called on exons 182 and 173, and conversely (Figure 1).

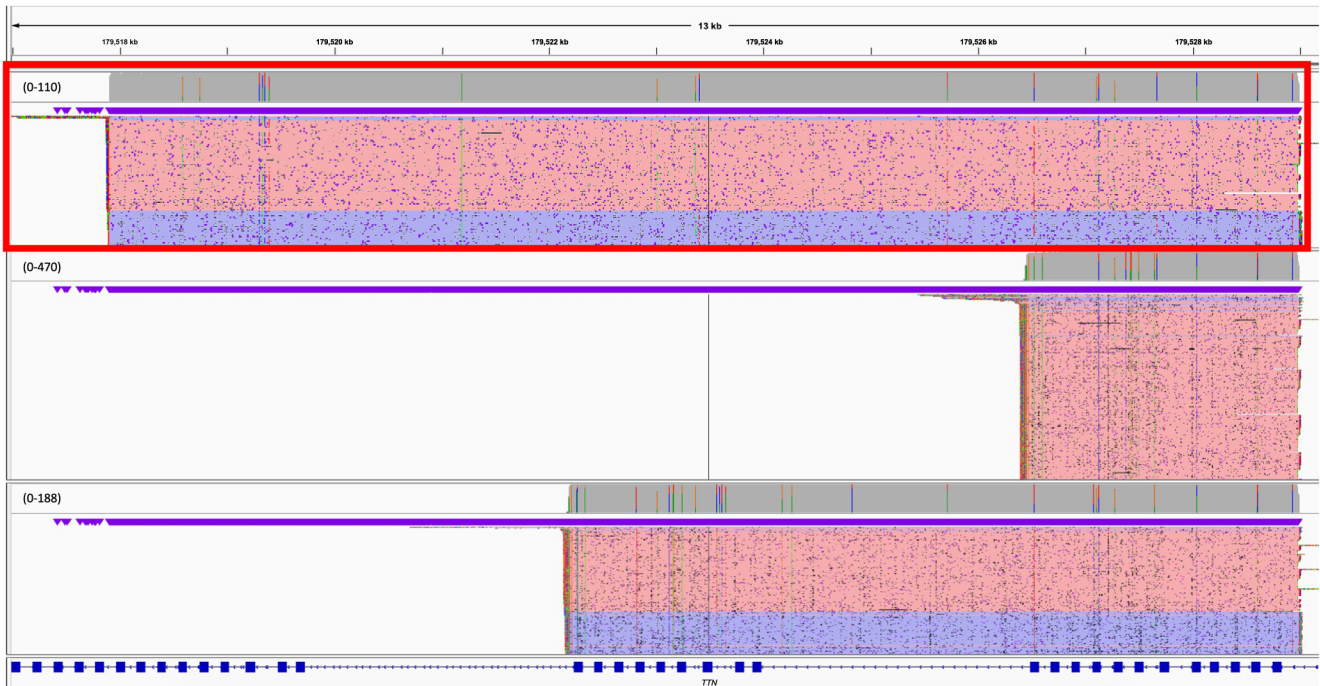
### Long-Range PCR and Long-Reads Sequencing Method Implementation

We first developed a strategy based on long-range PCR amplification using specific primers external to the repeats, followed by nested PCRs within the amplified amplicon. Unfortunately, this approach failed to localize precisely variants in the repeated regions of *TTN* because of the lack of specificity of the internal primers.

Therefore, we decided to evaluate another technique, the third-generation sequencing of long fragments, performed on the MinION sequencer (Oxford Nanopore Technologies) using the workflow shown in Figure 2A.

We implemented a long-range PCR by using oligonucleotides external to the three repeated blocks with Qiagen Long Range PCR kit (see *Materials and Methods* and Figure 2B). PCR amplification generated an amplicon of 11,088 bp at the expected size, as visualized by electrophoretic gel analysis (Figure 2C). Long-read sequencing, performed using the Nanopore MinION, leads to sequencing with a maximal depth at 1160×.

The sequencing data showed three populations of reads: one population corresponding to the 5' first block-only



**Figure 3** Bioinformatic sorting of the three populations of reads. Integrative Genomics Viewer screenshot, representing the three populations of reads identified after sequencing and bioinformatic sorting of the three reads populations. The population of reads of interest (with a size of >9000 bp) is surrounded by a red frame, showing an average coverage of 110 reads. The other two populations of reads covering the first repeat and the first two repeat areas were discarded for analysis. They included 470 and 188 reads of sequencing, respectively.

sequences (exons 173 to 181), a second population with the two 5' first blocks only (exons 173 to 190), and a third population containing the expected three blocks amplicon sequences (exons 173 to 199) (Figure 2D). The two additional populations are probably related to a lack of the 3' PCR primer specificity on intron 199 because sequence homologies extend beyond the three blocks of repeated exons.<sup>8</sup> We filtered all reads at <9000 bp, corresponding to the size of the entire repeat region, to eliminate artifactual reads that correspond to unspecific amplification. Although this filter decreases the number of reads in the region, the

resulting average depth after filtering (ie, 110 reads) was sufficient for reliable analysis of the results (Figure 3).

### Analysis of Control DNAs with Variants in the Repeated Region

We analyzed six genomic DNAs containing eight heterozygous genomic variants in the repeated region of *TTN* previously identified by short-read NGS, from patients diagnosed as having or suspected of having titinopathy. Among these variants, three were one-nucleotide deletions,

**Table 1** Control DNAs with Variants in the Repeated Region Identified by Short-Reads NGS

Patient ID	Exon/intron localization <sup>†</sup>	Variant(s) in <i>TTN</i> (NM_001267550.1)
I794	e174/e183/e192	c.36620delC: p.(Pro12207Glnfs*740)
I969	e174/e183/e192	c.36682delG: p.(Glu12228Lysfs*719)
I422	e199/e190/e181	c.38767A>T: p.(Lys12923*)
	e195/e186/e177	c.38386G>T: p.(Glu12796*)
I728	e197/e188/e178	c.38660delA: p.(Lys12887ArgfsTer60)
I796	e195/e186/e177	c.38421_38437delinsC: p.(Pro12808Lysfs*134)
I984	i199/e190/e181	c.38791+1G>T: p.(?)
	e195/e186/e177	c.38437A>T: p.(Lys12813*)

NM accession numbers from <https://www.ncbi.nlm.nih.gov/nucore/388998876> (last accessed April 20, 2022).

<sup>†</sup>The nomenclature of the variant was randomly attributed to one of the three repeats.

ID, identifier; NGS, next-generation sequencing.

four were substitutions, and one was a 17-bp deletion associated with the insertion of one nucleotide (Table 1).

Long-reads sequencing data showed that each of the substitutions was identified in one of the three repeats with an allelic balance of approximately 50% (Figure 4). The single-nucleotide deletions identified for patients I794, I969, and I728 were present at  $\geq 50\%$  allelic balance in only one exon. The largest deletion of 17 nucleotides in patient I796 was detected in exon 195, by a decrease in sequencing coverage of about 50%, not observed for homologous exons 186 and 177. Patients I422 and I984 carry two one-nucleotide substitutions. These substitutions were detected only in one exon of these repeats. Moreover, these results confirm the presence of these variants in the heterozygous state in these patients. Analysis of the coverage of each of these substitutions allows us to be confident about the location of the variants in the identified exon. For example, the c.38386G>T variant of patient I422 was detected at 43% in exon 195 and 1% and 2% in homologous exons 186 and 177, respectively. Some reads detected variants in low proportion in homologous exons despite sorting of reads at >9000 bp. Nevertheless, the detection threshold was not sufficient to raise doubt on the localization of variants identified by long-reads sequencing.

Long-reads sequencing also allowed us to determine the phase of variants localized in the repeated region, which could be important to confirm or not the diagnosis of a recessive titinopathy. As an example, we present an 11-year-old patient presenting a phenotype of congenital myopathy, for which the short-reads sequencing of *TTN* repeated regions identified three genetic variants in *TTN* gene: intron 199 c.38791+1G>T (or intron 190 or 181), exon 195 c.38437A>T, p.(Lys12813\*) (or exon 186 or 177), and exon 60 c.17521\_17522delAC p.(Thr5841Phefs\*4). The patient was adopted, without information about his family, and no possibility to perform variant familial segregation.

After long-reads sequencing of the long-range PCR product, we analyzed sequencing data on Integrative Genomics Viewer. We confirmed that the variant c.38791+1G>T was in intron 199 and the variant c.38437A>T was in exon 195. To determine the haplotype of this patient, we selected the reads containing the variant c.38791+1G>T and thus we could determine that these same reads also contain the variant c.38437A>T. These two variants were thus located on the same allele (Figure 5A). We verified this strategy in another patient (I422) for whom we had no information on inheritance. For this second patient, we established that he carried two variants located in exons 195 (c.38386G>T) and 199 (c.38767A>T), each located on one allele, therefore in *trans*.

## Discussion

For the first time, we implemented a long-reads sequencing protocol to localize precisely variants in repeated domains

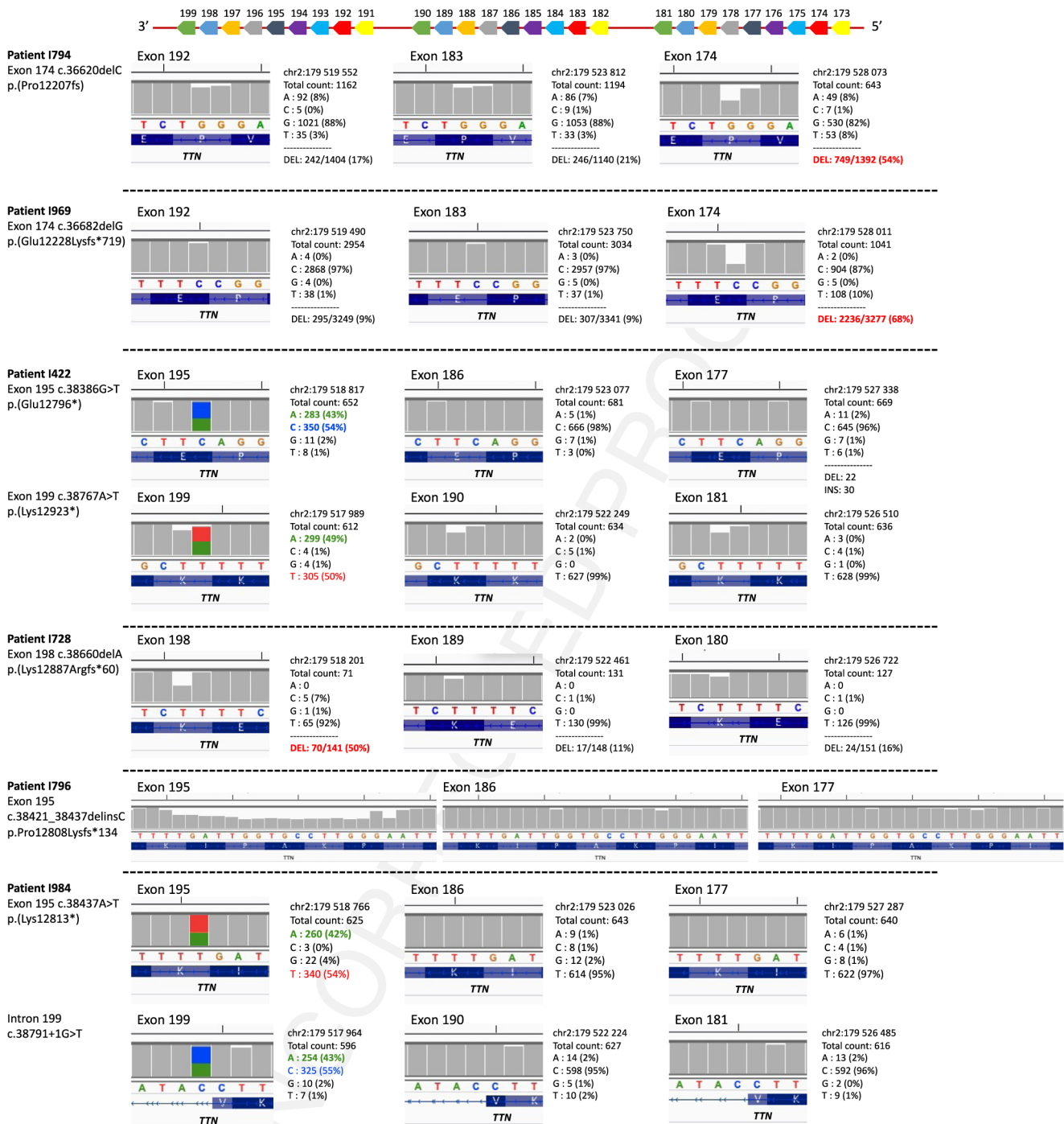
of the *TTN* gene and to quantify precisely the allelic frequency of each variant.

The prior amplification of the amplicon containing the three repeats by long-range PCR, which allows us to have high coverage rates, and the use of MinION technology ensure a confidence in the detection and localization of these variants. This strategy compensates the inability of targeted Sanger sequencing and short-reads NGS to precisely characterize variants on these repeated areas. It was described that long-reads sequencing, regardless of the technique, produces sequencing errors; however, this limitation can be mitigated using specific correction algorithms.<sup>9,10</sup> In our study, this limitation was not an issue because the objective of this approach was to verify the localization of variants already identified by NGS short reads to remove any ambiguity about location of *TTN* variants in the repeat areas, and determine haplotype segregation.

In addition to providing to patients an accurate diagnosis, this technology allows a reliable analysis of relatives and offers future prenatal diagnosis for carriers. Another interesting contribution of this technology is to define the phase of variants in the repeated region, essential to exclude or confirm a recessive titinopathy. This method could be useful for other genes that also contain a repeated sequence, such as the nebulin gene (*NEB*).<sup>11</sup>

The particularity of repeated region of *TTN*, beyond the sequence homology, is that is also a domain subject to strong alternative splicing during development.<sup>8</sup> According to the *TTN* database ([https://www.cardiodb.org/titin/titin\\_transcripts.php](https://www.cardiodb.org/titin/titin_transcripts.php), last accessed April 20, 2022) and the study by Savarese et al,<sup>8</sup> only a small subset of exons, 173 to 199, appears to be incorporated in post-natal skeletal muscle transcripts. However, the exons that are thought to be expressed only during fetal development (metatranscript-only exons) most likely have a role in skeletal muscle development, as truncating variants located in these exons are associated with recessive congenital myopathies. The few publications of congenital titinopathies due to *TTN* variants in the repeated regions have postulated that protein defects in the fetal period are responsible for the severe clinical picture at birth, but the observed stability of the disease after birth is thought to be due to the nonexpression of the mutated exon after birth.<sup>12,13</sup> An improvement of the patients' symptoms has been observed during their lifetime. Thus, accurately locating these variants is essential to evaluate their consequences on post-natal muscle transcripts, depending on whether they are included or skipped exons, and in the management of these patients and their relatives.

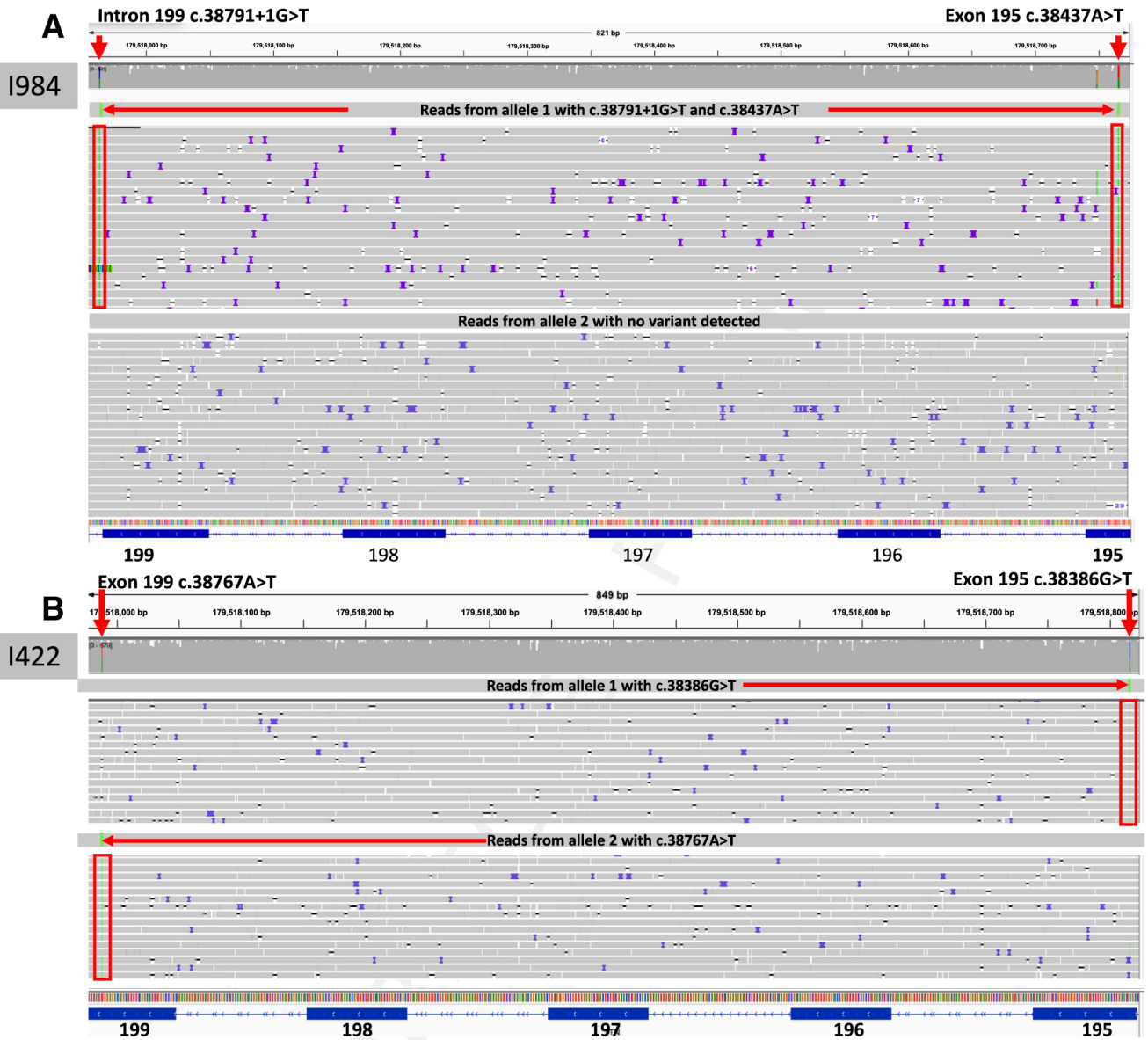
Other neuromuscular genetic diseases could benefit from the technological advances of long-read sequencing for the characterization of complex rearrangements, such as neurologic short tandem repeat expansion disorders<sup>14</sup> or facioscapulohumeral dystrophy, mostly related to the deletion of repeated elements at the 4q35 locus.<sup>15</sup> In



**Figure 4** Long-reads sequencing allows us to localize correctly genomic variants in *TTN* repeated domains. At the top of the panel, schematic representation of the exons of the repeated zones; each colored **arrow** represents an exon, and **arrows** of the same color are homologous exons. The analysis of the sequencing results shows for each patient the location of the variant(s) in the *TTN* repeat area. The variant is colored in red for deletion, and the sequencing coverage is represented by the gray histograms. The presence of two colors for the nucleotide substitutions represents the proportion of each nucleotide that is specified numerically in the associated numerical boxes. chr, chromosome.

infantile spinal muscular atrophy, mainly due to homozygous *SMN1* gene deletion, identification of heterozygous *SMN1* deletion in patients' relatives can be altered by the presence of two consecutive *SMN1* copies on the same allele.<sup>16,17</sup> Long-read sequencing could be interesting to detect these duplications located in the same haplotype.

Duchenne and Becker are muscular dystrophy resulting from variants in the *DMD* gene, mainly due to large genomic rearrangements (deletions or duplications) identified by quantitative methods, such as multiplex ligation-dependent probe amplification, comparative genomic hybridization array, or short-reads NGS. However, in rare



**Figure 5** Haplotyping variants in repeated area. **A:** Screen capture montage of Integrative Genomics Viewer (IGV) of Patient 1984. Sequencing reads were base sorted with the c.38791+1G>T variant selection in exon 199. The green bars show the position of the variants on each read. The reads containing the c.38791+1G>T variant also contain a second green bar in exon 195, corresponding to the c.38437A>T variant, establishing that the variants are in *cis*. **Bottom panel:** The absence of variants on the reads from the other allele. **B:** Screen capture montage of IGV of I422; the reads were sorted by base at the position of the variants c.38767A>T (exon 199; **bottom panel**) and at the position c.38386G>T (exon 195; **top panel**). The reads containing each of these variants (green bars in **red boxed area**) do not contain the other, establishing that the variants are in *trans*.

cases, structural variants are suspected; and long-reads sequencing of the *DMD* gene, including exonic and intronic regions, would be helpful to detect inversions or translocations with intronic breakpoints that escape conventional approaches.<sup>18</sup>

The contribution of long-read sequencing also has a promising future in the analysis of gene transcripts that are highly subject to alternative splicing under physiological and pathologic conditions to search for splicing defects in patients.<sup>19</sup>

Another interesting aspect is to search for new isoforms for large genes, to accurately determine the composition and

phasing of exons within transcripts, like in recent studies that identified and quantified previously unannotated transcripts.<sup>20</sup>

In conclusion, this long-read sequencing strategy is promising for the analysis of large genes, such as *TTN* or *NEB*, with a complex structure, important alternative splicing, or repeated sequences.

## Acknowledgments

We thank Emmanuelle Pion for proofreading the English version; and Gisèle Bonne and French *TTN* Consortium for discussion about *TTN* patients.



## References

1. Jäckel M, Witt C, Antonova O, Curdt I, Labeit S, Jockusch H: Deletion in the Z-line region of the titin gene in a baby hamster kidney cell line, BHK-21-Bi. *FEBS Lett* 1997, 408:21–24
2. Labeit S, Kolmerer B: The complete primary structure of human nebulin and its correlation to muscle structure. *J Mol Biol* 1995, 248: 308–315
3. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, et al: Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med* 2015, 7:270ra6
4. Chauveau C, Rowell J, Ferreiro A: A rising titan: TTN review and mutation update. *Hum Mutat* 2014, 35:1046–1059
5. Savarese M, Sarparanta J, Vihola A, Udd B, Hackman P: Increasing role of titin mutations in neuromuscular disorders. *J Neuromuscul Dis* 2016, 3:293–308
6. Juntas-Morales R, Perrin A, Cossée M: Corrélations phénotype-génotype dans les titinopathies. *Les Cah Myol* 2020:16–20
7. Zenagui R, Lacourt D, Pegeot H, Yauy K, Juntas Morales R, Theze C, Rivier F, Cances C, Sole G, Renard D, Walther-Louvier U, Ferrer-Monasterio X, Espil C, Arné-Bes MC, Cintas P, Uro-Coste E, Martin Negrier ML, Rigau V, Bieth E, Goizet C, Claustres M, Koenig M, Cossée M: A reliable targeted next-generation sequencing strategy for diagnosis of myopathies and muscular dystrophies, especially for the giant titin and nebulin genes. *J Mol Diagn* 2018, 20:533–549
8. Savarese M, Jonson PH, Huovinen S, Paulin L, Auvinen P, Udd B, Hackman P: The complexity of titin splicing pattern in human adult skeletal muscles. *Skelet Muscle* 2018, 8:11
9. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q: Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020, 21:30
10. Zhang H, Jain C, Aluru S: A comprehensive evaluation of long read error correction methods. *BMC Genomics* 2020, 21:889
11. Lehtokari VL, Pelin K, Sandbacka M, Ranta S, Donner K, Muntoni F, Sewry C, Angelini C, Bushby K, Van Den Bergh P, Iannaccone S, Laing NG, Wallgren-Pettersson C: Identification of 45 novel mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. *Hum Mutat* 2006, 27:946–956
12. 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
13. Oates EC, Jones KJ, Donkervoort S, Charlton A, Brammah S, Smith JE, et al: Congenital titinopathy: comprehensive characterization and pathogenic insights. *Ann Neurol* 2018, 83:1105–1124
14. Chintalaphani SR, Pineda SS, Deveson IW, Kumar KR: An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol Commun* 2021, 9:98
15. Mitsuhashi S, Nakagawa S, Takahashi Ueda M, Imanishi T, Frith MC, Mitsuhashi H: Nanopore-based single molecule sequencing of the D4Z4 array responsible for facioscapulohumeral muscular dystrophy. *Sci Rep* 2017, 7:14789
16. Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, Le Paslier D, Frézal J, Cohen D, Weissenbach J, Munnich A, Melki J: Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 1995, 80:155–165
17. Ogino S, Wilson RB, Gold B: New insights on the evolution of the SMN1 and SMN2 region: simulation and meta-analysis for allele and haplotype frequency calculations. *Eur J Hum Genet* 2004, 12: 1015–1023
18. Chin HL, O'Neill K, Louie K, Brown L, Schlade-Bartusiak K, Eydoux P, Rupps R, Farahani A, Boerkoel CF, Jones SJM: An approach to rapid characterization of DMD copy number variants for prenatal risk assessment. *Am J Med Genet A* 2021, 185:2541–2545
19. Savarese M, Qureshi T, Torella A, Laine P, Giugliano T, Jonson PH, Johari M, Paulin L, Piluso G, Auvinen P, Nigro V, Udd B, Hackman P: Identification and characterization of splicing defects by single-molecule real-time sequencing technology (PacBio). *J Neuromuscul Dis* 2020, 7:477–481
20. Uapinyoying P, Goecks J, Knoblach SM, Panchapakesan K, Bonnemann CG, Partridge TA, Jaiswal JK, Hoffman EP: A long-read RNA-seq approach to identify novel transcripts of very large genes. *Genome Res* 2020, 30:885–897