



HAL
open science

Identification of a CCG-enriched expanded allele in patients with myotonic dystrophy type 1 using amplification-free long-read sequencing

Yu-Chih Tsai, Laure de Pontual, Cheryl Heiner, Tanya Stojkovic, Denis Furling, Guillaume Bassez, Geneviève Gourdon, Stéphanie Tomé

► To cite this version:

Yu-Chih Tsai, Laure de Pontual, Cheryl Heiner, Tanya Stojkovic, Denis Furling, et al.. Identification of a CCG-enriched expanded allele in patients with myotonic dystrophy type 1 using amplification-free long-read sequencing. *Journal of Molecular Diagnostics*, In press, 10.1016/j.jmoldx.2022.08.003 . hal-03832574

HAL Id: hal-03832574

<https://hal.science/hal-03832574v1>

Submitted on 27 Oct 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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



TECHNICAL ADVANCE

Identification of a CCG-Enriched Expanded Allele in Patients with Myotonic Dystrophy Type 1 Using Amplification-Free Long-Read Sequencing

Q1

Q26

Yu-Chih Tsai,* Laure de Pontual,[†] Cheryl Heiner,* Tanya Stojkovic,[‡] Denis Furling,[†] Guillaume Bassez,^{†‡} Geneviève Gourdon,[†] and Stéphanie Tomé[†]

From the Pacific Biosciences,* Menlo Park, California; the Sorbonne Université,[†] Inserm, Institut de Myologie, Centre de Recherche en Myologie, Paris, France; and the Centre de Référence des Maladies Neuromusculaires Nord-Est/Île-de-France,[‡] Institut de Myologie, GHU Pitié-Salpêtrière, AP-HP, Paris, France

Accepted for publication
August 11, 2022.

Q5 Q6

Address correspondence to Stéphanie Tomé, Repeat Expansions & DM Centre de Recherche en Myologie, UMRS974 Sorbonne Université/INSERM/AIM47, bld de l'hôpital—G.H. Pitié-Salpêtrière—Bâtiment Babinski, 75651 Paris Cedex 13, France. E-mail: stephanie.tome@inserm.fr.

Myotonic dystrophy type 1 (DM1) exhibits highly heterogeneous clinical manifestations caused by an unstable CTG repeat expansion reaching up to 4000 CTG. The clinical variability depends on CTG repeat number, CNG repeat interruptions, and somatic mosaicism. Currently, none of these factors are simultaneously and accurately determined due to the limitations of gold standard methods used in clinical and research laboratories. An amplicon method for targeting the *DMPK* locus using single-molecule real-time sequencing was recently developed to accurately analyze expanded alleles. However, amplicon-based sequencing still depends on PCR, and the inherent bias toward preferential amplification of smaller repeats can be problematic in DM1. Thus, an amplification-free long-read sequencing method was developed by using CRISPR/Cas9 technology in DM1. This method was used to sequence the *DMPK* locus in patients with CTG repeat expansion ranging from 130 to >1000 CTG. We showed that elimination of PCR amplification improves the accuracy of measurement of inherited repeat number and somatic repeat variations, two key factors in DM1 severity and age at onset. For the first time, an expansion composed of >85% CCG repeats was identified by using this innovative method in a DM1 family with an atypical clinical profile. No-amplification targeted sequencing represents a promising method that can overcome research and diagnosis shortcomings, with translational implications for clinical and genetic counseling in DM1. (*J Mol Diagn* 2022, ■: 1–12; <https://doi.org/10.1016/j.jmoldx.2022.08.003>)

Q7

Short tandem repeats are an important source of genetic variation and phenotypic variability in disease and health that are not always well characterized and understood due to their complexities. Among these repeated elements, unstable repeat expansions are associated with >20 diseases, including the complex and variable myotonic dystrophy type 1 (DM1) disorder.¹ DM1 is caused by an unstable CTG repeat expansion in the 3' untranslated region of the DM1 protein kinase (*DMPK*) gene, which usually increases across generations and over time in somatic tissues.^{2,3} In DM1, longer expanded alleles are usually associated with a worsening of clinical severity and an earlier age of onset.⁴ This anticipation phenomenon is particularly obvious in DM1.⁵ DM1 is mainly characterized by an unusually broad

clinical spectrum of symptoms divided into five distinct clinical forms ranging from late onset to the congenital forms, which are often associated with the largest size of inherited disease—associated allele.⁴ Facial dysmorphism, muscle weakness, and cognitive impairment are more frequent symptoms at an earlier onset, whereas cardiac defects and cataracts are more common in DM1 patients with later forms of the disease.

Supported by the Institut National de la Santé et de la Recherche Médicale, Sorbonne Université, Institut Mérieux (Accelerator project) and the 2019 Targeted Sequencing SMRT Grant. Q3 Q3

Disclosures: Y.-C.T. is a full-time employee at Pacific Biosciences, a company developing single-molecule sequencing technologies.

In DM1, it is laborious, for several reasons, to diagnose patients and classify them into distinct clinical categories based exclusively on mutation status and size of CTG repeats when known. First, the long-expanded allele and the precise number of CTG repeats are complicated to identify and measure by using conventional methods, particularly for the larger repeat expansions.^{6,7} Second, patients with DM1 exhibit high clinical and genetic variability that cannot be exclusively explained by the size of the CTG repeats.⁴ Third, additional factors such as somatic mosaicism are important disease modifiers contributing to the high genotype and phenotype variability observed in DM1. Somatic mosaicism biased toward expansions contributes to the progressive nature of the various DM1 symptoms and also to the variation in the age of onset.^{8–11} Fourth, the majority of patients with DM1 inherit pure CTG repeat expansion. However, >8% of patients with known DM1 carry interruptions that vary in type (CCG, CAG, CTC, and CGG) and number between families and also among individuals of the same family.¹² Interruptions are frequently associated with intergenerational contractions and stabilization of the CTG repeat as well as milder DM1 symptoms and/or additional symptoms.^{11,13–22} Fifth, single-nucleotide polymorphisms in the DNA mismatch repair gene *MSH3*, required for maintenance of genomic integrity, have been shown to reduce levels of somatic mosaicism and are associated with delayed onset in patients with DM1.^{23,24} The contribution of CTG repeat length, somatic mosaicism, structural variants, and/or modifier genes on the DM1 genotype and phenotype remains poorly understood due to technical difficulties in analyzing these factors.

Currently, the molecular diagnostic procedure used in DM1 genetic testing is divided into two steps.^{6,7} The first is to detect or rule out possible DM1 expansions using PCR and fragment-length analysis (Table 1). This method cannot differentiate between individuals homozygous for a normal allele and individuals whose CTG repeat expansion could not be amplified by PCR. A second step is therefore necessary to identify large expanded alleles and estimate the approximate size by using Southern blot. All these methods are time-consuming procedures that do not simultaneously provide information on somatic mosaicism and structural variants of expansion, two important prognostic parameters. Furthermore, triplet-repeat primed PCR (TP-PCR) is also used to directly detect the presence of an expanded allele.²⁵ However, no size repeat can be estimated by this method, which leads to a loss of information necessary for performing genetic counseling. To date, TP-PCR is the only diagnostic method, which makes it possible to reveal variant repeats in the first 100 bp within the 5' and 3' end of the CTG array. However, this method is unable to identify interruptions at the middle of large repeat expansions of >200 to 300 CTG repeats and to provide the exact number and type of interruptions, regardless of the size of the repeats. Diagnostic laboratories generally use several methodologic approaches to detect CTG repeat expansion and roughly

estimate the size of the repeat, which leads to long diagnostic delays. Unfortunately, no diagnostic method can accurately estimate somatic mosaicism, which is an important parameter in predicting disease progression.^{8–11}

Currently, the most accurate method to assess the inherited CTG repeat expansion size and the level of somatic mosaicism is the small-pool PCR (SP-PCR).^{9,26} SP-PCR (PCR on successive DNA dilutions) is a time-consuming and fastidious method, used exclusively in research laboratories. The PCR amplification bias toward shorter repeats and the difficulties inherent in the amplification of the GC-rich *DMPK* locus by this method remain a concern, which must be addressed by a new, widely available method. In addition, neither SP-PCR nor Southern blot can detect CNG interruptions within the repeats, which can lead to inaccurate genetic analysis and genotype–phenotype associations. Currently, molecular diagnostic tests used in diagnostic laboratories to analyze the *DMPK* locus may induce loss of information and bias in the interpretation of genetic data, especially when patients carry large CTG repeats. It is therefore essential to accurately measure the number of CTG repeats, structural variants, and somatic mosaicism in patients with DM1 using a fast and accurate method to provide better diagnosis and prognosis with minimal delay.

Targeted single-molecule long-read real-time sequencing from PCR products by Pacific Biosciences (Menlo Park, CA) was recently described to analyze the *DMPK* locus in patients.^{16,27} This method allows accurate counting of repeat numbers in large expanded alleles, and it can simultaneously evaluate the degree of somatic mosaicism and identify interruptions.²⁷ However, amplicon-based sequencing relies on PCR and is still subjected to the inherent preferential amplification of smaller repeats, which introduces a confounding and undesirable bias in repeat sizing. To overcome this limitation, we performed a pilot study in 11 patients with DM1 using a robust amplification-free targeted long-read sequencing (Pacific Biosciences) to simultaneously analyze the size and sequence of CTG repeat expansions and somatic mosaicism. Using this method, measurement of CTG repeat length and somatic mosaicism is improved in DM1 patient samples. The second advantage of this method is that CNG interruptions are easily detected regardless of the size of the CTG repeat expansions. Strikingly, and for the first time, triplet repeat expanded alleles composed of >85% of CCG were identified in patients with DM1 for whom amplification of the triplet repeat expansion failed according to PCR and TP-PCR.

Materials and Methods

DM1 Patient Genomic DNA Samples

Participants with DM1 were recruited by the Genetics Department of Nantes Hospital, the Genetics Department of the Necker–Enfants Malades Hospital, the DM-Scope

Table 1 Genetic Testing Strategy for DM1

Information	PCR	Southern blot	TP-PCR
Mutation status	✓	✓	✓
CTG repeat size <200-300	✓	✓	✓
Large CTG repeat size	x	✓	x
Structural variants	x	x	x
Somatic mosaicism	x	✓	x

DM1, myotonic dystrophy type 1; TP-PCR, triplet-repeat primed PCR.

registry, and the Neuromuscular Disease Reference Center of Pitié-Salpêtrière Hospital in France. Written informed consent was obtained from all participants.

Genomic DNA samples were initially genotyped for DM1 by using conventional PCR, TP-PCR, or Southern blot.^{6,7} DNA was extracted in diagnostic laboratories. High-molecular weight DNA was extracted from immortalized lymphoblastoid cells by using the Monarch Genomic DNA Purification Kit (catalog no. T3010S; New England Biolabs, Evry-Courcouronnes, France).

Design of Guide RNAs for Cas9 Digestion

Human Hg19 reference sequences surrounding the DM1 repeat region were used to design Cas9 CRISPR RNA (crRNA) sequences. Candidate target sequences were generated by using the Genetic Perturbation Platform sgRNA Designer webtool on the Broad Institute website. The latest design tool, which provides results the same as the earlier version of the tool, is available through the link (Broad Institute, <https://portals.broadinstitute.org/gppx/crispick/public>, last accessed July 25, 2022). Two upstream and three downstream top ranking candidate sequences, which could generate approximately 2 kb target fragments, were custom synthesized and tested in targeted sequencing experiments with HEK293 genomic DNA. The combination of upstream and downstream crRNAs that produced the highest sequencing yield was selected for the current study. The final crRNA sequences are the DM1-L1-crRNA sequence (5'-

CCCCATCGGGACAACGCAGA-3') and the DM1-R1-crRNA sequence (5'-GGGCGTGTATAGACACCTGG-3'), which generate a target capture region of 2361 bp.

SMRTbell Library Preparation

Preparation of the SMRTbell library was performed according to the "Procedure and Checklist-No-Amp Targeted Sequencing utilizing the CRISPR-Cas9 system" (available on the Pacific Biosciences website) per manufacturer instructions.²⁸ Five patients were analyzed by using SMRT Cell. Approximately 3 to 5 µg of native genomic DNA was dephosphorylated with 0.05 U/µL of shrimp alkaline phosphatase for each patient (catalog no. M0371S; New England Biolabs) to reduce the amount of off-target molecules in the final SMRTbell library. Guide RNAs (gRNAs) were formed by annealing crRNAs (Integrated DNA Technologies, Leuven, Belgium) to tracrRNAs (catalog no. 1072533; Integrated DNA Technologies) at a 1:1 ratio and a 5 µM concentration. The sequences of each crRNA targeting DM1 and various control loci used in the experiments are shown in Table 2. The gRNA-Cas9 complex was prepared by incubating 400 nmol/L gRNA with 400 nmol/L Cas9 nuclease (catalog no. M0386M; New England Biolabs) for 10 minutes at 37°C. Dephosphorylated genomic DNAs were digested with the gRNA-Cas9 complex for 1 hour at 37°C. DM1 and control gRNA were multiplexed in the same digestion reaction. Cas9 digestion products were purified by using a 0.45X volume of AMPure PB Beads (catalog no. 100-265-900; Pacific Biosciences). Barcoded adapters recommended by Pacific Biosciences were ligated to purified Cas9 digestion products using 0.4 µmol/L barcode adapters and 0.9U/µL T4 DNA ligase (catalog no. EL0013; Thermo Fisher Scientific, Illkirch-Graffenstaden, France) to form a library of symmetric SMRTbell template molecules. The SMRTbell library was purified by using a 0.45X volume of AMPure PB Beads (catalog no. 100-265-900; Pacific Biosciences). Failed ligation products and gDNA fragments were removed by nuclease treatment using 1.2 U/µL Exonuclease III (catalog no. M0206L; New

Table 2 crRNA and Genomic Coordinate for *DMPK* and Control Loci

Locus	crRNA1	crRNA2	GRCh38	Size of the fragment,* bp (non-expanded allele)
<i>DMPK</i>	5'-CCCCATCGGGACAACGCAGA-3'	5'-GGGCGTGTATAGACACCTGG-3'	Chromosome 19 (45,769,144-45,771,505)	2328
<i>HTT</i>	5'-CTTATTAACAGCAGAGAACT-3'	5'-TAAACTTTGAAGACGAGACA-3'	Chromosome 4 (3,072,601-3,076,440)	3806
<i>C9orf72</i>	5'-TTGGTATTTAGAAAGGTGGT-3'	5'-GGAAGAAAGAATTGCAATTA-3'	Chromosome 9 (27,571,393-27,575,065)	3639
<i>FMR1</i>	5'-CGCGCGTCTGTCTTTTCGACC-3'	5'-CCTTTATGCAAAGTTAGCTC-3'	Chromosome X (147,911,739-147,915,434)	3662

*Fragment size after CRISPR/Cas9 digestion.

crRNA, CRISPR RNA; DM1, myotonic dystrophy type 1; GRCh38, Genome Reference Consortium Human Build 38.

Table 3 HiFi Reads in Patients with DM1

Individuals	#11	#18	#L1	#L3	#A3	#15969	#15841
Tissue	Blood	Blood	Blood	Blood	Lymphoblastoid cells	Lymphoblastoid cells	Lymphoblastoid cells
DNA extraction	ND	ND	ND	ND	Monarch Genomic DNA Purification Kit (NEB)	Monarch Genomic DNA Purification Kit (NEB)	Monarch Genomic DNA Purification Kit (NEB)
HiFi reads	841	1041	836	3731	7405	11,201	4429
<i>DMPK</i> locus	452	352	332	114	569	733	597
Control loci	None	None	None	None	4522	3073	3218
Expanded allele	206	205	65	21	303	398	361
Non-expanded allele	246	147	262	93	266	335	236
% Reads on-target	53.7	33.8	39.71	3.06	68.75*	33.98*	86.14*
% Reads off-target	46.3	66.2	60.29	96.94	31.25	66.02	13.86

*Includes *DMPK* and control loci reads.

DM1, myotonic dystrophy type 1; HiFi reads, highly accurate long-reads.

England Biolabs) and an Enzyme Clean Up Kit (catalog no. 101-746-400; Pacific Biosciences) for 2 hours at 37°C. After nuclease treatment, the SMRTbell library was incubated with 41 µg/mL SOLu-Trypsin (catalog no. EMS0004; Sigma-Aldrich; Saint-Quentin-Fallavier, France) for 20 minutes at 37°C. The nuclease-treated SMRTbell library was purified by using a 0.45X volume of AMPure PB beads for the first purification and a 0.42X volume of AMPure PB Beads for the second purification. Finally, the purified SMRTbell library was resuspended in 6.3 µL of elution buffer (catalog no. 101-633-500; Pacific Biosciences) for targeted SMRT sequencing.

Targeted SMRT Sequencing

Following the “Procedure and Checklist-No-Amp Targeted Sequencing Utilizing the CRISPR-Cas9 System” per manufacturer instructions,²⁸ final SMRTbell libraries were annealed with Sequencing Primer v4 from the No-Amp Accessory Kit (catalog no. 101-788-900; Pacific Biosciences) and bound with Sequel DNA Sequencing Polymerase 3.0 in the Sequel Binding Kit 3.0 (catalog no. 101-626-600; Pacific Biosciences). Polymerase-bound SMRTbell complexes were purified with 0.6X AMPure PB Beads and sequenced on the Sequel System using Sequel Sequencing Kit 3.0 (Pacific Biosciences) and a customized setting with 4 hours of complex immobilization and 20-hour movie collection.

Sequencing Data and Repeat Sequence Analysis

Sequencing data repeat sequence analysis was performed following the “Analysis Procedure—No-Amp Data Preparation and Repeat Analysis” (available on the Pacific Biosciences website) per manufacturer instructions.²⁸ Barcode demultiplexed highly accurate long-reads (ie, HiFi reads) were first generated by using the Pacific Biosciences SMRT Link software version 8.0 or later version. HiFi reads provide an accuracy of >99.9%. HiFi reads corresponding to each sample were mapped to the Genome Reference

Consortium Human Build 38 human reference sequence, allowing on-target read identification and visual inspection of sequencing reads using Integrative Genomics Viewer.²⁹ Repeat sequence analysis was then performed by using the RepeatAnalysisTools and instruction available through GitHub (GitHub, <https://github.com/PacificBiosciences/apps-scripts/tree/master/RepeatAnalysisTools>, last accessed July 25, 2022) to generate various repeat sequence statistics.

Amplicon Long-Read Targeted Sequencing

DMPK locus was characterized in DM1 individuals #11 and #18 according to the method described in Mangin et al.²⁷ Briefly, normal and expanded CTG repeat alleles were amplified by PCR using barcoded ST300-F (5'-GAACTGTCTTCGACTCCGGG-3') and ST300-R (5'-GCACTTTGCGAACCAACGAT-3') primers. To enrich the SMRTbell library with the expanded allele, the PCR products were purified by using a 0.45 to 0.5X volume of AMPure PB Beads (catalog no. 100-265-900; Pacific Biosciences) before SMRT sequencing. SMRTbell libraries were prepared by using the SMRTbell Express Template Prep Kit 2.0 (catalog no. 100-938-900; Pacific Biosciences) and sequenced with the Sequel System.

Small-Pool PCR

SP-PCR amplifications and PCR product electrophoretic analyses were performed according to the methods of Gomes-Pereira et al.³⁰ Briefly, blood DNA samples were digested with *HindIII* enzyme, and SP-PCR was performed by using DM-C (5'-AACGGGGCTCGAAGGGTCCT-3') and DM-BR (5'-CGTGGAGGATGGAACACGGAC-3') primers.²⁶ The conditions of PCR were as follows: denaturation at 96°C for 5 minutes followed by 30 cycles at 96°C for 45 seconds, 68°C for 30 seconds, and at 70°C for 3 minutes with a chase of 1 minute at 68°C and 10 minutes at 72°C. SP-PCR products were loaded onto a 1% agarose gel to compare instability between DM1 patients #11 and #18 and run at 300 V for 30 minutes followed by 16 hours at 160

Table 4 CTG Repeat Size and Interruptions in Patients with DM1

Individuals	#11	#18	#L1	#L3	#A3	#15969	#15841
Tissue	Blood	Blood	Blood	Blood	Lymphoblastoid cells	Lymphoblastoid cells	Lymphoblastoid cells
Age at sample (years)	20	42	29	26	40	36	58
Median expanded allele (No-Amp)	516	495	1378	1287	141	158	467
Median expanded allele (amplicon)	468	460	563	927	#NA	#NA	Failed
Interruptions (No-Amp)	2CCG	4CCG	Pure	Pure	1CAG	Pure	78CCG
Interruptions (amplicon)	2CCG	4CCG	Pure	Pure	#NA	#NA	Failed
CNG repeat size (diagnosis)	≅ 700	≅ 400	>1000	>1000	130	150	370

DM1, myotonic dystrophy type 1; No-Amp, no-amplification targeted sequencing.

V. The PCR products were then transferred to GeneScreen Plus Hybridization Transfer Membrane (catalog no. NEF988001PK; PerkinElmer SAS, Villebon-sur-Yvette, France) and detected by hybridization using a non-radioactive method as described in Tomé et al.³¹

Results

Amplification-free *DMPK* Locus Enrichment in Patients with DM1

Amplicon-based long-read sequencing depends on PCR, which has an inherent bias toward smaller repeats. To overcome this limitation, an amplification-free long-read sequencing method (No-Amp) was used to specifically characterize the *DMPK* locus. This method uses the CRISPR/Cas9 system to target and isolate the DNA fragment of interest from genomic DNA, in combination with

long-read single-molecule real-time (SMRT) sequencing. This method was performed in seven patients with DM1 using DNA isolated either from blood or lymphoblastoid cell lines. These patients carried different CTG repeat expansion lengths that were previously defined by using Southern blot and/or PCR. The number of generated on-target HiFi reads was sample dependent and ranged from 114 to 733 *DMPK* reads (Table 3). The CTG repeat region was successfully captured in all samples, with a better result for samples #A3, #15969, and #15841. For those samples, DNA was extracted with a high-molecular weight DNA extraction kit (New England Biolabs). Off-target reads were observed for all samples and represent an average number of approximately 54%. No significant accumulation of off-target sequencing reads mapping to a specific genome location was observed in the samples digested exclusively with *DMPK* locus gRNAs; all the off-target sequencing reads appeared to be randomly distributed across the

CNG repeat distribution

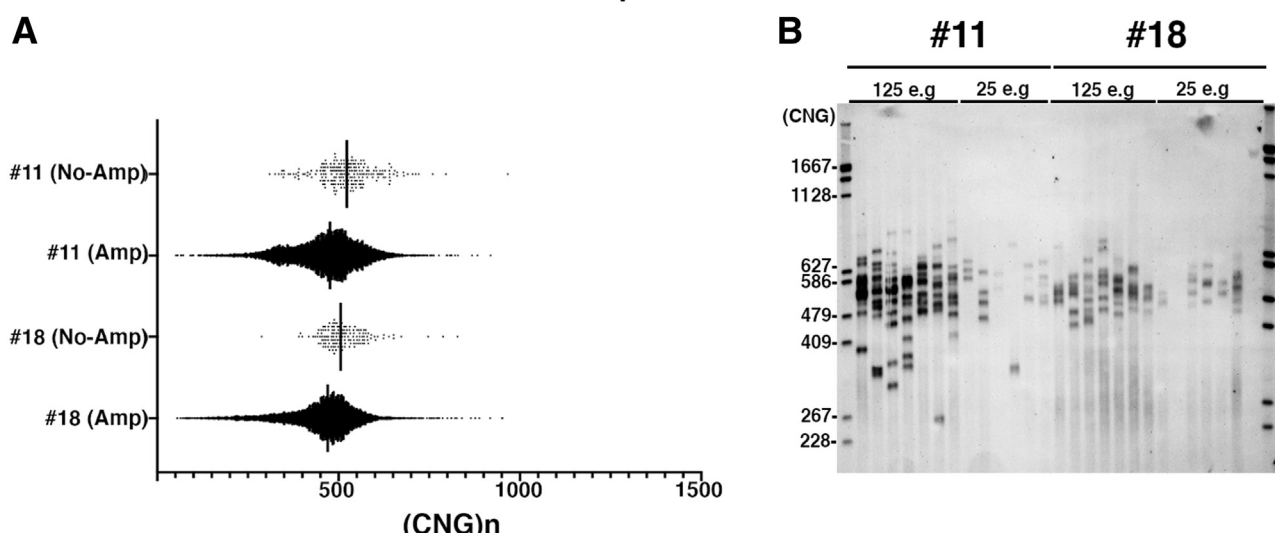


Figure 1 Somatic mosaicism in individuals #11 and #18. **A:** The distribution of the size of the repeat of each highly accurate long-read (HiFi read) generated by the Pacific Biosciences algorithm. The x axis shows the CTG repeat number. Each dot represents the CTG repeat size of one HiFi read. The black bar represents the median. A total of 10,000 to 30,000 single-molecule HiFi reads of expanded allele are analyzed in the amplicon (Amp) experiment, whereas >200 single-molecule HiFi reads of expanded allele are analyzed in the no-amplification (No-Amp) experiment. **B:** Trinucleotide repeat instability estimation by small-pool PCR in blood samples of patients with myotonic dystrophy type 1. The sizes (converted in CTG repeats) of DNA Molecular Weight Marker III, DIG-labeled are indicated on the left of the figure. About 25 to 125 DNA molecules are amplified for each individual. e.g., equivalent genome.

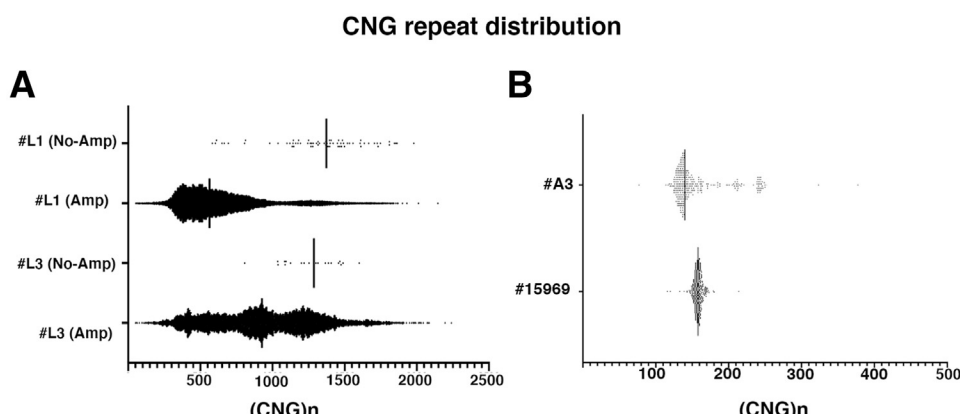


Figure 2 Distribution of the size of the repeat of each highly accurate long-read (HiFi read) generated by the Pacific Biosciences algorithm. **A:** Individuals #L1 and #L3. **B:** Individuals #A3 and #15969. The x axis shows the CTG repeat number. Each **dot** represents the CTG repeat size of one HiFi read. The **black line** represents the median. A total of 10,000 to 30,000 single-molecule HiFi reads of expanded allele are analyzed in the amplicon (Amp) experiment, whereas 21 to 398 single-molecule HiFi reads of expanded allele are analyzed in the no-amplification (No-Amp) experiment.

Genome Reference Consortium Human Build 38 reference (Supplemental Figure S1). However, some specific off-target reads on chromosome 2 (chr2:116779778-116785717) and chromosome 13 (chr13:89237016-89240938) were captured in two of the three samples digested by DM1 and control locus gRNAs (Table 2 and Supplemental Figure S1). These specific off-target reads are potentially produced by off-target gRNA sequence (DMPK-crRNA2, C9orf72-crRNA1, and C9orf72-crRNA2) matches with two to four mismatched bases (data not shown). Off-target as well as control reads were filtered out during the bioinformatics analysis and did not affect analysis of DM1-specific reads. For samples #L1 and #L3 carrying the largest CTG repeat sizes, the number of highly accurate long-reads from the expanded allele was lower than for the other samples in this study. The lower sequencing yields for these large alleles are likely the result of reduced efficiency of immobilization of the DNA polymerase complex into the SMRT Cell via diffusion and lower raw reads to HiFi reads conversion efficiency due to significantly longer template lengths.

Targeted Amplification-free Long-Read Sequencing Improves the Accuracy of CTG Repeat Size Measurement

This is the first time that amplification-free targeted sequencing has been described to generate long-read sequencing data analysis for the *DMPK* locus using the RepeatAnalysisTools developed by Pacific Biosciences (see Materials and Methods for details). By amplification-free sequencing, the number of CTG repeats was quantified in seven patients with DM1 and ranged from 141 to 1378 repeats (median) (Table 4). The expanded alleles showed a median CTG repeat of 516 and 495 in patients #11 and #18, respectively, which was a similar range to that observed for

amplicon target sequencing (Table 4). On the contrary, individuals #L1 and #L3, whose size was estimated to be >1000 CTGs in diagnosis, displayed different CTG repeat medians between amplicon sequencing and No-Amp targeted sequencing. The median CTG repeat size was estimated to be 563 and 927 CTG in individuals #L1 and #L3, respectively, by the amplicon sequencing experiment, and 1378 CTG and 1287 CTG by the No-Amp experiment (Table 4). These results strongly suggest that amplicon sequencing is subjected to the inherent preferential amplification of the smallest repeats, particularly in individuals #L1 and #L3 carrying the largest inherited expanded alleles. Thus, PCR may introduce a confounding and undesirable bias in repeat sizing in patients with DM1. The median CTG repeat size for the expanded allele in samples #A3 and #15969, based on No-Amp targeted sequencing, is consistent with previous amplicon targeted sequencing data from Mangin et al.²⁷ Interestingly, the number of repeats in patient #15841 was estimated to be 467 CNG repeats by the No-Amp method in lymphoblastoid cell lines. This size estimate could not be performed before by using PCR as it failed several times.

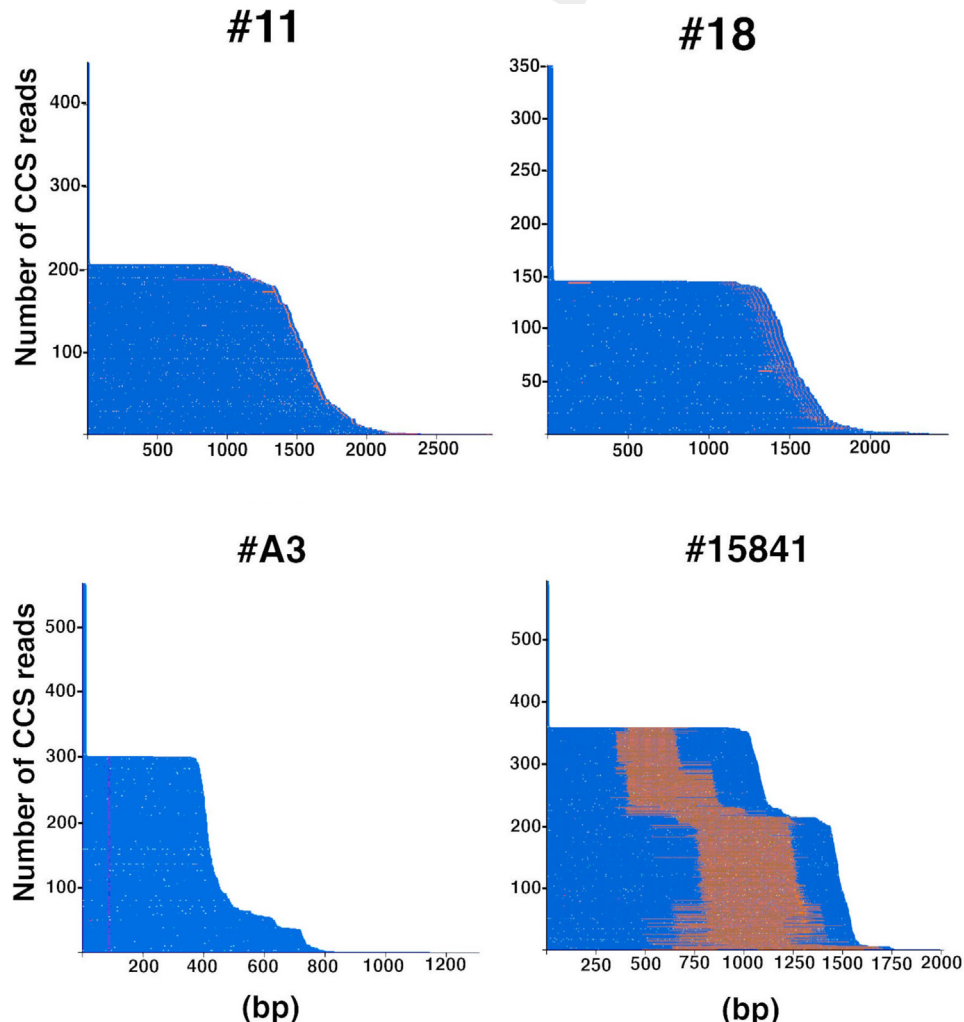
Targeted Amplification-free Long-Read Sequencing Exhibits Somatic Mosaicism in Patients with DM1

The distribution of CNG repeats was analyzed in individuals #11 and #18 carrying <500 interrupted CNG repeats by targeted long-read sequencing as well as by SP-PCR, which is usually used to estimate somatic mosaicism in patients with DM1. All these methods, including SP-PCR, showed that somatic mosaicism was lower in individual #18 than in individual #11 (Figure 1). Despite similar somatic mosaicism observed with the different methods, a PCR bias in which the shortest alleles were overrepresented using long-read sequencing of PCR products (amplicon) is clearly

745 observed. Interestingly, approximately 200 single-molecule
 746 HiFi reads of the expanded allele appear sufficient to
 747 accurately estimate somatic mosaicism in patients by no-
 748 amplification targeted sequencing (Figure 1). Somatic
 749 mosaicism was also observed in blood from patients #L1
 750 and #L3 carrying >1000 pure CTG repeats at the *DMPK*
 751 locus (Figure 2A). Despite a limited number of HiFi reads
 752^[F2] obtained, no-amplification targeted sequencing provided a
 753 better estimate of CTG repeat number distribution in pa-
 754 tients with large expansions in which PCR bias is extremely
 755 high using amplification targeted sequencing. To comple-
 756 ment our data, we have also shown that lymphoblastoid cell
 757 lines from individuals #A3 and #15969 exhibit somatic
 758 mosaicism to varying degrees, with the highest being
 759 observed in patient #A3 (Figure 2B).

760 Targeted Amplification-free Long-Read Sequencing 761 Detects CNG Interruptions

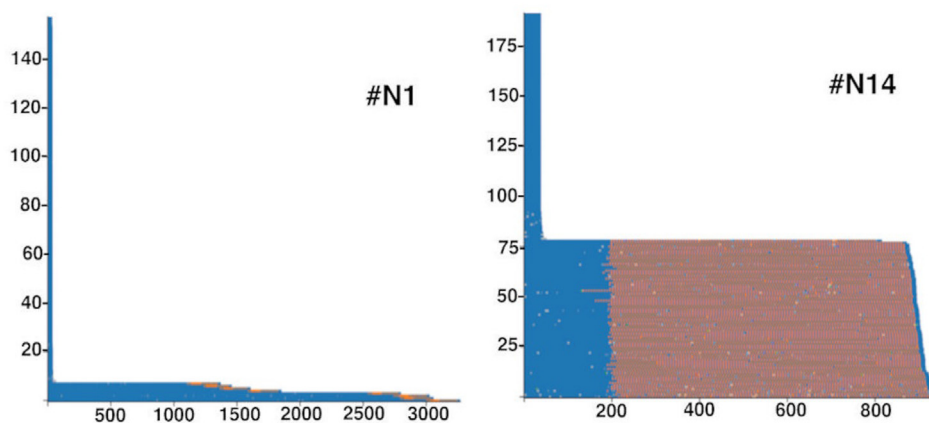
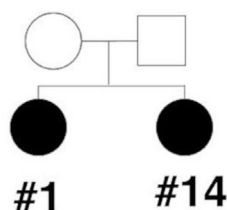
762 In the literature, the interrupted expanded allele has recently
 763 been shown to be a key factor in the severity of DM1
 764 symptoms and the age of onset of these symptoms.^{1,13–22}
 765 The identification of interruptions in patients with DM1 at
 766 diagnosis is therefore important to improve genetic coun-
 767 seling in DM1. Using the No-Amp method, we analyzed
 768 repeat expansion sequences to identify interrupted expanded
 769 alleles in patients with DM1. Samples #11 and #18 showed
 770 interruptions at the 3' end of the CTG repeat expansion,
 771 which were also detected by TP-PCR (Figure 3) (data not
 772 shown). Two CCG interruptions and four CCG interruptions
 773 were identified in samples #11 and #18, respectively, by
 774



775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868

Figure 3 Waterfall plots outline the repeat structure of the normal and expanded alleles. Triplet repeat sequences are sorted from shortest to longest, and each individual molecule is represented by a series of **colored dots** on a **horizontal line**. Each **dot** represents a single repeat unit. The CTG repeat is represented in blue, whereas the CCG repeat is represented in orange. The highest peaks represent the normal allele. The x axis shows the size of the repeat in base pair (bp).

Family 1



Family 2

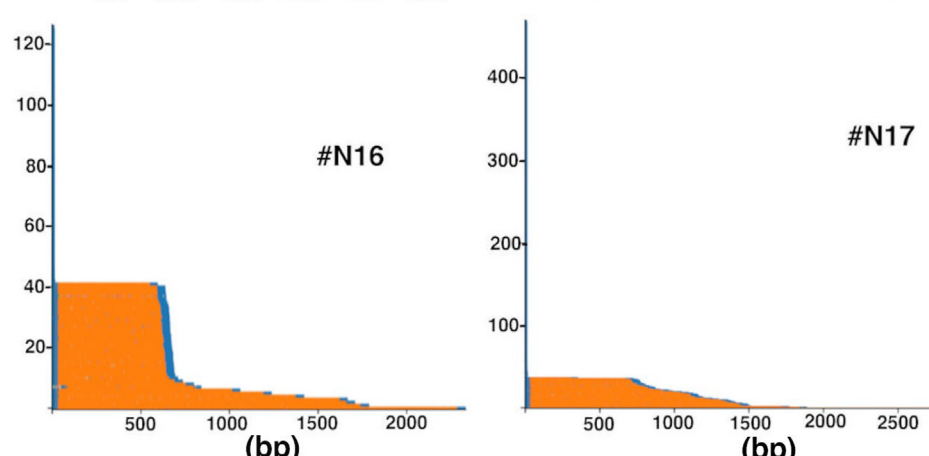
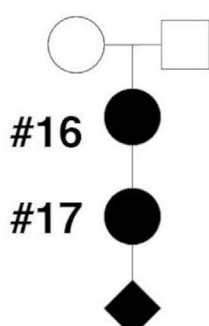


Figure 4 Waterfall plots outline the repeat structure of the normal and expanded alleles in myotonic dystrophy type 1 families 1 and 2 (PCR failed). Triplet repeat sequences are sorted from shortest to longest, and each individual molecule is represented by a series of colored dots on a horizontal line. Each dot represents a single repeat unit. The CTG repeat is represented in blue, whereas the CCG repeat is represented in orange. The highest peaks represent the normal allele. The x axis shows the size of the repeat in base pair (bp). Part of the pedigrees of families 1 and 2 are shown on the left of the figure according to standardized human pedigree nomenclature based on the methods of Bennet et al.³²

targeted sequencing of the *DMPK* locus, identification that was impossible with other gold standard methods (Figure 3). Interestingly, a single CAG repeat interruption at the 5' end of the CTG repeat expansion was also identified in the blood-derived lymphoblastoid cell line of sample #A3 (Figure 3). In sample #15841, amplification of the expanded allele failed with PCR, making it impossible to analyze the CNG repeat expansion sequence. By targeted No-Amp sequencing, we overcame this limitation and identified approximately 78 CCG interruptions in this patient's mutated allele located both at the 3' end of the CTG repeat expansion and in the middle of the sequence (Figure 3).

Identification of DM1 Families Carrying CCG-Enriched Expanded Allele

DM1 patients in families 1 and 2 were identified at the neuro-myology department (Institute of Myology, Paris, France) as patients with atypical DM1 in whom mild or no muscle symptoms were reported. In addition, no cardiac or

severe respiratory abnormalities were observed (Figure 4^[F4] and Table 5).³² In particular, individual #14 is completely asymptomatic even after 12 years of follow-up.^[T5] She has neither motor weakness nor myotonia, even on electromyography examination. Respiratory and cardiac investigations remain normal. These clinical aspects differ from what is usually observed in patients with DM1 with similar age and CTG repeat size (ranging from 360 to 1000 by Southern blot).⁴ To better characterize the *DMPK* mutation in DM1 members of these two families, targeted sequencing without amplification was performed in individuals #1, #14, #16, and #17 in whom conventional PCR had failed. The amplification-free *DMPK* locus enrichment is shown in Table 6; only eight expanded allele HiFi reads^[T6] were obtained in sample #1, making it difficult to estimate repeat size and somatic mosaicism in this patient (Table 6). The estimated size of the repeats in samples #14, #16, and #17 was 299, 224, and 360 CNG repeats, respectively, which is lower than the estimates according to Southern blot (360, 430, and 1000 CNG repeats) (Table 7).^[T7]

Table 5 Clinical Features in DM1 Families 1 and 2

	Age sample (years)	Age at onset (years)	CTG repeat number (Southern blot)	Weakness	Myotonia	Cardiac defect	Respiratory defect
#1	45	36	NA	Axial, proximal, distal, and facial	Yes (36 years)	No	Vital capacity 91% and hypercapnia
#14	35	Asymptomatic	360	No	No	No	No
#16	48	38	430	Very mild facial	No	No	No
#17	21	20	1000	Distal weakness of the hand (Hirayama disease)	~ 15 years	No	No

DM1, myotonic dystrophy type 1.

These results may be explained by an insufficient number of HiFi reads obtained for each sample, particularly for samples #16 and #17. Strikingly, large expanded alleles enriched in CCG triplet repeats were identified, which explains the difficulties in amplifying the CNG repeat expansions in these patients (Figure 4). The mutated alleles of individual #14 [(CTG)61-71(CTGCCG)115(CTG)1-5] are composed of >35% CCG repeat interruptions. The expanded allele structure observed in patient #14 differs from the expanded allele structure observed in her sister (individual #1). The presence of >35% hexamer (CTGCCG) appears to stabilize the repeat in the blood of patient #14 (aged 35 years) (Figure 5). In family 2, patients #16 and #17 carry an expanded allele with >85% CCG interruptions. The major allele structure observed in patient #16 is (CTG)7(CCG) (CTG)3(CCG)197(CTG)16 and differs from that observed in her daughter (patient #17), in whom the major allele structure is (CTG)7(CCG) (CTG) 3(CCG)330(CTG)19. No-Amp sequencing results revealed a maternal CNG repeat expansion between individuals #16 and #17, suggesting that the DM1 expanded allele with block CCG interruptions is unstable across generations in this family. Furthermore, somatic mosaicism was observed in these two patients with two different distributions of CTG repeat number (Figure 5). The low number of HiFi reads did not allow an accurate estimate of somatic mosaicism in these patients (Table 6). However, for the first time, a new variant at the *DMPK* locus composed of >85% CCG repeats was identified by using the No-Amp targeted sequencing method.

Discussion

In DM1, current routine molecular diagnosis approaches do not provide simultaneous characterization of the size and sequence of the DM1 mutation with precision as well as the dynamics of repeat instability in patients with DM1 (Table 1). These parameters are important DM1 disease modifiers.³ In some cases, incomplete genetic analyses may lead to an inaccurate prognosis in patients with DM1. New genetic tests for DM1 are becoming essential to solidify the diagnosis, aid prognosis, and improve care with shorter time

frames. Recently, we have shown that Pacific Biosciences long-read sequencing from amplicons can sequence large CTG repeat expansions in patients with DM1 and detect CTG repeat interruptions.²⁷ Nevertheless, this approach remains limited by PCR amplification. Here, we showed that the Pacific Biosciences No-Amp targeted sequencing method is effective to sequence through previously inaccessible *DMPK* locus (large expanded alleles and GC-enriched interrupted alleles). In the DM1 population, >30% of patients carry large or interrupted expanded alleles.^{4,10} This number is certainly underestimated by the fact that the size and sequence of the CTG repeat expansion are not systematically determined during diagnosis. Amplification-free targeted sequencing allowed highly accurate sequence determination of hundreds of expanded DNA molecules in several DM1 samples. We have observed an up to 50% DM1 on-target rate in HiFi sequencing reads when targeting only the *DMPK* locus with the amplification-free targeted sequencing method. The on-target rate is affected by the quality of genomic DNA in which high-molecular weight DNAs provided the highest No-Amp sequencing yield in this study (Table 3). Recently, it has also been shown that high-molecular weight DNA, yielding DNA ranging from 50 to 300+ Kb in length, generates the highest quality long-read sequencing data of complex metagenomes.³³ The amplification-free targeted sequencing method is also sensitive to the origin of DNA samples. Variable results with DNA samples extracted from different tissues or cell types have been observed (data not shown). To continue improving the methodology in the future, additional studies are needed to understand how

Table 6 HiFi Reads in DM1 Patients of Families 1 and 2

Individuals	#1	#14	#16	#17
HiFi reads	926	652	510	1916
<i>DMPK</i> locus	158	192	127	471
Control loci	393	230	220	1275
Expanded allele	8	79	42	38
Non-expanded allele	150	113	85	433
% Reads on-target	59.61	64.88	64.12	91.13
% Reads off-target	40.39	35.12	35.88	8.87

DM1, myotonic dystrophy type 1; HiFi reads, highly accurate long-reads.

Table 7 CTG Repeat Size and Interruptions in DM1 Patients of Families 1 and 2

Individuals	#1	#14	#16	#17
Median expanded allele (CNG)	NA	299	224	360
Interruptions	52 CCG	115 CCG	198 CCG	331 CCG
% CCG repeats	NA	39%	88%	92%
CNG repeat size (diagnosis)	NA	360	430	1000

DM1, myotonic dystrophy type 1.

different DNA sources and extraction methods are affecting this targeted sequencing application.

During library preparation without amplification, a distribution close to 50% to 50% of normal and mutated sequencing reads was observed in most sequencing data. However, an overrepresentation of the normal allele in patients with the largest expansion was observed. This phenomenon can be explained by the lower HiFi read conversion rate because of the sequencing read length limitation and lower immobilization efficiency of the DNA polymerase complex on the Sequel Sequencing Chip. Long DNA molecules also have higher probability to carry damages susceptible to exonuclease degradation during sequencing library preparation. The combination of these factors could lead to none 50% to 50% distribution of sequencing reads on normal and expanded alleles. Although including circular consensus reads with predicted read quality below QV20 (accuracy <99%) in data analysis could recover additional reads with long repeat expansion, the option was not explored in this study to strengthen the identification of CNG interruptions. Improved long template immobilization efficiency and better sequencing read length on the newer SMRT sequencing platform could help overcome these limitations and provide more even coverages between normal and expanded alleles.

Removal of the PCR amplification step improves the accuracy of the measurement of the inherited CTG repeat expansion as well as the measurement of somatic mosaicism observed in blood or immortalized cells in most patients with DM1. Using no-amplification targeted sequencing, we were also able to accurately detect interruptions and determine their number in blood and/or immortalized cells of all patients with DM1, regardless of the number of reads.

For the first time, DM1 patients with an expansion composed of >85% CCG associated with mild or no muscle symptoms and the absence of cardiac and severe respiratory abnormalities were identified by using this method. This finding is consistent with recent studies showing that interrupted expanded alleles are associated with a decrease in the severity of DM1 symptoms.^{10,11,17,19–21} It also raises question about systematic analysis of interruptions in patients with DM1. Altogether, our results confirm the need to revisit molecular diagnosis with No-Amp long-read sequencing to improve genotype–phenotype correlations and thus genetic counseling and prognosis in DM1. Interestingly, heterogeneity in the sequence of the expansion,

characterized by different expansion lengths and varying number of repeat interruptions, was detected in families 1 and 2. The heterogeneity in the number and type of interruptions observed in expanded interrupted alleles across generations (family 2) and in the sibling blood of family 1 suggests instability of the interruptions despite the fact that the same interruptions can stabilize the pure CTG repeat expansion.^{13–18,22} The emergence of new interruptions in the expansion may result from multiple processes, including DNA polymerase errors during replication, error-prone DNA repair, or recombination. This method opens new avenues to understand the role of interruptions in DM1. The different mechanisms involved in the formation of CNG interruptions and instability of CTG repeats at the *DMPK* locus remain unknown and need to be identified in patients carrying the interrupted expanded allele using new models. Several therapeutic strategies using gene editing techniques or molecular approaches targeting the CTG repeat expansion have been developed to induce a decrease in the repeat and thus slow down or reverse DM1 disease.^{34,35} The frequency of correction and adverse effects of gene editing at the DM1 locus (DNA rearrangements, large indels, and substitutions) are difficult to measure by current methods. Long-read SMRT sequencing would allow a better estimation of the frequency of correction but also a better understanding of the mechanisms induced by gene editing at the DM1 locus.

In summary, this proof-of-principle study showed that no-amplification targeted long-read sequencing developed by Pacific Biosciences allows simultaneous attainment of high-resolution information on the number of repeats, a complete and accurate sequence, and a measure of somatic mosaicism even for long repeats. These three parameters involved in the DM1 clinic can be analyzed in the same assay in <1 month from DNA extraction to bioinformatics analysis. Currently, no-amplification long-read sequencing remains expensive with 2 days of library preparation (5 patients). A

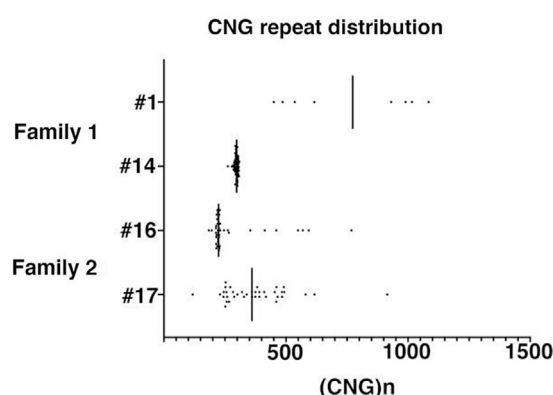


Figure 5 Distribution of the size of the repeat in families 1 and 2. The x axis shows the CTG repeat number. Each dot represents the CTG repeat size of one highly accurate long-read. The black line represents the median. Eight to 79 single-molecule highly accurate long-reads of expanded allele were analyzed in the no-amplification experiment.

single-molecule sequencing instrument should be readily available in next-generation sequencing platforms. However, the multiplexing of patients on a single SMRT Cell, improving the technology by Pacific Biosciences, and reducing its cost will allow SMRT sequencing to be implemented as a routine molecular diagnostic method offering the best diagnosis and prognosis for patients with DM1. This approach is highly relevant in DM1, which is one of the most complex trinucleotide disorders associated with the largest expansion and with high genetic and clinical variability. No-amplification targeted sequencing gives us the opportunity to better understand the dynamics of CTG repeat instability and genotype–phenotype association in DM1. This method will be extremely useful for validating innovative therapeutic strategies aimed at decreasing the CTG.CAG repeat length and thus stopping disease progression in patients with DM1.

Acknowledgments

We thank the DM1 patients and colleagues at the Myology Center for Research, Curie Institute, and Pacific Biosciences for helpful discussions and comments; and specially thank Sonia Lameiras, Sylvain Baulande, Tina Alaeitabar, Lori Aro, Deborah Moine, and David Stucki for their active involvement in this project.

Author Contribution

Conceptualization, Y.-C.T., C.H., T.S., D.F., G.B., G.G., and S.T. Formal analysis, Y.-C.T. and S.T. Funding acquisition, G.G. and S.T. Investigation, Y.-C.T., L.d.P., T.S., G.B., and S.T. Methodology, Y.-C.T., C.H., and S.T. Project administration, S.T. Resources, Y.-C.T., C.H., T.S., D.F., G.B., G.G., and S.T. Supervision, S.T. Validation, Y.-C.T. and S.T. Visualization, S.T. Writing—original draft, S.T. Writing—review and editing, Y.-C.T., L.d.P., C.H., T.S., D.F., G.B., G.G., and S.T.

Supplemental Data

Supplemental material for this article can be found at <http://doi.org/10.1016/j.jmoldx.2022.08.003>.

References

1. Khristich AN, Mirkin SM: On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J Biol Chem* 2020, 295:4134–4170
2. Harper PS: *Myotonic Dystrophy*. London, UK, W.B. Saunders, 2001
3. Tomé S, Gourdon G: DM1 phenotype variability and triplet repeat instability: challenges in the development of new therapies. *Int J Mol Sci* 2020, 21:457
4. De Antonio M, Dogan C, Hamroun D, Mati M, Zerrouki S, Eymard B, Katsahian S, Bassez G; French Myotonic Dystrophy Clinical Network: Unravelling the myotonic dystrophy type 1 clinical

- spectrum: a systematic registry-based study with implications for disease classification. *Revue Neurol (Paris)* 2016, 172:572–580
5. Harper PS, Harley HG, Reardon W, Shaw DJ: Anticipation in myotonic dystrophy: new light on an old problem. *Am J Hum Genet* 1992, 51:10–16
6. Kamsteeg E-J, Kress W, Catalli C, Hertz JM, Witsch-Baumgartner M, Buckley MF, van Engelen BGM, Schwartz M, Scheffer H: Best practice guidelines and recommendations on the molecular diagnosis of myotonic dystrophy types 1 and 2. *Eur J Hum Genet* 2012, 20:1203–1208
7. Savić Pavićević D, Miladinović J, Brkušanin M, Šviković S, Djurica S, Brajušković G, Romac S: Molecular genetics and genetic testing in myotonic dystrophy type 1. *Biomed Res Int* 2013, 2013:1391821
8. Morales F, Couto JM, Higham CF, Hogg G, Cuenca P, Braida C, Wilson RH, Adam B, del Valle G, Brian R, Sittenfeld M, Ashizawa T, Wilcox A, Wilcox DE, Monckton DG: Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum Mol Genet* 2012, 21:3558–3567
9. Morales F, Vásquez M, Corrales E, Vindas-Smith R, Santamaría-Ulloa C, Zhang B, Siritto M, Estecio MR, Krahe R, Monckton DG: Longitudinal increases in somatic mosaicism of the expanded CTG repeat in myotonic dystrophy type 1 are associated with variation in age-at-onset. *Hum Mol Genet* 2020, 29:2496–2507
10. Cumming SA, Jimenez-Moreno C, Okkersen K, Wenninger S, Daidj F, Hogarth F, Littleford R, Gorman G, Bassez G, Schoser B, Lochmüller H, van Engelen BGM, Monckton DG; OPTIMISTIC Consortium: Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort. *Neurology* 2019, 93:e995–e1009
11. Overend G, Légaré C, Mathieu J, Bouchard L, Gagnon C, Monckton DG: Allele length of the DMPK CTG repeat is a predictor of progressive myotonic dystrophy type 1 phenotypes. *Hum Mol Genet* 2019, 28:2245–2254
12. Peric S, Pesovic J, Savić-Pavicevic D, Rakocevic Stojanovic V, Meola G: Molecular and clinical implications of variant repeats in myotonic dystrophy type 1. *Int J Mol Sci* 2021, 23:354
13. Braida C, Stefanatos RKA, Adam B, Mahajan N, Smeets HJM, Niel F, Goizet C, Arveiler B, Koenig M, Lagier-Tourenne C, Mandel J-L, Faber CG, de Die-Smulders CEM, Spaans F, Monckton DG: Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Hum Mol Genet* 2010, 19:1399–1412
14. Musova Z, Mazanec R, Krepelova A, Ehler E, Vales J, Jaklova R, Prochazka T, Koukal P, Marikova T, Kraus J, Havlovicova M, Sedlacek Z: Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet A* 2009, 149A:1365–1374
15. Botta A, Rossi G, Marcaurelio M, Fontana L, D'Apice MR, Brancati F, Massa R, Monckton DG, Sanguiuolo F, Novelli G: Identification and characterization of 5' CCG interruptions in complex DMPK expanded alleles. *Eur J Hum Genet* 2017, 25:257–261
16. Cumming SA, Hamilton MJ, Robb Y, Gregory H, McWilliam C, Cooper A, Adam B, McGhie J, Hamilton G, Herzyk P, Tschannen MR, Worthey E, Petty R, Ballantyne B; Scottish Myotonic Dystrophy Consortium, Warner J, Farrugia ME, Longman C, Monckton DG: De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur J Hum Genet* 2018, 26:1635–1647
17. Santoro M, Masciullo M, Silvestri G, Novelli G, Botta A: Myotonic dystrophy type 1: role of CCG, CTC and CGG interruptions within DMPK alleles in the pathogenesis and molecular diagnosis. *Clin Genet* 2017, 92:355–364
18. Tomé S, Dandelot E, Dogan C, Bertrand A, Geneviève D, Péréon Y, DM Contraction Study Group, Simon M, Bonnefont J-P, Bassez G,

- 1365 Gourdon G: Unusual association of a unique CAG interruption in 5'
1366 of DM1 CTG repeats with intergenerational contractions and low
1367 somatic mosaicism. *Hum Mutat* 2018, 39:970–982
- 1368 19. Wenninger S, Cumming SA, Gutschmidt K, Okkersen K, Jimenez-
1369 Moreno AC, Daidj F, Lochmüller H, Hogarth F, Knoop H, Bassez G,
1370 Monckton DG, van Engelen BGM, Schoser B: Associations between
1371 variant repeat interruptions and clinical outcomes in myotonic dys-
1372 trophy type 1. *Neurol Genet* 2021, 7:e572
- 1373 20. Miller JN, van der Plas E, Hamilton M, Kosciak TR, Gutmann L,
1374 Cumming SA, Monckton DG, Nopoulos PC: Variant repeats within
1375 the *DMPK* CTG expansion protect function in myotonic dystrophy
1376 type 1. *Neurol Genet* 2020, 6:e504
- 1377 21. Ballester-Lopez A, Koehorst E, Almendrote M, Martínez-Piñeiro A,
1378 Lucente G, Linares-Pardo I, Núñez-Manchón J, Guanyabens N,
1379 Cano A, Lucia A, Overend G, Cumming SA, Monckton DG,
1380 Casadevall T, Isern I, Sánchez-Ojanguren J, Planas A, Rodríguez-
1381 Palmero A, Monlleó-Neila L, Pintos-Morell G, Ramos-Fransi A,
1382 Coll-Cantí J, Nogales-Gadea G: A DM1 family with interruptions
1383 associated with atypical symptoms and late onset but not with a
1384 milder phenotype. *Hum Mutat* 2020, 41:420–431
- 1385 22. Pešović J, Perić S, Brkušanić M, Brajušković G, Rakočević-
1386 Stojanović V, Savić-Pavičević D: Repeat interruptions modify age at
1387 onset in myotonic dystrophy type 1 by stabilizing *DMPK* expansions
1388 in somatic cells. *Front Genet* 2018, 9:601
- 1389 23. Morales F, Vásquez M, Santamaría C, Cuenca P, Corrales E,
1390 Monckton DG: A polymorphism in the *MSH3* mismatch repair gene
1391 is associated with the levels of somatic instability of the expanded
1392 CTG repeat in the blood DNA of myotonic dystrophy type 1 patients.
1393 *DNA Repair (Amst)* 2016, 40:57–66
- 1394 24. Flower M, Lomeikaite V, Ciosi M, Cumming S, Morales F, Lo K,
1395 Hensman Moss D, Jones L, Holmans P, TRACK-HD Investigators,
1396 OPTIMISTIC Consortium, Monckton DG, Tabrizi SJ: *MSH3* mod-
1397 ifies somatic instability and disease severity in Huntington's and
1398 myotonic dystrophy type 1. *Brain* 2019, 142:1876–1886
- 1399 25. Warner JP, Barron LH, Goudie D, Kelly K, Dow D,
1400 Fitzpatrick DR, Brock DJ: A general method for the detection of
1401 large CAG repeat expansions by fluorescent PCR. *J Med Genet*
1402 1996, 33:1022–1026
- 1403 26. Monckton DG, Wong L-J, Ashizawa T, Caskey CT: Somatic mosa-
1404 icism, germline expansions, germline reversions and intergenerational
1405 reductions in myotonic dystrophy males: small pool PCR analyses.
1406 *Hum Mol Genet* 1995, 4:1–8
- 1407 27. Mangin A, de Pontual L, Tsai Y-C, Monteil L, Nizon M, Boisseau P,
1408 Mercier S, Ziegler J, Harting J, Heiner C, Gourdon G, Tomé S: Robust
1409 detection of somatic mosaicism and repeat interruptions by long-read
1410 targeted sequencing in myotonic dystrophy type 1. *Int J Mol Sci*
1411 2021, 22:2616
- 1412 28. Tsai Y-C, Zafar F, McEachin ZT, McLaughlin I, Van Blitterswijk M,
1413 Ziegler J, Schüle B. Edited by Proukakis C: Multiplex CRISPR/Cas9-
1414 Guided No-Amp targeted sequencing panel for spinocerebellar Ataxia
1415 repeat expansions. *Genomic Structural Variants in Nervous System*
1416 *Disorders. Neuromethods*, vol. 182. New York, NY: Humana, 2022.
1417 pp. 95–120
- 1418 29. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M,
1419 Lander ES, Getz G, Mesirov JP: Integrative genomics viewer. *Nat*
1420 *Biotechnol* 2011, 29:24–26
- 1421 30. Gomes-Pereira M, Bidichandani SI, Monckton DG: Analysis of un-
1422 stable triplet repeats using small-pool polymerase chain reaction.
1423 *Methods Mol Biol* 2004, 277:61–76
- 1424 31. Tomé S, Nicole A, Gomes-Pereira M, Gourdon G: Non-radioactive
1425 detection of trinucleotide repeat size variability. *PLoS Curr* 2014, 6.
1426 ecurrents.md.ad50113b899fa1352ce70c087eead706
- 1427 32. Bennett RL, French KS, Resta RG, Doyle DL: Standardized human
1428 pedigree nomenclature: update and assessment of the recommenda-
1429 tions of the National Society of Genetic Counselors. *J Genet Couns*
1430 2008, 17:424–433
- 1431 33. Trigodet F, Lolans K, Fogarty E, Shaiber A, Morrison HG,
1432 Barreiro L, Jabri B, Eren AM: High molecular weight DNA extrac-
1433 tion strategies for long-read sequencing of complex metagenomes.
1434 *Mol Ecol Resour* 2022, 22:1786–1802
- 1435 34. Izzo M, Battistini J, Provenzano C, Martelli F, Cardinali B,
1436 Falcone G: Molecular therapies for myotonic dystrophy type 1: from
1437 small drugs to gene editing. *Int J Mol Sci* 2022, 23:4622
- 1438 35. Marsh S, Hanson B, Wood MJA, Varela MA, Roberts TC: Appli-
1439 cation of CRISPR-Cas9-mediated genome editing for the treatment of
1440 myotonic dystrophy type 1. *Mol Ther* 2020, 28:2527–2539

1489 Q15
1490

Supplemental Figure S1 Genome-wide coverage plots for patients with myotonic dystrophy type 1 (DM1) are shown in the graphic. The y axis shows the number of reads, and the x axis spans over all the chromosomes in the Hg38 genome.

1491
1492

UNCORRECTED PROOF