

# Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort

Sarah A. Cumming, PhD, Cecilia Jimenez-Moreno, PhD, Kees Okkersen, MD, Stephan Wenninger, MD, Ferroudja Daidj, MD, Fiona Hogarth, PhD, Roberta Littleford, PhD, Gráinne Gorman, MD, Guillaume Bassez, MD, Benedikt Schoser, MD, Hanns Lochmüller, MD, Baziel G.M. van Engelen, MD, PhD, and Darren G. Monckton, PhD, on behalf of the OPTIMISTIC Consortium

**Correspondence**  
Dr. Monckton  
darren.monckton@glasgow.ac.uk

Neurology® 2019;93:e995-e1009. doi:10.1212/WNL.00000000000008056

## Abstract

### Objective

To evaluate the role of genetic variation at the *DMPK* locus on symptomatic diversity in 250 adult, ambulant patients with myotonic dystrophy type 1 (DM1) recruited to the Observational Prolonged Trial in Myotonic Dystrophy Type 1 to Improve Quality of Life—Standards, a Target Identification Collaboration (OPTIMISTIC) clinical trial.

### Methods

We used small pool PCR to correct age at sampling biases and estimate the progenitor allele CTG repeat length and somatic mutational dynamics, and *Acil* digests and repeat primed PCR to test for the presence of variant repeats.

### Results

We confirmed disease severity is driven by progenitor allele length, is further modified by age, and, in some cases, sex, and that patients in whom the CTG repeat expands more rapidly in the soma develop symptoms earlier than predicted. We revealed a key role for variant repeats in reducing disease severity and quantified their role in delaying age at onset by approximately 13.2 years (95% confidence interval 5.7–20.7, 2-tailed *t* test  $t = -3.7$ ,  $p = 0.0019$ ).

### Conclusions

Careful characterization of the *DMPK* CTG repeat to define progenitor allele length and presence of variant repeats has increased utility in understanding clinical variability in a trial cohort and provides a genetic route for defining disease-specific outcome measures, and the basis of treatment response and stratification in DM1 trials.

---

From the Institute of Molecular, Cell and Systems Biology (S.A.C., D.G.M.), University of Glasgow; Institute of Genetic Medicine (C.J.-M., H.L.) and Institute of Neurosciences (G.G.), Newcastle University, Newcastle upon Tyne, UK; Department of Neurology (K.O., B.G.M.v.E.), Donders Institute for Brain Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, the Netherlands; Department of Neurology (S.W., B.S.), Friedrich-Baur-Institute, Ludwig-Maximilians-Universität München, Munich, Germany; Neuromuscular Reference Centre (F.D., G.B.), Assistance Publique-Hôpitaux de Paris, France; and Tayside Clinical Trials Unit (F.H., R.L.), The University of Dundee, UK.

Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

OPTIMISTIC Consortium coinvestigators are listed at [links.lww.com/WNL/A962](http://links.lww.com/WNL/A962).

The Article Processing Charge was funded by Wellcome Trust.

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Glossary

**6-MWT** = 6-Minute Walk Test; **ASBQ** = Adult Social Behavior Questionnaire; **BDI-fs** = Beck Depression Inventory—fast screen; **CBT** = cognitive behavioral therapy; **CI** = confidence interval; **CIS-fatigue** = Checklist Individual Strength Fatigue score; **DM1** = myotonic dystrophy type 1; **ePAL** = estimated progenitor allele length; **FDSS** = Fatigue and Daytime Sleepiness Scale; **MIRS** = Muscle Impairment Rating Scale; **OPTIMISTIC** = Observational Prolonged Trial in Myotonic Dystrophy Type 1 to Improve Quality of Life—Standards, a Target Identification Collaboration; **RP-PCR** = repeat-primed PCR; **SP-PCR** = small pool PCR; **SSL** = van Sonderen Social Support List; **SSL-D** = van Sonderen Social Support List—Discrepancies; **SSL-I** = van Sonderen Social Support List—Interactions; **SSL-N** = van Sonderen Social Support List—Negative Interactions; **TMT** = Trail-Making Test.

Myotonic dystrophy type 1 (DM1) is a highly variable autosomal dominant inherited disorder affecting individuals of both sexes and all ages.<sup>1</sup> DM1 is caused by the expansion of a CTG repeat in the *DMPK* gene that varies in length from under 40 repeats in the general population to more than 1,000 repeats in some patients.<sup>2–7</sup> Longer CTG tracts are associated with more severe disease and earlier age at onset.<sup>8–11</sup> However, genotype to phenotype relationships determined using the traditional diagnostic test have low predictive value and the International Myotonic Dystrophy Consortium recommends that families are not offered predictive phenotypic information based on the size of the CTG repeat.<sup>12</sup>

Expanded disease-causing alleles are highly unstable and usually increase in length during intergenerational transmission,<sup>8,13–15</sup> explaining the decrease in age at onset of 20–30 years per generation typically observed in DM1 families.<sup>16</sup> The expanded CTG is also highly unstable in the soma in a process that is expansion-biased, age-dependent, and tissue-specific.<sup>17–24</sup> Thus, the average length of the CTG repeat measured using the traditional diagnostic depends on the age at which the sample was taken. Previously, we demonstrated that a sensitive small pool PCR (SP-PCR) analysis of blood DNA could be used to estimate the progenitor allele length (ePAL), i.e., the length of the CTG repeat transmitted to the individual by the affected parent.<sup>18</sup> In addition, we demonstrated that ePAL could reduce the confounding effects of age at sampling inherent in the traditional test and we could dramatically improve the ability to predict age at onset in DM1.<sup>25,26</sup>

Observational Prolonged Trial in Myotonic Dystrophy Type 1 to Improve Quality of Life—Standards, a Target Identification Collaboration (OPTIMISTIC) was a large international, multicenter, model-based randomized trial designed to compare the utility of tailored cognitive behavioral therapy (CBT) with an optional graded exercise plan against standard care.<sup>27,28</sup> In order to provide additional insight into the genetic basis of variability in DM1 and as a route toward genetic stratification in clinical trials, we sought to estimate the progenitor allele length and define the repeat dynamics of the DM1 CTG repeat in the blood DNA of participants recruited to OPTIMISTIC.

## Methods

### Standard protocol approvals, registrations, and patient consents

Full details of the OPTIMISTIC protocol, including power calculations for sample size, have been reported previously (Clinicaltrials.gov, NCT02118779).<sup>27</sup> All participants were recruited with informed consent. The protocol was approved by the National Research Ethics Service Committee North East—Sunderland (UK), the Comité de Protection des Personnes Ile de France V (France), the Ethikkommission bei der LMU München (Germany), and the Concernstaf Kwaliteit en Veiligheid Commissie Mensgebonden Onderzoek Regio Arnhem-Nijmegen (Netherlands).

### Phenotypic data

Ambulant adult patients (>18 years of age) with genetically proven DM1 capable of providing informed consent with severe fatigue (Checklist Individual Strength–Fatigue [CIS-Fatigue] score  $\geq 35$ ) were recruited between April 2014 and May 2015 through 4 clinical centers: Nijmegen, the Netherlands; Paris, France; Munich, Germany; and Newcastle, United Kingdom (n = 255 [46.3% female] [118/255]).<sup>27,28</sup> Exclusion criteria included neurologic or orthopedic comorbidity possibly influencing the intervention or outcomes; use of psychotropic drugs (except modafinil, Ritalin, and antidepressants, where the dosing regimen was stable for at least 12 months prior to screening); severe depression; participation in another interventional clinical trial; or inability to complete study questionnaires. Age at onset of DM1 symptoms was obtained during the trial recruitment screening visit in answer to the question “At what age did the first medical problems occur that may be related to your myotonic dystrophy?” With the exception of patients recruited through Munich, age at sampling was calculated based on known date of birth and the date of visit 2, at which the blood sample was provided. As a local regulatory requirement, date of birth was not recorded for participants recruited in Munich, therefore the age at sampling used was self-reported by participants at visit 2. Additional baseline (visit 2) clinical measures were obtained as described in the OPTIMISTIC protocol and reported previously.<sup>27,28</sup> The majority of participants (n = 225) were not known to be related. Eight parent–offspring pairs, 4 sibling pairs, and 3 trios of 2 siblings and 1 offspring were also recruited.

## DNA samples

A 10-mL venous blood sample was taken from OPTIMISTIC participants at baseline (time point V2). Blood was collected into EDTA tubes and stored locally at  $-80^{\circ}\text{C}$ . Frozen blood was shipped on dry ice to the Newcastle Biobank for Research of Neuromuscular Disorders,<sup>29</sup> where DNA was isolated and shipped to the University of Glasgow. Baseline DNA samples were available from 250 out of 255 OPTIMISTIC participants. Historical DNA samples from OPTIMISTIC participants collected during routine molecular diagnosis (from 6 months to 21 years previously) were available for 71 participants and obtained from local diagnostic laboratories via the Newcastle Biobank. Historical DNA samples were not available for the 5 participants without baseline DNA samples.

## SP-PCR and Acil digestion

SP-PCR was carried out using flanking primers DM-C and DM-DR as previously described<sup>18,30</sup> using Custom PCR Master Mix (Thermo Fisher Scientific [Waltham, MA] #SM-0005) supplemented with 69 mM 2-mercaptoethanol. Taq polymerase (Sigma-Aldrich UK, Gillingham) was used at 1 unit per 10  $\mu\text{L}$ . Where required, reactions were supplemented with 10% DMSO and the annealing temperature reduced to  $63.5^{\circ}\text{C}$ . PCR products were digested with Acil (New England Biolabs UK, Hitchin) in accordance with the manufacturer's instructions. DMSO was removed prior to Acil digestion using the QIAquick (Qiagen, Venlo, the Netherlands) PCR purification kit. DNA fragments were resolved by agarose gel electrophoresis and Southern blot hybridized as described.<sup>18,30</sup> Autoradiographic images were scanned and ePAL and modal allele lengths were estimated from the lower boundary<sup>18</sup> and the densest part of the expanded allele distribution respectively by comparison against the molecular weight ladder, using CLIQS 1D gel analysis software (TotalLab UK, Newcastle upon Tyne).

## Repeat-primed PCR (RP-PCR)

RP-PCR was performed at the 5'-end of the repeat tract using 1  $\mu\text{M}$  DM-A, 0.1  $\mu\text{M}$  TAG-CAG<sub>5</sub>, and 1  $\mu\text{M}$  TAG primer and at the 3'-end of the repeat tract using 1  $\mu\text{M}$  DM-DR, 0.1  $\mu\text{M}$  TAG-CTG<sub>5</sub>, and 1  $\mu\text{M}$  TAG primer, as described previously.<sup>31</sup>

## Statistical analysis

Statistical analyses were undertaken in R (version 3.4.3)<sup>32</sup> using RStudio (version 1.0.153).<sup>33</sup> There were no repeated measures and all individuals are represented once in each statistical analysis. All statistical analyses were 2-sided. Log transformations of the age at onset, degree of somatic instability, and Stroop interference score were used to better approximate a normal distribution and minimize the influence of extreme values. Participants with missing data were excluded from relevant analyses. Multiple linear regressions were performed using the `lm` function from the R Stats package. In all regression analyses, we report the adjusted  $r^2$ , the  $p$  value for the overall model, and the coefficient ( $\beta$ ) and  $p$  value for each measure. Model selection was based on the

Akaike information criterion for each model using a backwards stepwise selection procedure implemented using the `stepAIC` function in R.<sup>34</sup> Confidence intervals (CI) (95%) for point estimates were calculated where appropriate.

## Data availability

The de-identified participant genetic data presented here are available online ([datadryad.org/review?doi=doi:10.5061/dryad.t063q70](https://datadryad.org/review?doi=doi:10.5061/dryad.t063q70)). Requests to access the de-identified participant data from the main OPTIMISTIC trial should be addressed to the trial chief investigator Dr. Baziel van Engelen ([baziel.vanengelen@radboudumc.nl](mailto:baziel.vanengelen@radboudumc.nl)).

## Results

### Estimation of progenitor and modal allele CTG repeat lengths in the OPTIMISTIC cohort

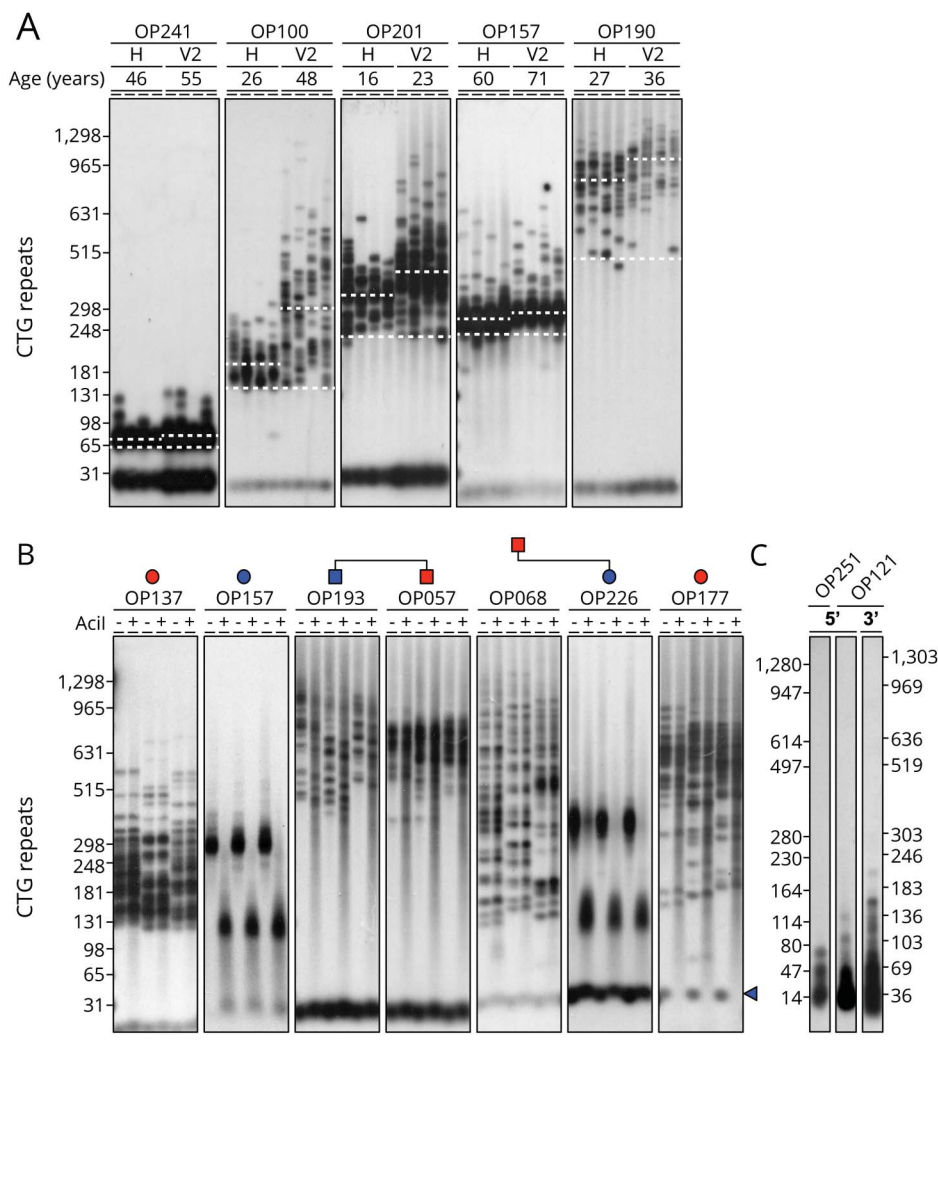
In order to investigate genetic variation in the OPTIMISTIC cohort, we used SP-PCR<sup>18</sup> to determine the ePAL, modal allele length at recruitment, and where available (71 participants), modal allele length at molecular diagnosis for the expanded CTG repeat *DMPK* allele (figure 1A). An expanded *DMPK* CTG repeat allele was readily detected in 241 out of 250 participants using our standard SP-PCR procedure. In an additional 7 participants, an expanded allele was only detected after the addition of 10% DMSO to the PCR, consistent with the presence of GC-rich variant repeats.<sup>31</sup> In one of these individuals, even in the presence of DMSO, only a small number of expanded alleles was successfully amplified relative to the non-disease-associated allele and it was not possible to define the ePAL or modal allele length. In 2 remaining individuals, we could not successfully amplify an expanded allele, although in both cases it was possible to amplify a single nonexpanded allele, consistent with the presence of a non-amplifiable expanded allele containing a large number of GC-rich variant repeats (see below for additional analyses). In the cases where we were able to obtain a historical DNA sample, the modal allele length had nearly always noticeably increased in size between the time of diagnosis and recruitment to the trial (figure 1A). As expected, it was easier to estimate ePAL from the diagnostic DNA sample obtained at an earlier age, confirming the value of these historical DNA samples.

### Identification of variant repeats within the CTG arrays of the OPTIMISTIC cohort

Although the majority of patients with DM1 likely inherit a pure CTG repeat expansion at the *DMPK* locus, it is known that a subset of patients contains variant repeat interruptions within the disease allele.<sup>31,35</sup> Such variant repeats, most commonly CCG and CGG, have been shown to reduce both germline and somatic instability<sup>31</sup> and appear to be associated with less severe symptoms.<sup>31,35-37</sup> Thus, we used SP-PCR to amplify the CTG repeat tract and test for the presence of variant repeats by post-PCR digestion with Acil (figure 1B). Acil recognizes the sequence CCGC and cleaves both CCG and CGG variant repeats.<sup>31</sup> Using this approach, we identified



**Figure 1** Estimation of progenitor and modal CTG repeat length and detection of Acil-sensitive variant repeats



(A) Shown are representative small pool PCR (SP-PCR) analyses of repeat length variation at the *DMPK* CTG repeat in blood DNA from 5 individuals. For each participant, the analysis of 2 DNA samples is shown: historical sample taken at time of initial diagnosis (H) and at baseline (V2). The age at sampling for each sample is indicated in years. The approximate positions of the molecular weight standards converted into the number of CTG triplet repeats are shown on the left. The position of the estimated progenitor allele length for each individual is indicated with the lower dashed white line. The modal allele length at each time point is indicated by the upper dashed white line. For each sample, 4 replicate PCRs were performed with ~180–300 pg DNA. (B) Shown are representative SP-PCR and Acil digestions of the *DMPK* CTG repeat in blood DNA from 7 individuals. For each participant, 3 replicate PCRs were performed with ~500 pg DNA, digested (+) or undigested (-) with Acil, resolved by agarose gel electrophoresis and Southern blot hybridized with a repeat unit probe. The approximate positions of the molecular weight standard converted into the number of CTG triplet repeats of undigested products is shown on the left. For each participant the presence (blue) or absence (red) of Acil-sensitive variant repeats within the expanded allele is indicated by filled pedigree symbols. Note OP177 contains Acil-sensitive variant repeats within the non-disease-associated allele as indicated by the blue triangle. (C) Detection of expanded alleles using repeat-primed PCR (RP-PCR). Shown are 5'- and 3'-RP-PCR assays for 2 participants (OP251 and OP121) in which we were not able to amplify the expanded allele using flanking primers. For both participants, a ladder of products extending beyond 50 repeats was observed using RP-PCR, confirming the presence of a disease-causing expansion. Notably, the ladders were discontinuous, consistent with the presence of variant repeat blocking amplification from the (CAG)<sub>5</sub> or (CTG)<sub>5</sub> repeat primer at some positions within the array. The approximate positions of the molecular weight standard converted into the number of triplet repeats are shown.

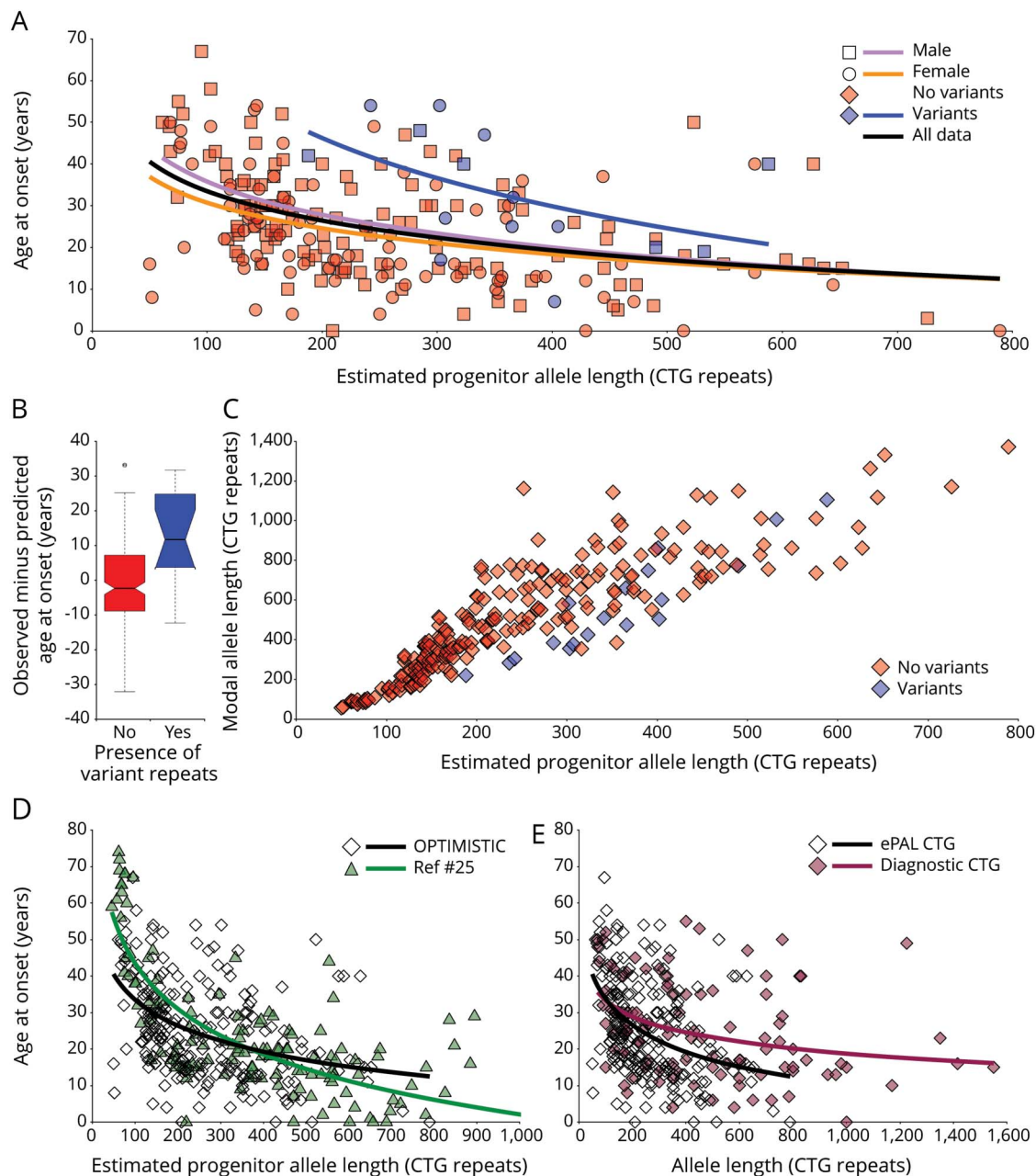
variant repeats in 19 out of 248 expanded alleles (7.7%). This included all the samples that required the addition of DMSO to facilitate PCR amplification. For the 2 samples for which we could not amplify the expanded allele, we used RP-PCR<sup>31,38,39</sup> to confirm the presence of an expanded allele (figure 1C). The presence of gaps within the repeat ladder of these 2 participants is consistent with the existence of variant repeats within the expanded allele. Thus, overall, the total frequency of participants identified with variant repeats in the expanded allele was 8.4% (21/250 [13/21 female], 95% CI 5.4%–12.7%). This included 1 pair of siblings, in both of whom we detected variant repeats. Interestingly, we also detected variant repeats in the daughter of a male transmitting parent in whom we did not detect variant repeats, consistent with de novo gain of variant repeats (figure 1B). Similarly, we

also detected variant repeats in only 1 of a pair of siblings, consistent with either a de novo gain or loss of variant repeats (figure 1B). Although the proportion of participants with variant repeats appeared to differ between centers (Munich 5/65 [7.7%, 95% CI 2.8%–17.8%], Newcastle 4/52 [7.7%, 95% CI 2.5%–19.4%], Nijmegen 9/66 [13.6%, 95% CI 6.8%–24.8%], Paris 3/67 [4.5%, 95% CI 1.2%–13.4%]), these differences were not statistically significant (Fisher exact test,  $p = 0.32$ ). In one participant, we detected variant repeats in a 34-triplet large non-disease-associated allele (figure 1B).

### Age at onset correlations in the OPTIMISTIC cohort

Self-reported age at onset of symptoms was available for 229 participants and we used linear regression modeling to

**Figure 2** Age at onset and somatic instability correlations



(A) Age at onset is highly correlated with estimated progenitor allele length (ePAL). The scatterplots show the relationship between ePAL and age at onset. The relevant line of best fit under a logarithmic model for female individuals (circles, orange line), male individuals (squares, lavender line), and sex-averaged for all individuals (black line, model 1a) (table 1) are shown. Individuals with (blue) and without (red) Acil-sensitive variant repeats are also depicted, along with the logarithmic regression line for Acil-sensitive variant repeat carriers (blue line). (B) Effect of variant repeats on age at onset. Shown are boxplots for the difference in observed age at onset minus predicted age at onset for Observational Prolonged Trial in Myotonic Dystrophy Type 1 to Improve Quality of Life—Standards, a Target Identification Collaboration (OPTIMISTIC) participants with (yes, blue) and without (no, red) Acil-sensitive variant repeats in their expanded *DMPK* allele. Predicted age at onset was derived using model 1a ( $Age_o = 80.3 + [-23.4 \times \log(ePAL)]$ ) (table 1). The mean difference in age at onset for participants carrying Acil-sensitive variant repeats was 13.2 years (95% confidence interval [CI] 5.7 to 20.7, 2-tailed  $t$  test  $t = -3.7$ ,  $p = 0.0019$ ). The bottom and top of the box are the lower and upper quartiles, respectively. The band near the middle of the box is the median and the notches approximate to the 95% CI for the medians. The whiskers represent the full range of observations bounded by an upper limit equal to the upper quartile plus  $1.5 \times$  the interquartile range, and a lower limit equal to the lower quartile minus  $1.5 \times$  the interquartile range. Any points outside these bounds are displayed individually as small circles. (C) Modal allele length, ePAL, and presence of Acil variant repeats. The scatterplots show the relationship between ePAL and modal allele length at the V2 time point. Individuals with (blue) and without (red) Acil-sensitive variant repeats are indicated. (D) Sampling bias in the OPTIMISTIC cohort. The scatterplots show the relationship between ePAL and age at onset for the OPTIMISTIC cohort (open diamonds, black line) and a family-based DM1 population characterized by Morales et al.<sup>25</sup> (green triangles and line). The relevant line of best fit under a logarithmic model (model 1a, table 1) is shown for each population (for the OPTIMISTIC cohort  $n = 222$ ,  $r^2 = 0.178$ ,  $p = 3.1 \times 10^{-11}$  and for Morales et al.,  $n = 137$ ,  $r^2 = 0.640$ ,  $p < 2.2 \times 10^{-16}$ ). Note that the Morales et al. cohort has more very mildly affected participants with small expansions, and more severely affected patients with large expansions. (E) ePAL is more informative than the diagnostic measure in predicting age at onset. The scatterplots show the relationship between age at onset and ePAL (ePAL, open diamonds and black line) and the diagnostic allele length (diagnostic CTG, maroon diamonds and line) for the OPTIMISTIC cohort. The relevant line of best fit under a logarithmic model (model 1a, table 1) is shown for each CTG measure (for ePAL  $n = 222$ ,  $r^2 = 0.178$ ,  $p = 3.1 \times 10^{-11}$  and for diagnostic CTG  $n = 105$ ,  $r^2 = 0.123$ ,  $p = 0.00014$ ).

**Table 1** Age at onset regression models

Model	Adjusted $r^2$	$p$ Value	Measure	Coefficient	Standard error	$t$ Statistic	$p$ Value
<b>1a</b> Age <sub>o</sub> = $\beta_0 + \beta_1 \text{Log(ePAL)}$ , n = 222	0.178	$3.1 \times 10^{-11}$	Intercept, $\beta_0$	80.3	7.9	10.2	$<2 \times 10^{-16}$
			Log(ePAL), $\beta_1$	-23.4	3.3	-7.0	$3.1 \times 10^{-11}$
<b>1b</b> Age <sub>o</sub> = $\beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Variant repeats}$ , n = 222	0.240	$3.4 \times 10^{-14}$	Intercept, $\beta_0$	86.6	7.7	11.2	$<2 \times 10^{-16}$
			Log(ePAL), $\beta_1$	-26.5	3.3	-8.0	$5.5 \times 10^{-14}$
			Variant repeats, $\beta_2$	13.8	3.2	4.3	$2.2 \times 10^{-5}$
<b>1c</b> Age <sub>o</sub> = $\beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Variant repeats} + \beta_3 \text{Sex}$ , n = 222	0.257	$1.3 \times 10^{-14}$	Intercept, $\beta_0$	85.6	7.7	11.2	$<2 \times 10^{-16}$
			Log(ePAL), $\beta_1$	-27.0	3.3	-8.3	$1.4 \times 10^{-14}$
			Variant repeats, $\beta_2$	14.6	3.2	4.6	$7.0 \times 10^{-6}$
			Sex, $\beta_3$	3.8	1.6	2.4	0.015
<b>1d</b> Age <sub>o</sub> = $\beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Variant repeats} + \beta_3 \text{Sex} + \beta_4 \text{Standardised residuals (SI, model 2b)}$ , n = 222	0.260	$2.9 \times 10^{-14}$	Intercept, $\beta_0$	85.7	7.6	11.2	$<2 \times 10^{-16}$
			Log(ePAL), $\beta_1$	-27.0	3.3	-8.3	$1.2 \times 10^{-14}$
			Variant repeats, $\beta_2$	14.5	3.2	4.6	$8.2 \times 10^{-6}$
			Sex, $\beta_3$	3.8	1.6	2.4	0.017
			Standardized residuals (SI), $\beta_4$	-1.0	0.8	-1.3	0.186
<b>1e</b> Age <sub>o</sub> = $\beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Variant repeats} + \beta_3 \text{Sex} + \beta_4 \text{Standardised residuals (SI, model 2b)} + \beta_5 \text{Standardised residuals (SI, model 2b)} \times \text{Log(ePAL)}$ , n = 222	0.304	$<2.2 \times 10^{-16}$	Intercept, $\beta_0$	104.0	8.8	11.8	$<2 \times 10^{-16}$
			Log(ePAL), $\beta_1$	-34.8	3.8	-9.2	$<2 \times 10^{-16}$
			Variant repeats, $\beta_2$	16.6	3.1	5.3	$2.7 \times 10^{-7}$
			Sex, $\beta_3$	3.6	1.5	2.3	0.020
			Standardized residuals (SI), $\beta_4$	25.0	6.8	3.7	$2.9 \times 10^{-4}$
Standardized residuals (SI) $\times$ log(ePAL), $\beta_5$	-11.3	2.9	-3.9	$1.5 \times 10^{-4}$			

Abbreviations: Age<sub>o</sub> = age at onset; SI = somatic instability.

The table shows the number of observations (n), adjusted squared coefficient of correlation (adjusted  $r^2$ ), and statistical significance ( $p$ ) for each model, and the coefficient ( $\beta$ ), standard error,  $t$  statistic, and statistical significance ( $p$ ) associated with each measure in the model. The coefficient ( $\beta$ ) provides an indication of the relative weight of the contribution of each measure to the model and its associated standard error. The  $t$  statistic and corresponding  $p$  value provide an indication of the statistical significance that the measure is adding explanatory power to the model.

investigate genotype–phenotype relationships. As expected,<sup>25,26,40</sup> ePAL was a major modifier of age at onset (figure 2A and table 1, model 1a,  $r^2 = 0.178$ ,  $p = 3.1 \times 10^{-11}$ ). The presence of variant repeats was also highly significantly ( $p = 2.2 \times 10^{-5}$ ) associated with a later age at onset (figure 2A and table 1, model 1b,  $r^2 = 0.240$ ,  $p = 3.4 \times 10^{-14}$ ). In order to better quantify these effects, we calculated the age at onset predicted by ePAL alone (i.e., model 1a) and compared this to the observed age at onset. The mean difference in age at onset between individuals with and without variant repeats was 13.2 years (95% CI 5.7 to 20.7, 2-tailed  $t$  test  $t = -3.7$ ,  $p = 0.0019$ , figure 2B). Recently, it has been reported that there are substantial sex-dependent differences in the relative incidence and severity of some aspects of DM1.<sup>41</sup> Similarly, we recently reported some evidence for sex-dependent genotype–phenotype correlations in age at onset, although these appeared to be driven largely by an excess of mildly affected transmitting grandfathers.<sup>40</sup> Here, sex was revealed as having a marginally

significant effect ( $p = 0.015$ ) on age at onset (table 1, model 1c,  $r^2 = 0.257$ ,  $p = 1.2 \times 10^{-14}$ ).

### Determinants of somatic instability in the OPTIMISTIC cohort

We defined the degree of somatic instability as the difference between ePAL and the modal allele at recruitment (time point V2). As expected,<sup>25,40</sup> the major determinants of the degree of somatic instability were ePAL (figure 2C), age at sampling, and an interaction between them (table 2, model 2a,  $r^2 = 0.560$ ,  $p < 2.2 \times 10^{-16}$ ). Previously, we also demonstrated in one family that the presence of variant repeats reduced the degree of somatic instability.<sup>31</sup> Likewise, here the presence or absence of variant repeats (figure 2C) was revealed as a highly significant ( $p = 4.9 \times 10^{-8}$ ) additional measure in the regression model (table 2, model 2b,  $r^2 = 0.610$ ,  $p < 2.2 \times 10^{-16}$ ). In contrast, sex of the participant ( $p = 0.48$ ) was found not to

**Table 2** Regression models of the determinants of the degree of SI

Model	Adjusted $r^2$	$p$ Value	Measures	Coefficient	Standard error	$t$ Statistic	$p$ Value
<b>2a</b> $\text{Log(SI)} = \beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Age}_s + \beta_3 \text{Log(ePAL)} \times \text{Age}_s$ , $n = 247$	0.560	$<2.2 \times 10^{-16}$	Intercept, $\beta_0$	1.72	0.73	2.37	0.019
			$\text{Log(ePAL)}$ , $\beta_1$	0.20	0.30	0.66	0.509
			$\text{Age}_s$ , $\beta_2$	-0.05	0.01	-3.73	$2.4 \times 10^{-4}$
			$\text{Log(ePAL)} \times \text{Age}_s$ , $\beta_3$	0.02	0.01	3.90	$1.3 \times 10^{-4}$
<b>2b</b> $\text{Log(SI)} = \beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Age}_s + \beta_3 \text{Log(ePAL)} \times \text{Age}_s + \beta_4 \text{VariantRepeats}$ , $n = 247$	0.610	$<2.2 \times 10^{-16}$	Intercept, $\beta_0$	1.50	0.69	2.19	0.029
			$\text{Log(ePAL)}$ , $\beta_1$	0.29	0.28	1.03	0.304
			$\text{Age}_s$ , $\beta_2$	-0.05	0.01	-3.94	$1.1 \times 10^{-4}$
			$\text{Log(ePAL)} \times \text{Age}_s$ , $\beta_3$	0.02	0.01	4.17	$4.2 \times 10^{-5}$
			Variant repeats, $\beta_4$	-0.39	0.07	-5.63	$4.9 \times 10^{-8}$
<b>2c</b> $\text{Log(SI)} = \beta_0 + \beta_1 \text{Log(ePAL)} + \beta_2 \text{Age}_s + \beta_3 \text{Log(ePAL)} \times \text{Age}_s + \beta_4 \text{VariantRepeats} + \beta_5 \text{Sex}$ , $n = 247$	0.609	$<2.2 \times 10^{-16}$	Intercept, $\beta_0$	1.49	0.69	2.17	0.031
			$\text{Log(ePAL)}$ , $\beta_1$	0.30	0.28	1.06	0.292
			$\text{Age}_s$ , $\beta_2$	-0.05	0.01	-3.92	$1.2 \times 10^{-4}$
			$\text{Log(ePAL)} \times \text{Age}_s$ , $\beta_3$	0.02	0.01	4.16	$4.5 \times 10^{-5}$
			Variant repeats, $\beta_4$	-0.39	0.07	-5.67	$4.1 \times 10^{-8}$
			Sex, $\beta_5$	-0.02	0.04	-0.71	0.481

Abbreviation: SI = somatic instability.

The table shows the number of observations ( $n$ ), adjusted squared coefficient of correlation (adjusted  $r^2$ ), and statistical significance ( $p$ ) for each model, and the coefficient ( $\beta$ ), standard error,  $t$  statistic, and statistical significance ( $p$ ) associated with each measure in the model.

improve the regression model (table 2, model 2c,  $r^2 = 0.609$ ,  $p < 2.2 \times 10^{-16}$ ).

### The role of somatic instability on age at onset in the OPTIMISTIC cohort

As longer inherited allele lengths precipitate an earlier age at onset in DM1, and as somatic instability is highly biased toward expansions, it seems logical to expect that individual-specific differences in the rate of expansion should modify age at onset. Indeed, we previously established evidence for such an effect in a mixed cohort of patients with DM1 from Scotland, the United States, and Costa Rica,<sup>25</sup> and confirmed this effect in a large Costa Rican cohort.<sup>40</sup> In the OPTIMISTIC cohort, incorporating the standardized residuals of somatic instability derived from model 2b into the age at onset model (table 1, model 1d,  $r^2 = 0.260$ ,  $p = 3.0 \times 10^{-14}$ ) improved the  $r^2$  value slightly from model 1c, but the standardized residuals of somatic instability did not reach statistical significance as a measure ( $p = 0.19$ ). However, when the standardized residuals of somatic instability were included with an interaction with ePAL, the model was improved further (table 1, model 1e,  $r^2 = 0.304$ ,  $p < 2.2 \times 10^{-16}$ ) and both the standardized residuals of somatic instability alone ( $p = 2.9 \times 10^{-4}$ ) and the interaction with ePAL ( $p = 1.5 \times 10^{-4}$ ) were revealed as highly significant measures.

### Genetic correlations with progressive phenotypes in the OPTIMISTIC cohort

As part of the OPTIMISTIC protocol, several DM1 phenotypes were also quantified at baseline using a variety of direct

and indirect (i.e., questionnaires) assessment tools.<sup>27</sup> To investigate the genotype–phenotype relationships of these progressive phenotypes, we applied a model selection process incorporating ePAL, residual variation in somatic instability, variant repeats, sex and age at baseline, and interactions between ePAL, residual variation in somatic instability, and age at baseline (table 3). Highly significant associations with effect sizes comparable to those observed with age at onset were revealed for many measures (table 3 and figure 3), including the Muscle Impairment Rating Scale (MIRS,  $r^2 = 0.32$ ,  $p = 2.6 \times 10^{-19}$ ), the OPTIMISTIC primary outcome, the DM1-Activ-c score ( $r^2 = 0.23$ ,  $p = 2.9 \times 10^{-13}$ ), and key secondary outcomes such as the 6-Minute Walk Test (6-MWT,  $r^2 = 0.31$ ,  $p = 2.3 \times 10^{-18}$ ), physical activity (accelerometry, mean,  $r^2 = 0.25$ ,  $p = 3.9 \times 10^{-10}$ , and most active 5-hour,  $r^2 = 0.25$ ,  $p = 8.9 \times 10^{-12}$ ), the Trail-Making Test (TMT-A,  $r^2 = 0.23$ ,  $p = 2.2 \times 10^{-14}$ , and TMT-B,  $r^2 = 0.24$ ,  $p = 4.5 \times 10^{-12}$ ), and the Stroop interference score ( $r^2 = 0.21$ ,  $p = 2.8 \times 10^{-12}$ ). Moderate associations were observed for the Adult Social Behavior Questionnaire (ASBQ,  $r^2 = 0.11$ ,  $p = 1.6 \times 10^{-4}$ ), the Apathy Evaluation Scale–Clinician Version ( $r^2 = 0.10$ ,  $p = 5.1 \times 10^{-5}$ ), the Fatigue and Daytime Sleepiness Scale (FDSS) ( $r^2 = 0.09$ ,  $p = 2.9 \times 10^{-5}$ ), the Myotonic Dystrophy Health Index total score ( $r^2 = 0.09$ ,  $p = 3.4 \times 10^{-5}$ ), and Self-Efficacy–Scale 28 for Fatigue ( $r^2 = 0.07$ ,  $p = 2.2 \times 10^{-4}$ ). Only weak or statistically nonsignificant associations were observed for a number of measures, including for van Sonderen Social Support List (SSL) (SSL—Discrepancies [SSL-D], SSL—Interactions



**Table 3** Regression models of the determinants of progressive phenotypes

Dependent variable	Direction	Overall model			Parameter			
		n	r <sup>2</sup>	p Value	Age		Log(ePAL)	
					β	p Value	β	p Value
Muscle impairment rating scale	L	247	0.32	3 × 10 <sup>-19</sup>	-0.06	0.08	-0.1	0.92
6-min walk test distance (m)	H	246	0.31	2 × 10 <sup>-18</sup>	2.89	0.57	-62.7	0.56
Physical activity (most active 5-h)	H	198	0.25	9 × 10 <sup>-12</sup>	-1.11	5 × 10 <sup>-9</sup>	-60.0	5 × 10 <sup>-11</sup>
Physical activity (mean)	H	198	0.25	4 × 10 <sup>-10</sup>	-0.10	0.87	-16.4	0.23
Trail making test—A	L	234	0.24	5 × 10 <sup>-12</sup>	3.14	0.20	97.6	0.06
Trail making test—B	L	242	0.23	2 × 10 <sup>-14</sup>	0.82	2 × 10 <sup>-15</sup>	22.4	3 × 10 <sup>-6</sup>
DM1-Activ-c	H	246	0.23	3 × 10 <sup>-13</sup>	0.43	0.57	-3.3	0.84
Log (Stroop interference score)	L	238	0.21	3 × 10 <sup>-12</sup>	0.02	3 × 10 <sup>-12</sup>	0.7	2 × 10 <sup>-5</sup>
Adult social behavioural questionnaire	L	201	0.11	0.0002	0.99	0.17	31.5	0.03
Apathy evaluation scale—clinician score	L	241	0.10	5 × 10 <sup>-5</sup>	0.45	0.38	20.0	0.07
Fatigue and daytime sleepiness scale	L	242	0.09	3 × 10 <sup>-5</sup>	1.04	0.04	15.6	0.15
Myotonic dystrophy health index	L	245	0.09	3 × 10 <sup>-5</sup>	0.43	2 × 10 <sup>-5</sup>	17.7	0.0003
Self-efficacy scale 28 for fatigue	H	241	0.07	0.0002	-0.04	0.02	-2.4	0.002
Van Sonderen social support—discrepancies	L	196	0.05	0.01	0.22	0.02	9.7	0.04
Illness cognition questionnaire	H	237	0.04	0.02	0.17	0.45	3.5	0.47
Van Sonderen social support—interactions	H	216	0.04	0.02	-0.20	0.06	-10.7	0.09
Individualised neuromuscular quality of life questionnaire—subscale quality of life	L	245	0.04	0.03	-3.56	0.07	-63.2	0.13
Checklist individual strength—fatigue subscale	L	247	0.03	0.01	0.07	0.04		
Sickness impact profile	L	156	0.03	0.02			78.6	0.02
McGill pain score	L	233	0.03	0.02			15.4	0.03
Jacobsen fatigue catastrophizing scale	L	241	0.02	0.03	0.09	0.06	5.4	0.01
Apathy evaluation scale—informant version	L	99	0.02	0.32				
Beck depression inventory—fast screen	L	243	0.01	0.18	0.16	0.36	3.8	0.31

Continued



**Table 3** Regression models of the determinants of progressive phenotypes (continued)

Dependent variable	Direction	Overall model			Parameter			
		n	r <sup>2</sup>	p Value	Age		Log(ePAL)	
					β	p Value	β	p Value
Illness management questionnaire	L	241	0.01	0.28	0.44	0.42	11.1	0.33
Van Sonderen social support—negative interactions	L	218	0.00	0.35			1.4	0.22
Physical activity (least active 5-hour)	L	198	−0.02	0.82				
Caregiver strain index	L	98	−0.07	0.98				

Dependent variable	Standardized residuals (SI, model 2b)		Variant repeats		Sex (Male)		Log(ePAL) × Age		Log(ePAL) × Standardized residuals (SI, model 2b)		Standardized residuals (SI, model 2b) × Age		Log(ePAL) × Standardized residuals (SI, model 2b) × Age	
	β	p Value	β	p Value	β	p Value	β	p Value	β	p Value	β	p Value	β	p Value
	Muscle impairment rating scale	0.3	2 × 10 <sup>−8</sup>	−0.8	6 × 10 <sup>−5</sup>			0.04	0.02					
6-min walk test distance (m)	−25.2	0.0001	115.2	1 × 10 <sup>−5</sup>	51.9	0.0001	−3.24	0.13						
Physical activity (most active 5-h)	−4.3	0.03	22.2	0.003	9.0	0.02								
Physical activity (mean)	33.6	0.12	9.6	0.001	2.2	0.13	−0.14	0.60	−14.5	0.12	−0.64	0.14	0.26	0.17
Trail making test—A	−158.8	0.06	−30.2	0.007			−0.45	0.66	71.4	0.05	2.93	0.09	−1.32	0.08
Trail making test—B			−8.9	0.04										
DM1-Activ-c	−1.6	0.10	12.4	0.001			−0.48	0.14						
Log (Stroop interference score)			−0.5	0.001										
Adult social behavioural questionnaire	−88.5	0.0004	−5.6	0.07	4.4	0.007	−0.40	0.19	37.0	0.0007	1.88	0.001	−0.80	0.001
Apathy evaluation scale—clinician score	−46.8	0.009	−7.0	0.002			−0.15	0.49	22.5	0.004	0.81	0.02	−0.39	0.01
Fatigue and daytime sleepiness scale			−6.5	0.01	−3.7	0.005	−0.36	0.09						
Myotonic dystrophy health index	1.7	0.10	−6.6	0.11										
Self-efficacy scale 28 for fatigue	−0.5	0.01			0.8	0.02								
Van Sonderen social support—discrepancies			−6.1	0.08	3.6	0.08								
Illness cognition questionnaire	16.7	0.04			1.4	0.007	−0.08	0.44	−7.5	0.03	−0.33	0.04	0.15	0.04
Van Sonderen social support—interactions	21.9	0.03	6.2	0.15	−5.6	0.01			−9.6	0.03				

Continued

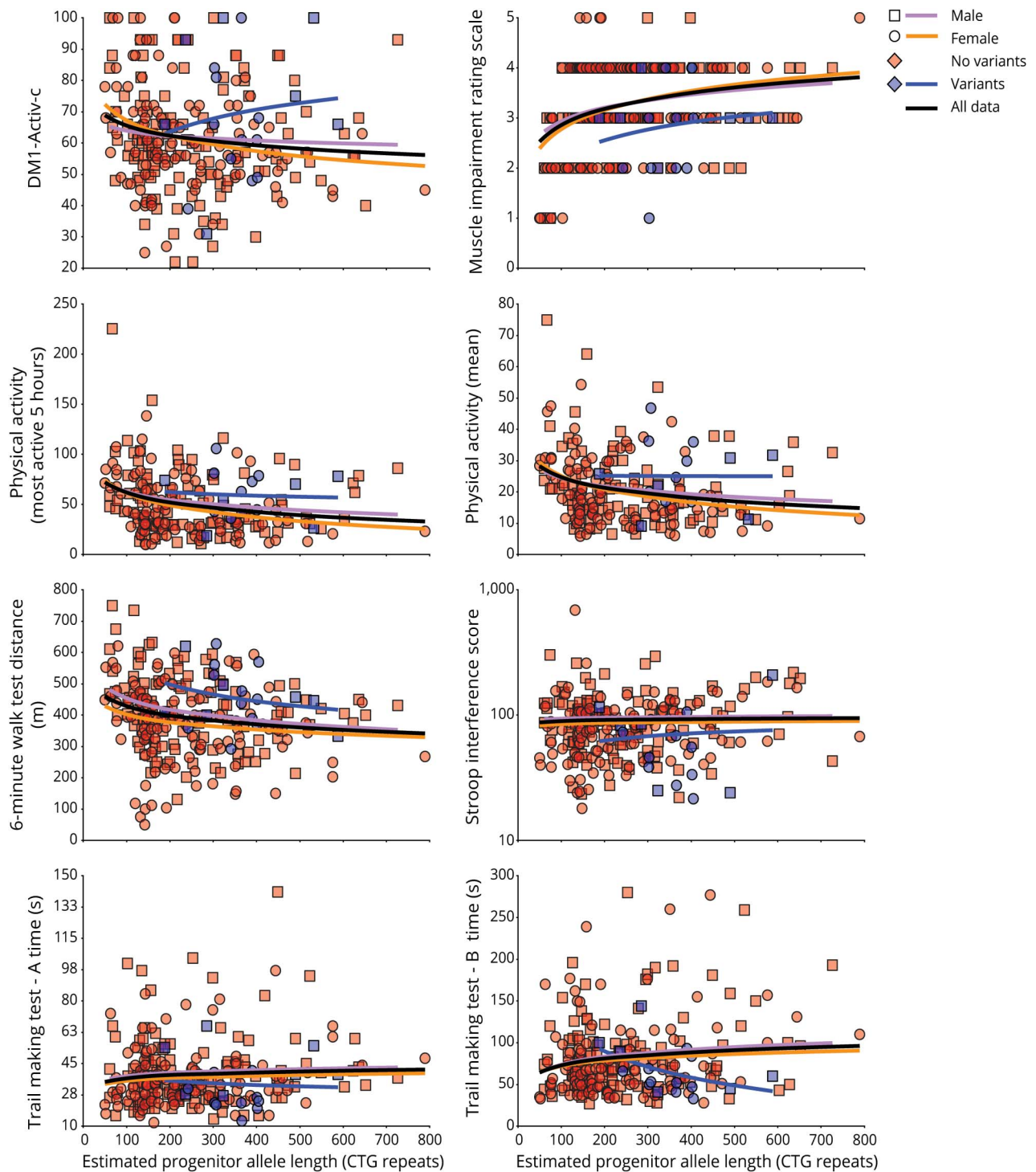
**Table 3** Regression models of the determinants of progressive phenotypes (continued)

Dependent variable	Standardized residuals (SI, model 2b)		Variant repeats		Sex (Male)		Log(ePAL) × Age		Log(ePAL) × Standardized residuals (SI, model 2b)		Standardized residuals (SI, model 2b) × Age		Log(ePAL) × Standardized residuals (SI, model 2b) × Age	
	β	p Value	β	p Value	β	p Value	β	p Value	β	p Value	β	p Value	β	p Value
<b>Individualised neuromuscular quality of life questionnaire—subscale quality of life</b>	90.1	0.19					1.78	0.04	−42	0.16	−2.26	0.10	1.07	0.08
<b>Checklist individual strength—fatigue subscale</b>	−4.0	0.01									0.09	0.006		
<b>Sickness impact profile</b>														
<b>McGill pain score</b>					−6.5	0.06								
<b>Jacobsen fatigue catastrophizing scale</b>														
<b>Apathy evaluation scale—informant version</b>														
<b>Beck depression inventory—fast screen</b>	−17.1	0.007					−0.06	0.42	7.6	0.005	0.29	0.02	−0.13	0.02
<b>Illness management questionnaire</b>	−25.6	0.17					−0.18	0.44	11.5	0.15	0.64	0.09	−0.29	0.08
<b>Van Sonderen social support—negative interactions</b>	−3.6	0.10							1.6	0.09				
<b>Physical activity (least active 5-hour)</b>														
<b>Caregiver strain index</b>														

Abbreviation: SI = somatic instability.

The table shows the number of observations (n), adjusted squared coefficient of correlation (adjusted  $r^2$ ), and statistical significance (p) for each overall model for each dependent variable (phenotype), and the coefficient (β) and statistical significance (p) associated with each measure in the model and their interactions (×). The direction indicates whether a higher (H) or lower (L) score for the phenotype represents a better (i.e., “healthier”) result for the individual. Note that the blank cells correspond to independent variables that were rejected by the model selection procedure, as their inclusion did not significantly improve the overall regression model for the given phenotype.

**Figure 3** Estimated progenitor allele length (ePAL) and variant repeat correlations with age-dependent phenotypes



The scatterplots show the relationship between ePAL and a number of age-dependent phenotypes. The relevant line of best fit under a logarithmic model (dependent variable =  $\beta_0 + \beta_1 \text{Log[ePAL]}$ ) for female individuals (circles, orange line), male individuals (squares, lavender line), and sex-averaged for all individuals (black line) are shown. Individuals with (blue) and without (red) Acil-sensitive variant repeats are also depicted, along with the logarithmic regression line for Acil-sensitive variant repeat carriers (blue line).

[SSL-I], and SSL—Negative Interactions [SSL-N]), CIS-Fatigue, the McGill Pain Questionnaire, the Sickness Impact Profile, and the Beck Depression Inventory—fast screen (BDI-fs), all  $r^2 < 0.05$ ,  $p > 0.01$ , and  $n > 150$  (table 3). For all

of the highly associated measures ( $r^2 > 0.15$ ), age, ePAL, and presence of variant repeats were revealed as important explanatory variables (table 3). The relative degree of somatic mosaicism was also an important explanatory variable for all

bar the Stroop interference score and TMT-A. Sex was also an important modifier for the 6-MWT and physical activity (accelerometry, mean, and most active 5-hour) (table 3).

## Discussion

We analyzed the length and dynamics of the disease-causing CTG repeat expansion in the *DMPK* gene to reveal the primary genetic determinants of symptomatic variation in the OPTIMISTIC cohort. As expected, we were able to establish ePAL as a major determinant of self-reported age at onset (table 1, model 1a). However, it is notable that the proportion of variance in age at onset explained by ePAL ( $r^2 = 0.178$ ) was much lower than observed in previous studies (e.g.,  $r^2 = 0.640$  in Morales et al.<sup>25</sup>). This is attributable to the eligibility constraints for recruitment to OPTIMISTIC that dictated that participants must be aged at least 18 years, have severe fatigue (CIS-Fatigue severity score  $\geq 35$ ), but nonetheless retain the ability to walk independently, and be able to provide informed consent and be motivated to undergo CBT. Thus, OPTIMISTIC has fewer very mildly affected participants with small expansions, and fewer severely affected patients with very large DM1 expansions (figure 2D). Selection of moderately affected patients is likely to be a common feature of many of the early DM1 trials, suggesting that the genetic characteristics and phenotypic relationships of the OPTIMISTIC cohort will likely be representative of other DM1 trial cohorts. Although age at onset is an important aspect of the disease phenotype, it is somewhat subjective in its assessment as it may depend on patient recall, patient knowledge and relative cognition, and insight.<sup>42</sup> Moreover, age at onset is unlikely in the near term to be an outcome measure in clinical trials. We have also revealed that ePAL is an important contributor to multiple age-dependent phenotypes that were also assessed as part of the OPTIMISTIC protocol, including key measures of muscle function and activity such as the MIRS, 6MWT, DM1-Activ-c, and accelerometry data, and key CNS phenotypes such as the TMT and Stroop test. In contrast, other phenotypes such as SSL-D/I/N, CIS-Fatigue, McGill pain, and BDI-fs were not well-explained by genetic factors. It seems reasonable to assume that in general symptoms that are strongly associated with the causative mutation in DM1 are most directly linked to the underlying disease process and may be the most responsive to therapeutic interventions targeting the underlying pathology. It is also possible that some of the more subjective self-reported measures may also be compromised by reduced patient insight,<sup>42</sup> further confounding genotype to phenotype associations.

For many years it was assumed that expanded alleles at the *DMPK* locus comprised pure CTG repeat arrays. However, it has become apparent that a subset of DM1 expansions contains variant repeats.<sup>31,35,37,43,44</sup> Previous estimates for the frequency of such variants in unselected DM1 cohorts vary from ~3% to 5%.<sup>31,35–37</sup> Here, we determined the overall frequency of variant repeats as 8.4% of the total cohort (21/250). In the Nijmegen subcohort, the frequency of variants

was 13.6% (9/66, 95% CI 6.8%–24.8%), while in the Paris subcohort it was only 4.5% (3/67, 95% CI 1.2%–13.4%). Although there was 1 pair of siblings from Nijmegen who shared variant repeats, none of the other participants from Nijmegen with variant repeat containing expanded alleles were known to be related to each other. Thus, chance sampling of one or a few large families segregating variant repeat alleles does not explain the higher frequency of variant repeat alleles observed in the Nijmegen subcohort. While it is possible that this may reflect some sort of site-specific recruitment bias, this difference was not statistically significant and may simply represent random sampling error. Regardless, the presence of variant repeats clearly influences disease severity, with individuals carrying such alleles having an age at onset delayed by an average of ~13 years relative to that predicted using the 1a regression model. It is possible that the reduced disease severity observed in DM1 carriers of repeat alleles led to their overselection in OPTIMISTIC, conceivably mediated by a reduced neuropsychological effect of the disease and higher motivation among this subcohort. In addition to modifying age at onset, these data also reveal the protective effect of variant repeats on many of the progressive phenotypes likely to be outcomes in clinical trials. If variant repeat carriers are over-recruited, this may reduce sensitivity to correctly evaluate efficacy of a therapeutic intervention in a clinical trial. These considerations suggest that testing for the presence of variant repeats should be included in DM1 trial design.

Recently, it has been demonstrated that the absolute frequency and severity of different aspects of the complex DM1 phenotype are differentially expressed in male and female patients.<sup>41</sup> Similarly, we have presented preliminary evidence that suggest there may also be subtle repeat-length-independent sex effects on age at onset.<sup>40</sup> Here, we also observed a marginally significant sex effect ( $p = 0.015$ ) when including sex as a measure in the age at onset model (table 1, model 1c). However, the direction of this effect was the opposite to that previously observed. These observations suggest sex has only a subtle effect on overall age at onset. Nonetheless, sex was revealed as an important cofactor for some phenotypes such as the 6MWT, the ASBQ, and the FDSS (table 3) and should be considered as an important factor in clinical trials.

We recently demonstrated that residual variation in somatic instability not accounted for by age at sampling and ePAL was inversely correlated with residual variation in age at onset not accounted for by ePAL, i.e., patients in whom the repeat expands more rapidly in the soma have earlier ages at onset than expected.<sup>25,40</sup> Here, we have used SP-PCR to calculate the modal length change during the lifetime of the patient as a measure of somatic instability and shown variation in this measure is explained by the expected measures, i.e., ePAL, age at sampling, and a strong interaction between them (table 2, model 2a). In addition, we have confirmed that somatic instability is reduced by the presence of variant repeats. We have also provided additional evidence that somatic mutational dynamics directly modify disease severity, with residual



variation in somatic instability accounting for approximately 4.7% of the total variation in age at onset and detectably contributing to many age-dependent phenotypes and confirming somatic expansion as an important therapeutic target.

Traditionally, DM1 is diagnosed using Southern blot hybridization of restriction digested genomic DNA. The low predictive value of this measure and additional complications mediated by a failure to fully consider the effects of age-dependent somatic expansion on interpreting intergenerational length changes prompted the International Myotonic Dystrophy Consortium to recommend that families not be offered predictive phenotypic information based on the number of CTG repeats.<sup>12</sup> While this recommendation is not universally observed, this suggestion, coupled with the technical challenges of Southern blot hybridizations and the availability of a simple yes/no RP-PCR test,<sup>39</sup> has led to a situation where many diagnostic laboratories do not even attempt to measure the number of CTG repeats. For example, within the OPTIMISTIC cohort, a diagnostic CTG repeat length was available for only 121 out of 255 participants. As expected, the predictive value of this measure was relatively low, the diagnostic CTG length measure accounting for only 12% of the variation in age at onset (figure 2E,  $n = 103$ ,  $r^2 = 0.123$ ,  $p = 0.00014$ ). This contrasts sharply with the ~30% of variation accounted for by estimating progenitor allele length, quantifying somatic mosaicism, and determining the presence or absence of variant repeats (table 1, model 1e). These data suggest that it may be time to revisit the recommended diagnostic criteria for DM1 and the potential value of reporting more informative prognostic information to families by estimating progenitor allele length and testing for the presence of variant repeats.

We have defined the genetic characteristics of the *DMPK* expansion in the cohort of patients with DM1 recruited to the OPTIMISTIC clinical trial. These baseline data have already yielded important insights into genotype to phenotype relationships in DM1 and should provide a route to determining the possible effect of genotype on intervention response and a basis for genetic stratification of DM1 trial participants.

## Acknowledgment

The authors thank Berit A. Adam, Josephine McGhie, Stephanie L. Baross, Roxanna J. Boyd, and Stuart S. Martin (all University of Glasgow) for acquisition of some of the genetic data; the participants and family members of the OPTIMISTIC cohort; colleagues at all 4 clinical recruitment centers for collection and transfer of the blood samples and clinical data; Mojgan Reza and staff at the Newcastle Biobank for Research of Neuromuscular Disorders for preparing and shipping the OPTIMISTIC DNA samples; the local diagnostic laboratories for provision of the historical DNA samples; Shaun Treweek (University of Aberdeen) for trial design and implementation; the Tayside Clinical Trials Unit for collation and provision of participant data; and colleagues in the DGM group at the University of Glasgow for discussions on this work.

## Study funding

Study funded by European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 305697 (the OPTIMISTIC project) and donations to the DGM group from the Myotonic Dystrophy Support Group. The funders had no role in the study design, data collection, analysis, interpretation of data, writing the report, or decisions regarding when to submit publications.

## Disclosure

S. Cumming reports no disclosures relevant to the manuscript. D. Monckton has been a consultant for Biogen Idec and Vertex; has received an honorarium from Charles River for delivering a lecture; is a consultant for, and had a research contract with, AMO Pharma; is on the scientific advisory board of the Myotonic Dystrophy Foundation; is a medical advisor for the Myotonic Dystrophy Support Group; and received research funding from the European Union and the Myotonic Dystrophy Support Group. B. Schoser is a member of the neuromuscular advisory board of Audentes Therapeutics, San Francisco; received speaker honoraria from Sanofi Genzyme, Biomarin, Amicus Therapeutics, Lupin Pharmaceuticals, Recordati Pharma, and CSL Behring during the last 3 years; and received an unrestricted research grant from Sanofi Genzyme (2016–2017). K. Okkersen, C. Jimenez-Moreno, S. Wenninger, F. Daidj, R. Littleford, H. Lochmüller, and G. Gorman report no disclosures relevant to the manuscript. G. Bassez reports grants from AFM-Telethon, personal fees and other from Lupin Pharmaceuticals, non-financial support and other from the Myotonic Dystrophy Foundation, and nonfinancial support from AMO Pharma, outside the submitted work. B. van Engelen reports grants from European Union's Horizon 2020 research and innovation programme (Murab), grants from Netherlands Organisation for Scientific Research (NWO), grants from The Netherlands Organisation for Health Research and Development (ZonMw), grants from Global FSH, grants from Prinses Beatrix Spierfonds, grants from Stichting Spieren voor Spieren, grants from Association Francaise contre les Myopathies, and grants from Dutch FSHD Foundation during the conduct of the study. Go to [Neurology.org/N](http://Neurology.org/N) for full disclosures.

## Publication history

Received by *Neurology* July 2, 2018. Accepted in final form April 10, 2019.

## Appendix 1 Authors

Name	Location	Role	Contribution
Sarah A. Cumming, PhD	University of Glasgow	Author and site investigator	Major role in acquisition of genetic data, analysis and interpretation of data, critical revision of manuscript for intellectual content

Continued

## Appendix 1 (continued)

Name	Location	Role	Contribution
<b>Cecilia Jimenez-Moreno, PhD</b>	Newcastle University	Author and site investigator	Major role in recruitment of participants and acquisition of biological samples and clinical data, critical revision of manuscript for intellectual content
<b>Kees Okkersen, MD</b>	Radboud University	Author and site investigator	Major role in recruitment of participants and acquisition of biological samples and clinical data, critical revision of manuscript for intellectual content
<b>Stephan Wenninger, MD</b>	Ludwig-Maximilians-Universität München	Author and site investigator	Major role in recruitment of participants and acquisition of biological samples and clinical data, critical revision of manuscript for intellectual content
<b>Ferroudja Daidj</b>	Assistance Publique-Hôpitaux de Paris	Author and site investigator	Major role in recruitment of participants and acquisition of biological samples and clinical data, critical revision of manuscript for intellectual content
<b>Fiona Hogarth, PhD</b>	University of Dundee	Author and site investigator	Major role in collation of clinical data, critical revision of manuscript for intellectual content
<b>Roberta Littleford, PhD</b>	University of Dundee	Author and site investigator	Major role in study design and collation of clinical data, critical revision of manuscript for intellectual content
<b>Gráinne Gorman, MD, PhD</b>	Newcastle University	Author and partner lead	Supervision of recruitment, data and sample acquisition and clinical study design, critical revision of manuscript for intellectual content
<b>Guillaume Bassez, MD</b>	Assistance Publique-Hôpitaux de Paris	Author and partner lead	Supervision of recruitment, data and sample acquisition, clinical study design, critical revision of manuscript for intellectual content
<b>Benedikt Schoser, MD</b>	Ludwig-Maximilians-Universität München	Author and partner lead	Supervision of recruitment, data and sample acquisition, clinical study design, critical revision of manuscript for intellectual content
<b>Hanns Lochmüller, MD</b>	Newcastle University	Author and site investigator	Coordination of DNA sample collection, clinical study design, critical revision of manuscript for intellectual content

## Appendix 1 (continued)

Name	Location	Role	Contribution
<b>Baziel G.M. van Engelen, MD, PhD</b>	Radboud University	Author and chief investigator	Supervision of recruitment, data and sample acquisition, and the overall study, clinical study design, critical revision of manuscript for intellectual content
<b>Darren G. Monckton, PhD</b>	University of Glasgow	Author and partner lead	Genetic study concept and design, study supervision, data analysis, drafting of manuscript, critical revision of manuscript for intellectual content

## Appendix 2 Coinvestigators

Coinvestigators are listed at [links.lww.com/WNL/A962](https://links.lww.com/WNL/A962)

## References

- Harper PS. Myotonic Dystrophy. London: WB Saunders; 2001.
- Aslanidis C, Jansen G, Amemiya C, et al. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* 1992;355:548–551.
- Brook JD, McCurrach ME, Harley HG, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 1992;68:799–808.
- Buxton J, Shelbourne P, Davies J, et al. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 1992;355:547–548.
- Fu YH, Pizzuti A, Fenwick RG, et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 1992;255:1256–1258.
- Harley HG, Brook JD, Rundle SA, et al. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 1992;355:545–546.
- Mahadevan M, Tsilfidis C, Sabourin L, et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 1992;255:1253–1255.
- Harley HG, Rundle SA, MacMillan JC, et al. Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am J Hum Genet* 1993;52:1164–1174.
- Hunter A, Tsilfidis C, Mettler G, et al. The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *J Med Genet* 1992;29:774–779.
- Redman JB, Fenwick RG, Fu YH, Pizzuti A, Caskey CT. Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *J Am Med Assoc* 1993;269:1960–1965.
- Salehi LB, Bonifazi E, Stasio ED, et al. Risk prediction for clinical phenotype in myotonic dystrophy type 1: data from 2,650 patients. *Genet Test* 2007;11:84–90.
- The International Myotonic Dystrophy Consortium; Gonzalez I, Ohsawa N, et al. New nomenclature and DNA testing guidelines for myotonic dystrophy type 1 (DM1). *Neurology* 2000;54:1218–1221.
- Ashizawa T, Dubel JR, Dunne PW, et al. Anticipation in myotonic dystrophy: II: complex relationships between clinical findings and structure of the GCT repeat. *Neurology* 1992;42:1877–1883.
- Brunner HG, Brüggerwirth HT, Nillesen W, et al. Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am J Hum Genet* 1993;53:1016–1023.
- Lavedan C, Hofmann-Radvanyi H, Shelbourne P, et al. Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. *Am J Hum Genet* 1993;52:875–883.
- Höweler CJ, Busch HFM, Geraedts JPM, Niermeijer MF, Staal A. Anticipation in myotonic dystrophy: fact or fiction? *Brain* 1989;112:779–797.
- Jansen G, Willems P, Coerwinkel M, et al. Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)<sub>n</sub> variation and selection against extreme expansion in sperm. *Am J Hum Genet* 1994;54:575–585.
- Monckton DG, Wong LJ, Ashizawa T, Caskey CT. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum Mol Genet* 1995;4:1–8.
- Wong LJ, Ashizawa T, Monckton DG, Caskey CT, Richards CS. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet* 1995;56:114–122.

20. Anvret M, Ahlberg G, Grandell U, Hedberg B, Johnson K, Edström L. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum Mol Genet* 1993;2:1397–1400.
21. Ashizawa T, Dubel JR, Harati Y. Somatic instability of CTG repeat in myotonic dystrophy. *Neurology* 1993;43:2674–2678.
22. Thornton CA, Johnson K, Moxley RT. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann Neurol* 1994;35:104–107.
23. Ishii S, Nishio T, Sunohara N, et al. Small increase in triplet repeat length of cerebellum from patients with myotonic dystrophy. *Hum Genet* 1996;98:138–140.
24. Jinnai K, Mitani M, Futamura N, Kawamoto K, Funakawa I, Itoh K. Somatic instability of CTG repeats in the cerebellum of myotonic dystrophy type 1. *Muscle Nerve* 2013; 48:105–108.
25. Morales F, Couto JM, Higham CF, et al. 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.
26. Morales F, Vásquez M, Cuenca P, et al. Parental age effects, but no evidence for an intrauterine effect in the transmission of myotonic dystrophy type 1. *Eur J Hum Genet* 2015;23:646–653.
27. van Engelen B; The OPTIMISTIC Consortium. Cognitive behaviour therapy plus aerobic exercise training to increase activity in patients with myotonic dystrophy type 1 (DM1) compared to usual care (OPTIMISTIC): study protocol for randomised controlled trial. *Trials* 2015;16:224.
28. Okkersen K, Jimenez-Moreno C, Wenninger S, et al. Cognitive behavioural therapy with optional graded exercise therapy in patients with severe fatigue with myotonic dystrophy type 1: a multicentre, single-blind, randomised trial. *Lancet Neurol* 2018;17:671–680.
29. Reza M, Cox D, Phillips L, et al. MRC Centre Neuromuscular Biobank (Newcastle and London): supporting and facilitating rare and neuromuscular disease research worldwide. *Neuromuscul Disord* 2017;27:1054–1064.
30. Gomes-Pereira M, Bidichandani SI, Monckton DG. Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods Mol Biol* 2004;277:61–76.
31. Braida C, Stefanatos RK, Adam B, et al. 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.
32. The R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available at: [R-project.org/](http://R-project.org/). Accessed January 15, 2018.
33. The RStudio Team. RStudio: Integrated Development for R. Boston: RStudio, Inc. Available at: [rstudio.com/](http://rstudio.com/). Accessed August 21, 2017.
34. Venables WN, Ripley BD. *Modern Applied Statistics with S*. New York: Springer; 2002:495, xi.
35. Musova Z, Mazanec R, Krepelova A, et al. Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet A* 2009;149A:1365–1374.
36. Botta A, Rossi G, Marcaurelio M, et al. Identification and characterization of 5' CCG interruptions in complex *DMPK* expanded alleles. *Eur J Hum Genet* 2017;25:257–261.
37. Santoro M, Masciullo M, Pietrobono R, et al. Molecular, clinical, and muscle studies in myotonic dystrophy type 1 (DM1) associated with novel variant CCG expansions. *J Neurol* 2013;260:1245–1257.
38. Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 1991;354:204–209.
39. Warner JP, Barron LH, Goudie D, et al. A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J Med Genet* 1996;33:1022–1026.
40. Morales F, Vásquez M, Santamaria C, Cuenca P, Corrales E, Monckton DG. A polymorphism in the *MSH3* mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. *DNA Repair* 2016;40:57–66.
41. Dogan C, De Antonio M, Hamroun D, et al. Gender as a modifying factor influencing myotonic dystrophy type 1 phenotype severity and mortality: a nationwide multiple databases cross-sectional observational study. *PLoS ONE* 2016;11:e0148264.
42. Baldanzi S, Bevilacqua F, Lorio R, et al. Disease awareness in myotonic dystrophy type 1: an observational cross-sectional study. *Orphanet J Rare Dis* 2016;11:34.
43. Santoro M, Fontana L, Masciullo M, et al. Expansion size and presence of CCG/CTC/CGG sequence interruptions in the expanded CTG array are independently associated to hypermethylation at the *DMPK* locus in myotonic dystrophy type 1 (DM1). *Biochim Biophys Acta* 2015;1852:2645–2652.
44. Tome S, Dandelot E, Dogan C, et al. Unusual association of a unique CAG interruption in 5' of DM1 CTG repeats with intergenerational contractions and low somatic mosaicism. *Hum Mutat* 2018;39:970–982.