Myotonic dystrophy type 1 (DM1) clinical sub-types and CTCF site methylation status flanking the CTG expansion are mutant allele length-dependent

Fernando Morales[1*], Eyleen Corrales[1,#], Baili Zhang[2,#], Melissa Vásquez[1], Carolina Santamaría-Ulloa[1], Hazel Quesada[1], Mario Sirito[2], Marcos R. Estecio[3,4], Darren G. Monckton[5] and Ralf Krahe[2,4]

[1]Instituto de Investigaciones en Salud (INISA), Universidad de Costa Rica, San José, 2060, Costa Rica

[2]Department of Genetics, [3]Department of Epigenetics & Molecular Carcinogenesis and [4]Center for Cancer Epigenetics, University of Texas MD Anderson Cancer Center, Houston, Texas, 77030-4009, USA

[5]Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom

[#] These authors contributed equally to this work

***Corresponding author:**

Dr. Fernando Morales

Ciudad Universitaria Rodrigo Facio

Instituto de Investigaciones en Salud

Universidad de Costa Rica, 2060

San Pedro de Montes de Oca

San José

Costa Rica

Tel: +506 2511 2124

Fax: +506 2511 5130

Email: fernando.moralesmontero@ucr.ac.cr

Abstract

Myotonic dystrophy type 1 (DM1) is a complex disease with a wide spectrum of symptoms. The exact relationship between mutant CTG repeat expansion size and clinical outcome remains unclear. DM1 congenital patients (CDM) inherit the largest expanded alleles, which are associated with abnormal and increased DNA methylation flanking the CTG repeat. However, DNA methylation at the *DMPK* locus remains understudied. Its relationship to DM1 clinical subtypes, expansion size and age-at-onset is not yet completely understood. Using pyrosequencing-based methylation analysis on 225 blood DNA samples from Costa Rican DM1 patients, we determined that the size of the estimated progenitor allele length (ePAL) is not only a good discriminator between CDM and non-CDM cases (with an estimated threshold at 653 CTG repeats), but also for all DM1 clinical subtypes. Secondly, increased methylation at both CTCF sites upstream and downstream of the expansion was almost exclusively present in CDM cases. Thirdly, levels of abnormal methylation were associated with clinical subtype, age and ePAL, with strong correlations between these variables. Fourthly, both ePAL and the intergenerational expansion size were significantly associated with methylation status. Finally, methylation status was associated with ePAL and maternal inheritance, with almost exclusively maternal transmission of CDM. In conclusion, increased DNA methylation at the CTCF sites flanking the DM1 expansion could be linked to ePAL, and both increased methylation and the ePAL could be considered biomarkers for the CDM phenotype.

Introduction

Myotonic dystrophy type 1 (DM1) (OMIM #160900) is a complex multisystem autosomal-dominant disease. The most frequent and disabling symptoms include myotonia, cataracts, muscle weakness and wasting, cardiac problems and central nervous system-related symptoms (1). Clinically, the disease is highly variable within and between families, as is the age-at-onset with recognition of first symptoms as early as birth or as late as 70+ years of age. Based on symptoms and age-at-onset, DM1 patients have been classified into multiple different clinical sub-types. Initially, patients were classified in to four clinical subtypes: late-onset (> 40 years), classical (symptoms > 10 – 40 years), pediatric (≤ 10 years) and congenital (at birth) (2), but in 2016, a new classification composed of five clinical sub-types (late, adult, juvenile, infantile and congenital) was proposed (3). In both classifications, DM1 mutant allele carriers asymptomatic at the time of sampling are not taken into account. The most severe sub-type of the disease is congenital myotonic dystrophy (CDM), which shows a distinct clinical presentation compared to the other recognized DM1 clinical sub-types. Congenital cases present at birth with respiratory distress, feeding difficulties, a "tented" appearance of the mouth, bilateral facial weakness due to neonatal hypotonia, and intellectual disability, which is the most disabling symptom after the child survives the life-threatening neonatal problems during the first year (1). Most CDM children that survive the first year subsequently develop the DM1 classical symptoms by the age of ten (4).

DM1 is caused by the expansion of an unstable CTG repeat localized in the 3′-untranslated region (UTR) of the DM1 protein kinase (*DMPK*) gene (5-7). The repeat is polymorphic in the general population with 5 – 37 CTG repeats (g.17294_17296(5_37)). Affected individuals inherit from ~50 to several thousand CTG repeats (g.17294_17296(50_3000)) (5-7). Overall, mutant allele length displays a positive correlation with disease severity and an inverse correlation with the age-at-onset (8-11), which coupled with a bias toward expansion during germline transmission, provides the molecular basis for the observed genetic anticipation in DM1 families (12, 13).

The CDM sub-type is transmitted almost exclusively through an affected mother (14), with only about 10 cases reported to have been transmitted by an affected father (15-20). DM1 congenital cases generally inherit the largest CTG repeat expansions, however, there are CDM cases with relatively small expansions and non-CDM cases with relatively large expansions (21-23). This suggests that there may be other important factors contributing to or determining the CDM phenotype.

In addition to the clinical presentation and repeat expansion size, another major difference between CDM and the other recognized DM1 clinical sub-types is the methylation status of genomic DNA sequences flanking the CTG repeat tract. The first detailed analysis of DNA methylation at the DM1 locus was performed by Steinbach *et al.* in peripheral blood DNA from DM1 patients using methylation-sensitive restriction enzymes, which suggested that hypermethylation upstream of the CTG repeat was restricted to congenital cases (24). In 2001, Filipova *et al.* described two CTCF-binding sites flanking the CTG repeat on either side, which together with the CTG repeat, appeared to establish an insulator element around the CTG repeat (25). In congenital DM1 cases, the two CTCF sites were abnormally methylated, and increased methylation seemed to prevent the binding of the CTCF protein, leading to disruption of the insulator element (25), implicating DNA methylation as a potential epigenetic mechanism in the genesis of the CDM phenotype (26). However, the insulator function of the CTCF sites has been called into question, and the mechanistic role of the CTCF binding sites flanking the CTG repeat remains to be clarified (27-29). Subsequently, Lopez Castel *et al.* (30) demonstrated abnormal methylation upstream of the CTG repeat in several tissues of DM1 patients, including adult and CDM cases. Using human DM1 embryonic stem cells (hESCs), Yanovsky-Dagan *et al.* (31) showed that the regulatory element present at the DM1 locus could be disrupted due to epigenetic changes as a consequence of the CTG repeat size. Notably, Yanovsky-Dagan *et al.* (31) also showed that longer CTG repeats appeared to drive hyper-methylation from regions upstream of the repeat toward the CTG repeat in a polarized manner. More recently, Barbe *et al.* observed CpG hypermethylation upstream and downstream of the expanded CTG repeat mainly in CDM cases and suggested abnormal CTCF-site methylation as the molecular basis for the parent-of-origin effect for maternal-biased transmission characteristic of CDM (32). Interestingly, the abnormal methylation

observed flanking the CTG repeat expansion has been associated with several clinical symptoms, such as muscle strength, respiratory profile and cognitive functions (33, 34), and more recently, the analysis of methylation flanking the *DMPK* locus has gained interest as a diagnostic biomarker for the CDM clinical sub-type and its prognostic potential for non-CDM forms (35).

Despite these different studies, no definitive relationships between the CDM phenotype and levels of DNA methylation, expansion size, age-at-onset have yet been established. Thus, to provide a more in-depth analysis of methylation at the DM1 locus, we analyzed the methylation status and levels upstream and downstream of the CTG repeat in 225 blood DNA samples from Costa Rican DM1 patients encompassing all the different clinical sub-types. We determined that the size of the estimated progenitor allele length (ePAL) transmitted from parent to offspring is a good discriminator of the different DM1 clinical sub-types, that abnormal methylation is mostly observed in the congenital DM1 cases, and that patients with larger expanded alleles are more likely to show abnormal methylation flanking the DM1 expansion. By combining ePAL and methylation status at the CTCF sites flanking the DM1 expansion, we demonstrate their utility as biomarkers for the CDM phenotype.

Results

Estimated progenitor allele length (ePAL) can discriminate between different DM1 clinical sub-types

The DM1 patients involved in this study (n = 225) belonged to 62 apparently unrelated families. Age-at-sampling ranged from just after birth (within the first five months) to 83 years. For some of these families, only one sample (the proband) was available, whereas for others, multiple samples from two or three generations were accessible. Some data regarding this Costa Rican DM1 study population have been published previously (10, 11, 36). The progenitor allele length (ePAL, estimated progenitor allele length transmitted by the affected parent to their offspring) could be estimated in 223 samples. At sampling, 16 individuals remained asymptomatic and 209 were symptomatic. However, age-at-onset in ten affected individuals could not be established with certainty, but

based on available clinical information they could still be classified into one of the four/five clinical subtypes of the disorder. Detailed information about each DM1 patient involved in this study can be found in Supplementary Table S1.

First, we sought to analyze the relationship between age-at-onset and disease severity using the ePAL. The ePAL showed a positive correlation with the disease severity and a negative correlation with the age-at-onset. ePALs were much larger in the CDM population than in the non-CDM (Table 1, Supplementary Table S1). By performing a logistic regression, we determined that the ePAL is significantly associated with the CDM phenotype (Table 2, Model LR1). To determine if estimation of the PAL in any given DM1 patient could facilitate a more accurate prediction of the CDM1 phenotype, we used receiver operating characteristics (ROC) and precision-recall curve (PRC) analyses. The PRC approach is generally preferred when there is an imbalance in the relative proportion of observations in each class of the data to be analyzed (37), such as in this case (21 CDM cases vs. 202 non-CDM cases). Both ROC and PRC analysis showed that the model presented good quality (0.91), with an accuracy of 95% and a cutoff point defined at 653 CTG repeats (p-value < 0.0001, n = 223) (Supplementary Figure S1). Notably, we observed a highly significant difference in the proportion of CDM cases above and below this threshold: 85.7% (18 of 21) of the CDM cases carry an ePAL ≥ 653 CTG repeats, while only 7.43% (15 of 202) of the non-CDM cases carry an ePAL ≥ 653 CTG repeats (Fisher's exact test, p-value < 0.00001, n = 223).

As the ePAL is a good predictor for the CDM and non-CDM cases, we next aimed to answer whether the ePAL could also be used to discriminate among DM1 clinical sub-types, in other words, if the ePAL differed by clinical sub-type. To test this possibility, we performed a series of multinomial regression models in which, one at a time, we tested each clinical sub-type as the reference category. As there are four different clinical sub-types, we performed four multinomial regression analyses. Multinomial regression models showed a consistent significant association between ePAL and clinical sub-type, indicating that the ePAL is different between them, even after adjusting for multiple comparisons (p-value = 0.0125) (Table 3, Models MNR1 – 4). Our results indicate that the ePAL is a good discriminator of DM1 clinical sub-types. Using the Hecht *et al.* clinical classification (2), a graphical distribution of the ePAL range by clinical

sub-type is shown in Figure 1, which shows less overlap between the ePAL ranges and the clinical sub-types than when considering the modal expansion size obtained by Southern blot hybridization of restriction digested genomic DNA (2, 3). Table 1 shows the 25th and 75th percentiles for the ePAL range for each clinical sub-type.

In 2016, De Antonio and colleagues proposed an alternative clinical classification for DM1, with five instead of four clinical sub-types (3). To determine to which clinical classification our data had a better fit, we ran a sensitivity analysis by categorizing all Costa Rican DM1 patients according to the aforementioned classification (Supplementary Table S1). Using the five clinical sub-types proposed by De Antonio *et al.* (3), multinomial regression models consistently showed a significant association between ePAL and clinical sub-type, even after correcting for multiple comparisons (p-value = 0.01) (Supplementary Table S2, Models MNR1 – 5). Using this five-categories clinical classification, we also visualized the distribution of the ePAL range for each clinical sub-type (Supplementary Figure S2), but in this case, despite the fact that the ePAL is the best discriminator for the different DM1 clinical sub-types, the overlap between clinical sub-types was larger as compared to the Hecht four-categories classification, especially between the adult, juvenile and infantile clinical sub-types. Although both sets of models show a good fit, the four-categories classification model had a better fit as indicated by a greater Mcfadden pseudo r^2 coefficient as compared to the five-categories model (0.4549 vs. 0.2664, respectively). Also, the four-category model had a greater correct classification rate than the five-category model (79.1% vs. 58.4%, respectively), indicating that with fewer categories, DM1 cases can be better predicted by the model.

Increased methylation around the CTG repeat expansion is preferentially present in congenital DM1 cases

Having demonstrated that the ePAL is the major determinant of DM1 disease severity and predictor for DM1 clinical sub-type, we analyzed the relationship between ePAL and methylation flanking the CTG repeat expansion. By carrying out a Pyrosequencing-based Methylation Analysis (PMA), we were able to determine the genomic DNA methylation status and levels flanking the CTG repeat expansion (at CTCF1 and CTCF2) in 225 blood DNA samples from Costa Rican DM1 patients (Figure 2, Supplementary Figure S3). For upstream methylation (MetUS), we were able to quantify methylation levels at the 11 CpG sites

interrogated in all 225 DM1 samples. For downstream methylation (MetDS), we interrogated six CpG sites and obtained results for these six CpGs in most of the samples (170 of 217) (Figure 2, Supplementary Table S1).

Samples were considered abnormally methylated if methylation levels were > 10%. Thus, we observed MetUS in 13.78% (31 of 225) and MetDS in 11.52% (25 of 217) of the DM1 patients analyzed (Table 1, Figure 2). Co-occurring methylation at CTCF1 and CTCF2 was observed in 24 samples. In order to determine if MetUS levels were different than MetDS, we only used the samples with methylation at both CTCF1 and CTCF2. By Wilcoxon Rank Sum test, MetUS levels (mean = 28.84, SD = 11.64) were significantly higher than MetDS (mean = 20.60, SD = 7.80) (Wilcoxon Z = -3.043, p-value = 0.002, n = 24).

We also observed that methylation appeared to be confined to patients with large CTG repeat expansions (> 518 CTG repeats, see below), including classical, pediatric and congenital patients. To determine if the presence of methylation differed by DM1 clinical sub-types, we performed a Fisher´s exact test. We observed that the presence of methylation (either upstream or downstream) differed by DM1 clinical sub-types (classical, pediatric, congenital), being most frequently observed in the CDM cases, upstream (61% (19/31), p-value < 0.00001) and downstream (76% (19/25), p-value < 0.00001) of the CTG repeat expansion. Nonetheless, considering only the methylated samples (31 for MetUS and 25 for MetDS), we observed MetUS in 39% (12/31) and MetDS in 24% (6/25) of non-CDM cases, suggesting that increased methylation flanking the CTG repeat expansion, although preferentially present in CDM cases, it is also present in a significant number of non-CDM cases (Figure 2).

Abnormal methylation flanking the CTG repeat expansion is allele-length dependent

Increased methylation around the CTG repeat expansion is preferentially present in the CDM cases, which carry the largest CTG repeat expansions, suggesting that abnormal methylation could be a consequence of the size of the CTG repeat expansion. In order to test whether methylation status around the CTG repeat is associated to the size of the repeat expansion, we first determined the major modifiers of increased methylation levels upstream and downstream of the CTG repeat expansion. By performing a multivariate regression analysis, we found that about 61% of the variation on levels of MetUS is explained by the log

transformation of age, the log transformation of ePAL and by disease (CDM vs. non-CDM), but not by sex, (Table 4, Model 1), while about 67% of the variation on levels of MetDS is explained by the log transformation of age, the log transformation of ePAL and by disease (CDM vs. non-CDM), but not by sex, (Table 4, Model 2). This analysis indicated an apparent negative association between increased methylation levels and age, and a positive association between methylation levels and the ePAL. In the DM1 cases with increased methylation, ePAL showed a significant positive correlation between the log transformation of ePAL and the levels of MetUS (Spearman's = 0.582, *p*-value = 0.001, n = 29), but not with the levels of MetDS (Spearman's = 0.182, *p*-value = 0.395, n = 24) (Figure 3A). For the age-at-sampling of each DM1 patient, we observed a significant negative correlation between the log transformation of age-at-sampling and the levels of MetUS (Pearson's = -0.378, *p*-value = 0.036, n = 31), but not with the levels of MetDS (Spearman's = -0.350, *p*-value = 0.086, n = 25) (Figure 3B).

Because of the apparent association between ePAL and abnormal methylation levels, we next analyzed if the ePAL was also associated with the methylation status in DM1 patients. For this, we performed a logistic regression, and the results indicated that the ePAL is positively associated with MetUS (Table 2, Model LR2) and MetDS (Table 2, Model LR3). These results suggested that the larger the ePAL, the more likely abnormal methylation flanking the CTG repeat expansion is to occur. To determine a possible ePAL threshold at which a mutated DM1allele is more likely to be abnormally methylated, we used ROC and PRC analyses. For MetUS, the analysis showed that the model presented good quality (0.96), with an accuracy of 98% and a cutoff point defined at 518 CTG repeats (*p*-value < 0.0001, n = 223) (Supplementary Figure S4). Using a cutoff of 518 repeats, we observed that 100% (29 of 29) of the cases showing increased methylation had an ePAL ≥ 518 CTG repeats, while only 7.43% (28 of 194) of the unmethylated cases carried an ePAL ≥ 518 CTG repeats (Fisher's exact test, *p*-value < 0.00001, n = 223). For MetDS, both ROC and PRC analyses showed that the model presented good quality (0.95), with and accuracy of 97% and a cutoff of 558 CTG repeats (*p*-value < 0.0001, n = 223) (Supplementary Figure S5). Using a cutoff of 558 repeats, we observed that 100% (24 of 24) of the cases with increased methylation carried an ePAL ≥ 558 CTG repeats, while only 13% (25 of 192) of the unmethylated cases carried an ePAL ≥ 518 CTG repeats (Fisher's exact test, *p*-value < 0.00001, n = 216). Interestingly, there were four CDM cases in which we did not observe abnormal methylation upstream

and/or downstream of the CTG repeat. Two of them had an ePAL < 518 CTG repeats and did not show abnormal upstream and downstream methylation. The other two had an ePAL > 518, one of which showed increased upstream methylation only, while the other showed increased downstream methylation only (Supplementary Table S1).

Intergenerational CTG expansion contributes to abnormal methylation flanking the CTG repeat expansion

We hypothesized that the size of the intergenerational expansion of the CTG repeat expansion (i.e., the difference between the offspring and the parental ePAL) could be another factor that affects the methylation status of mutant DM1 alleles. In order to test this possibility, we performed a multivariate logistic regression on our data using methylation status as the dependent variable, and ePAL and the intergenerational expansion (IE) as the independent variables. Our analyses indicated that both ePAL and IE significantly contribute to the probability of the mutated allele being abnormally methylated upstream (Table 2, Models LR4 and LR5) of the CTG repeat expansion. To determine a possible intergenerational DM1 expansion size threshold for increased upstream methylation, we again used ROC and PRC analyses. For MetUS, the model presented good quality (0.63) with an accuracy of 78% and a cutoff of 541 CTG repeats (p-value < 0.0001, n = 46) (Supplementary Figure S6). By using a cutoff of 541 repeats, we observed that 61% (14 of 23) of the abnormally methylated cases showed an intergenerational expansion ≥ 541 CTG repeats, while only 4% (1 of 23) of the unmethylated cases showed an intergenerational expansion ≥ 541 CTG repeats (Fisher's exact test, p-value < 0.00001, n = 46). Interestingly, the two CDM cases in which we did not detect increased methylation upstream and downstream of the expanded CTG repeat showed an intergenerational expansion of < 260 CTG repeats. In contrast, the intergenerational expansion of the CDM case with increased MetUS but without MetDS was 567 CTG repeats, and of the CDM case with abnormal MetDS but without MetUS was 469 CTG repeats.

Both maternal inheritance and ePAL are associated with the CDM phenotype

Our combined molecular and family registry data indicated that in all CDM cases (n = 19) abnormally methylated mutant DM1 alleles were maternally transmitted. Interestingly, with the exception of one paternally inherited case in each sub-type, all classical cases and pediatric cases also showed maternal

inheritance of the abnormally methylated DM1 allele. Thus, regardless of clinical sub-type, methylation and parental lineage were significantly associated (MetUS, Fisher's exact test, *p*-value = 0.00001, n = 225; MetDS, Fisher's exact test, *p*-value = 0.00001, n = 217), with increased methylation almost exclusively maternally transmitted (29 of 31 cases, 94%) as is the case for CDM. Figure 4 shows one of the rare CR families with paternal transmission of abnormal methylation flanking the CTG repeat expansion.

Increased methylation around the DM1 mutant repeat has previously been suggested as the underlying mechanism for the observed parent-of-origin effect for the maternal-biased transmission of CDM (32). To confirm this observation, we performed a multivariate logistic regression on our data using methylation status as the dependent variable, and ePAL and inheritance (paternal or maternal) as the independent variables. Our analyses indicated that both maternal inheritance and ePAL in the affected patient contribute significantly to the probability of the mutated allele to being abnormally methylated upstream (Table 2, Model LR6) and downstream (Table 2, Model LR7) of the CTG repeat expansion.

Discussion

Myotonic dystrophy type 1 is considered one of the clinically most highly variable inherited human diseases (1, 38). Clinical variation can range from a few to many organ systems affected, each with a wide spectrum of symptoms from mild to severe. Based on the age-at-onset and the symptoms observed, DM1 patients have been classified into four (2) or five (3) different clinical sub-types, with congenital DM1 (CDM) being the most severe form in both classifications. However, with the exception of the CDM phenotype, which clearly shows distinctive clinical features, DM1 presents as a continuum of clinical phenotypes with no hard boundaries, where no single symptom has been reported as unique to any single sub-type. In addition to the considerable clinical overlap, there also appears to be a large overlap in terms of the DM1 CTG expansion sizes among different clinical sub-types in either classification scheme. Based on the significant overlap between DM1 clinical sub-types and corresponding expansion sizes, including CDM cases with small expansions, and non-CDM cases with

(very) large expansions (21-23), it has been suggested that factors other than mutant allele repeat size are possibly more important in determining the clinical phenotype. Consequently, it is not recommended to provide clinical classification based solely on expansion size (3, 9, 21-23, 32, 39, 40). Importantly, the CTG expansion size used in these studies was measured by determining the mid-point of the smear observed in the standard Southern blot hybridization analysis of restriction digested genomic blood DNA assay as has customarily been done (17, 21, 41). However, it is now well-established that this expansion size is age and size-dependent, and a rather poor indicator of overall disease course and severity (as determined by age-at-onset) (10, 12, 42-48). However, by using larger cohorts of DM1 patients and ePAL measurements to provide better estimates of the inherited expansion size in a given DM1 patient, it may be possible to improve genotype/phenotype correlations and better predict clinical outcome(s), facilitating better clinical classification and management of patients, and more accurate stratification for clinical trials (when they become available).

By using the ePAL (measured in blood DNA by carrying out SP-PCR), we have shown that clinical correlations in DM1 can be significantly improved (10, 11, 49). These data also suggest that ePAL could be a good predictor of the clinical outcome(s) of DM1 patients. In the present study, we have used a reasonably large number of representative patients for each DM1 clinical sub-type, and, we have demonstrated that the ePAL can discriminate between CDM and non-CDM cases (Table 2, Model LR1), although a small overlap remains. In fact, we identified a threshold (653 CTG repeats) above which the probability of developing a CDM phenotype increases significantly (Supplementary Figure S1), even though a previous study has indicated that the CTG repeat expansion cannot determine the CDM phenotype (32). We found non-CDM cases with ePAL > 653 repeats, but these cases are a minority in our study population (Supplementary Table S1). Interestingly, eight of these cases (of 15 in total) were classified as pediatric cases with very early age-at-onset. Since DM1 is a complex disease and the exact age-at-onset in many patients is difficult to establish, it is possible if not likely that some of these constitute CDM cases rather than pediatric cases. In addition to this, our data indicate that the ePAL can serve as the major classifier of DM1 clinical sub-types (Table 3), reducing the overlap between all DM1 clinical sub-types, in terms of the mutated allele size (Table 1, Figure 1). This indicates that the ePAL can be used as a prognostic

biomarker of disease severity in DM1. As the sample size for each clinical sub-type was not large, we were initially surprised by the statistically significant differences when comparing the expansion size between different clinical sub-types, but our results indicate that ePAL is a sufficiently robust measure to distinguish and predict clinical outcome. Similar results have been reported previously (3). However, in our case, we have used a much more robust statistical approach in combination with the ePAL measure that has enabled us to further refine the range of, and differences between, the different clinical sub-types of DM1 (Table 3, Figure 1). Importantly, our data show a better fit for the four-category clinical classification scheme by Hecht *et al.* (2) than the five-category scheme by De Antonio *et al.* (3) It would be interesting to determine if data from other larger DM1 populations also show a better fit for the Hecht classification. For most diagnostic, clinical or research laboratories, ePAL measurements should not present major technical difficulties, as it involves two routine techniques, namely PCR and Southern blot hybridization. As such, determining ePAL by SP-PCR should be easier and faster than obtaining results by traditional Southern blotting of restriction digested genomic DNA, as the former is less demanding, requires less DNA, is easier, overall, more flexible and lends itself to higher throughput.

In spite of these results, ePAL is likely not an adequate single measure to capture the entire spectrum of clinical features for all DM1 clinical sub-types and additional factors may contribute to the observed clinical variation. One of these additional factors could be abnormal DNA methylation around the CTG repeat expansion, which has been suggested as one of the underlying differences between CDM and non-CDM cases (24, 25, 30, 32).

To assess the contribution of increased DNA methylation in the two CTCF sites flanking the CTG repeat expansion, we determined the genomic DNA methylation status and levels flanking the CTG repeat expansion at two CTCF sites immediately flanking the repeat expansion. Consistent with previous studies, especially Barbe *et al.* (32), we also observed DNA hypermethylation preferentially in CDM cases, although it was also present in an appreciable fraction of non-CDM cases at the upstream CTCF1 site (MetUS 12 of 31 cases, 39%) and the downstream CTCF2 site (MetDS six of 25 cases, 24%), respectively, as also reported recently by Hildonen *et al.* (50). As described before (30, 32), we also observed higher MetUS methylation levels compared to MetDS, suggesting a

polarized pattern in terms of methylation levels upstream relative to downstream of the repeat, which may be driven by the expansion as previously suggested (30). MetUS methylation was slightly more common than MetDS (14% vs. 12% of cases Table 1), and most patients with increased MetUS levels showed concordantly increased MetDS (24 of 29, 83%). Only one patient with MetDS methylation was discordant for MetUS (3%). Almost complete concordance between abnormal upstream and downstream methylation in our sample set (Figure 2) appears to be contrary to the previously suggested methylation boundary (i.e., lack of abnormal downstream methylation, when methylation upstream is abnormal (30)). Taken together, our data indicate that when interrogating aberrant methylation flanking the CTG repeat, analysis of MetUS is sufficient to capture most of the variation in methylation for the region (Figure 2).

Barbe *et al.* attempted to identify relationships between hypermethylation flanking the CTG expansion, patient age and expansion size, but did not find any (32). However, by performing robust statistical analyses, we were able to identify the major modifiers of MetUS and MetDS (Table 4) and a relationship between increased DNA methylation around the DM1 expansion, patient age and expansion size (Figure 3). Although our, and the previous study (32), used different methodologies, both are NGS-based approaches that quantify methylation levels with high sensitivity and specificity; and the reported aberrant levels of hypermethylation were generally comparable. This suggests that any observed differences may be, at least in part, due to differences in patients' expansion size and age in the respective studies. In our study, we used the ePAL, which has been demonstrated to be the major modifier of age-at-onset, disease severity and clinical outcome in DM1 ((10, 11, 49, 51), and this study). In contrast, Barbe *et al.* (32) used an expansion size that is age-dependent and subject to patient-specific mutational dynamics and rates, factors that can confound clinical correlations in DM1 (10, 44-46, 51-53). As for the age-effect differences, the explanation may lie in the different populations studied, as changes in methylation levels may also be age-dependent (54-56). However, a recent study using non-CDM patients reported that CTG repeat flanking methylation levels in blood DNA were stable over time (50). In this context, it is worth noting that there is only a limited number of studies on DM1 methylation that have also taken into account patient age. Therefore, additional studies with larger population sizes are required to confirm these data.

For methylation status (hypermethylated vs. non-hypermethylated) and ePAL, we observed a significant positive association between the two variables -- that is the larger the expanded allele, the more likely the expanded allele shows increased methylation levels (Table 2, Models LR2 and LR3). Although Yanovsky-Dagan *et al.* analyzed a different DNA region flanking the CTG repeat, they also observed a positive relationship between aberrant upstream methylation and expansion size (31). Interestingly, using a DM1 mouse model, Brouwer *et al.* also reported a positive relationship between repeat size and increased methylation at upstream and downstream sites (27). Thus, it is possible that analysis of additional CpG sites within the *DMPK* gene could provide greater insights into the relationships described in this and previous studies. These studies contrast with previous findings (32), likely due to the statistical approach and expanded allele size used. Interestingly, our data suggest that increased methylation at CTCF sites flanking the CTG repeat could be a consequence of ePAL. Furthermore, we identified a CTG repeat threshold (518 for MetUS and 558 MetDS, respectively) over which the probability that the mutant allele becomes aberrantly methylated significantly increases (Supplementary Figures S4 and S5). In fact, 100% of patients that showed DM1 locus hypermethylation were carriers of expanded DM1 alleles with ePAL > 518 CTG. Interestingly, similar observations were recently made for expanded alleles at the FRAXA locus underlying fragile-X syndrome (FXS) (OMIM #300624), linking DNA hypermethylation with expansion size (57). Of the other repeat expansion diseases studies to date, only Friedreich ataxia (FRDA) (OMIM #229300) has been shown to present an association between DNA methylation and expansion size (58). Several of the genes with large, non-coding repeat expansions underlying myotonic dystrophy type 2 (DM2) (OMIM #602668), spinocerebellar ataxia type 8 (SCA8) (OMIM #608768), and spinocerebellar ataxia type 10 (SCA10) (OMIM #603516) also contain CTCF-binding sites (CTCFBSDB 2.0), but no studies on aberrant CTCF site methylation have been performed to date. In light of the data for DM1, it would be interesting to analyze CTCF site methylation status around these pathogenic expansions, in order to ascertain if there are similar genotype/phenotype associations in these diseases. For diseases caused by small repeat expansions, there is no clear relationship between expansion size and increased DNA methylation in the respective disease locus, although globally altered epigenetic profiles have been described in Huntington's disease (HD) (OMIM #143100) and several spinocerebellar ataxias (SCAs) (59, 60).

Interestingly, however, there also is a small fraction of DM1 patients with ePALs >518 or 588 CTG repeats that are devoid of methylation upstream and/or downstream the CTG repeat. While the methylation threshold is established by the PAL, estimating PAL may be confounded by the ongoing somatic expansion. It is possible that some patients have been sampled at an age where the PAL is no longer present in any blood cells and we have overestimated the true PAL. Indeed, many of these patients were only sampled when they were >35 years old. Another possibility underlying the lack of aberrant methylation are other yet unidentified patient-specific factors, including possibly, environmental factors.

Additionally, and for the first time, we identified in this study the size of the intergenerational expansion change as a contributing risk factor for the increased methylation of the mutated DM1 allele (Supplementary Figure S6). Interestingly, the threshold was similar to the ePAL, approximately 520 CTG repeats. To the best of our knowledge, this has not been reported previously for DM1 or for any other repeat expansion disorder. Based on our results and on the documented somatic and intergenerational expansion changes in DM1, even in the absence of altered methylation levels, increased methylation would not appear necessary for expansion to occur. Instead, our data suggest that increased methylation levels around the DM1 CTG repeat are a consequence of the PAL first, with a lesser contribution of the CTG repeat intergenerational size change (Supplementary Figures S4-6; Table 2, Models LR2 – 5). Collectively, these data suggest that expansion size drives altered methylation flanking the DM1 CTG repeat expansion, consistent with previous observations that expansion promotes altered methylation (31).

We analyzed the association between DNA methylation at the DM1 locus and the CDM phenotype the same way we analyzed that between ePAL and CDM. Since CDM cases generally carry the largest CTG repeat expansions (Table 1) and increased methylation at the two CTCF sites flanking the CTG repeat expansion is positively associated with ePAL (Table 2, Models LR2 and LR3, Figures 2 and 3, Supplementary Figures S4 and S5), one would expect abnormal methylation to also be associated with the CDM phenotype. Indeed, we observed abnormal methylation around the CTG repeat expansion preferentially in CDM cases (Table 1, Figure 2). Our observation agrees with that by Barbe *et al.* (32) who suggested that hypermethylation flanking the CTG repeat expansion is a

stronger indicator of the CDM phenotype than expansion size. However, our data, when taken together with those of Barbe *et al*. (32) do not provide unequivocal evidence for either methylation status or CTG repeat expansion size being more important in determining the CDM phenotype. Rather, our data suggest that the PAL, first, and abnormal methylation, to a lesser degree, contribute to the CDM phenotype, and therefore, both should be considered jointly as biomarkers for CDM. Based on the threshold for CTCF1 methylation (> 518 CTG repeats) we determined, it is tempting to speculate that large expansion alleles favor increased methylation at CTCF sites flanking the CTG mutant repeat expansion, leading to a CDM phenotype, but very likely not through altered chromatin conformational changes (29). However, these two molecular factors may not be the only or even the most important "drivers" in determining the CDM phenotype, and additional contributors likely exist. In fact, in our cohort we identified two CDM cases that lack hypermethylation around the CTG repeat expansion, with ePALs of < 500 CTG repeats and intergenerational expansion sizes < 300 CTG repeats (Supplementary Table S1). These data indicate that a minority of CDM cases (2 of 21, ~ 10%) can manifest in the absence of very large CTG expansions with increased methylation flanking the expansion. In fact, several factors in DM1 patients and pedigrees may be critical to the increased risk for transmission, inheritance and eventual development of the CDM phenotype: (1) the clinical status of the mother (61); (2) the age of the mother (36); (3) the CTCF-site methylation status upstream of the CTG repeat ((32), this study); (4) the maternal inheritance of hypermethylated mutant alleles as possible epimutations ((32), this study); (5) the ePAL and the intergenerational expansion size (this study), and other yet to be determined *cis/trans* factors, which may be identified and characterized through deep *omics* approaches and using larger sample sizes.

Finally, based on the observed exclusive maternal transmission of mutant CTG repeat expansion alleles showing hypermethylation at CpG-sites flanking the repeat in the CDM cases studied, Barbe *et al*. proposed a parent-of-origin effect for the DNA methylator phenotype as the underlying cause for the almost exclusive maternal inheritance of CDM (32). However, in our study we identified two non-CDM cases (Figure 4 shows the genealogy of one of these families) in which the hypermethylated mutant expansion alleles in the offspring were paternally transmitted (Supplementary Table S1). Thus, transmission of

hypermethylated mutant CTG expansion alleles is principally through the maternal germline (29 of 31 hypermethylated cases, 94%), but in rare instances can also occur through the paternal germline. These findings are consistent with results from human embryonic stem cells (hESC) derived from DM1-affected preimplantation embryos, where DNA hypermethylation near the CTG repeat expansion showed paternal transmission (31). As DNA hypermethylation flanking the CTG repeat appears closely correlated with expansion size, most (if not all) of the paternally transmitted CDM cases could be expected to show hypermethylation around the CTG expansion. Although inheritance and methylation status are indicators of the CDM phenotype (Table 2, Models LR6 and LR7), our analysis could not establish which is more important in determining mutant expansion allele abnormal methylation and development of CDM, but stated that both contribute to the CDM phenotype.

In conclusion, our data strongly indicate that ePAL can be used as an effective indicator of the most probable clinical outcome for DM1 patients, by defining a more specific measure of expansion size range and reducing the overlap between all DM1 clinical sub-types. In addition, we demonstrated a tight relationship between age, expansion size, inheritance and altered methylation status at CTCF sites flanking the CTG repeat expansion at the *DMPK* locus, indicating that both ePAL together with methylation status can be used as biomarkers for the CDM phenotype.

Materials and Methods

Patient population. The project was approved by the Ethics Scientific Committee of the Universidad de Costa Rica. All samples were collected after obtaining signed informed consent in accordance with the ethical protocols approved by the Ethical Scientific Committee of the Universidad de Costa Rica. In all, 225 Costa Rican (CR) DM1 patients were involved in this study. Based on their clinical manifestations, age-at-onset and following the Hecht clinical classification (2), patients were classified as belonging to one of four different DM1 clinical sub-types, as previously done by our group (10, 11, 36): 17 late onset, 140 classical, 30 pediatric and 22 congenital cases, in addition to 16 asymptomatic cases (at the moment of sampling). In ten cases, age-at-onset could not be

precisely determined, but nonetheless each participant could still be assigned to one of the above clinical sub-types. According to their clinical history and symptoms, all of these ten cases were non-CDM cases. For all study subjects, we collected and purified DNA from peripheral blood leukocytes (PBL) following standard procedures. The age-at-sampling, age-at-onset and the size of the estimated progenitor allele length (ePAL, sized in 223 patients) for most of the DM1 patients have been previously reported (10, 11, 36). All DNA samples were previously tested for the presence of Acil sensitive CCG/CGG variant repeats and none were detected.

Pyrosequencing-based methylation analysis (PMA) of methylation upstream and downstream of the CTG repeat. Using PMA as previously described (51, 62), we interrogated the methylation status and levels upstream, including the CTCF-binding site (designated CTCF1), and downstream, including the CTCF-binding site (CTCF2), of the DM1 mutation. At CTCF1, we interrogated 11 CpG sites, seven of which are located within the CTCF-binding site; at CTCF2, we interrogated six CpGs, three of which are within the CTCF-binding site (51, 62). For this, 300 ng of PBL genomic DNA was bisulfite-converted using the EZ DNA Methylation-Gold kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. Bisulfite treated-DNA was amplified using ZymoTaq Premix (Zymo Research, Irvine, CA), and PMA was carried out using the PyroMark Q96 MD platform (Qiagen) and PyroMark Gold Q96 reagents (Qiagen) according to the supplier's instructions. For each sample, a minimum of two replicate runs was performed. Results were analyzed with PyroMark CpG 1.0.11 software (Qiagen), which calculates the mean methylation ((mC/(mC+C)) for each CpG site and the methylation index (MI) (average value of mC/(mC+C)) for each CTCF site. Samples were considered to be abnormally methylated when the methylation index was > 10% (63). The final methylation index for each CTCF site and each CpG interrogated, was calculated as an average of the independent runs.

Statistical analyses. Dichotomous variables including the occurrence of the congenital phenotype (vs. non-congenital DM1), the occurrence of abnormal methylation upstream of the CTG repeat (MetUS) (vs. upstream non-methylation), and the occurrence of abnormal methylation downstream of the CTG (MetDS) (vs. downstream non-methylation) were modeled as dependent variables using logistic regression models. Estimated progenitor allele length (ePAL),

intergenerational expansion (IE) size, and parent of origin were included as independent variables in the logistic regression models. A categorical variable, DM clinical sub-type (congenital, pediatric, classical and late-onset, or congenital, infantile, juvenile, adult and late-onset) was modeled as the dependent variable of a multinomial regression model. ePAL and age were included as independent variables of the multinomial regression model. Mcfadden pseudo r^2 coefficient and correct classification rate (64) were used to compare the four vs. five DM1 clinical sub-type models. Continuous characteristics such as upstream and downstream abnormal methylation levels were modeled as dependent variables in multivariate regression models. ePAL, age, sex, and congenital DM were included as independent variables of multivariate regression models. Receiver operating characteristics curve (ROC curve) and the precision recall curve (PRC) analyses were used to determine the efficacy of tests in correctly classifying two clinical conditions (*e.g.*, congenital vs. non-congenital) or methylation status (abnormal methylated vs. non-methylated). Fisher's exact test was used to compare expected and observed results. The Wilcoxon rank sum test was used to compare the averages of two variables and Spearman and Pearson's correlations to study the relationship between two variables. Significance level was set at 5%. When necessary, a Bonferroni correction for multiple comparisons was applied, with the *p*-value adjusted accordingly.

Acknowledgements

We thank all the DM1 families participating in this study. Also, we want also to thank Mr. William Araya for his technical support and assistance provided to this project. The Muscular Dystrophy Association (www.mda.org) (MDA200568) and the Universidad de Costa Rica (UCR) (742-A8-306) funded this research project. Partial support was also provided by the NIH/NCI under award number P30 CA016672 grant, awarded to the University of Texas MD Anderson Cancer Center Epigenomics Profiling Core and the Center for Cancer Epigenetics.

Conflict of Interest Statement

DGM has been a scientific consultant and/or received an honoraria or stock options from Biogen Idec, AMO Pharma, Charles River, Vertex Pharmaceuticals, Triplet Therapeutics, LoQus23 and Small Molecule RNA and has had research contracts with AMO Pharma and Vertex Pharmaceuticals. The other authors declare no conflict of interest with the publication of this article.

References

1. Harper, P.S. (1989) *Myotonic Dystrophy*. W B Saunders Company.
2. Hecht, B.K., Donnelly, A., Gedeon, A.K., Byard, R.W., Haan, E.A. and Mulley, J.C. (1993) Direct Molecular Diagnosis Of Myotonic Dystrophy. *Clin. Genet.*, **43**, 276-285.
3. De Antonio, M., Dogan, C., Hamroun, D., Mati, M., Zerrouki, S., Eymard, B., Katsahian, S., Bassez, G. and French Myotonic Dystrophy Clinical, N. (2016) Unravelling the myotonic dystrophy type 1 clinical spectrum: A systematic registry-based study with implications for disease classification. *Rev. Neurol. (Paris)*, **172**, 572-580.
4. Harris, S., Moncrieff, C. and Johnson, K. (1996) Myotonic dystrophy: will the real gene please step forward! *Hum. Mol. Genet.*, **5**, 1417-1423.
5. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T. *et al.* (1992) 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*, **68**, 799-808.
6. Fu, Y.H., Pizzuti, A., Fenwick, R.G., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., de Jong, P. *et al.* (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**, 1256-1258.

7. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'hoy, K. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*, **255**, 1253-1255.
8. Harley, H.G., Rundle, S.A., MacMillan, J.C., Myring, J., Brook, J.D., Crow, S., Reardon, W., Fenton, I., Shaw, D.J. and Harper, P.S. (1993) Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am. J. Hum. Genet.*, **52**, 1164-1174.
9. Lavedan, C., Hofmann-Radvanyi, H., Shelbourne, P., Rabes, J.-P., Duros, C., Savoy, D., Dehaupas, I., Luce, S., Johnson, K. and Junien, C. (1993) Myotonic dystrophy: size and sex dependent dynamics of CTG meiotic instability, and somatic mosaicism. *Am. J. Hum. Genet.*, **52**, 875-883.
10. Morales, F., Couto, J.M., Higham, C.F., Hogg, G., Cuenca, P., Braida, C., Wilson, R.H., Adam, B., Del Valle, G., Brian, R. *et al.* (2012) 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.*, **21**, 3558-3567.
11. Morales, F., Vasquez, M., Santamaria, C., Cuenca, P., Corrales, E. and Monckton, D.G. (2016) 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 (Amst)*, **40**, 57-66.
12. Ashizawa, T., Dubel, J.R., Dunne, P.W., Dunne, C.J., Fu, Y.H., Pizzuti, A., Caskey, C.T., Boerwinkle, E., Perryman, M.B., Epstein, H.F. *et al.* (1992) Anticipation In Myotonic Dystrophy .2. Complex Relationships Between Clinical Findings and Structure Of the GCT Repeat. *Neurology*, **42**, 1877-1883.
13. Ashizawa, T., Dunne, C.J., Dubel, J.R., Perryman, M.B., Epstein, H.F., Boerwinkle, E. and Hejtmancik, J.F. (1992) Anticipation In Myotonic Dystrophy .1. Statistical Verification Based On Clinical and Haplotype Findings. *Neurology*, **42**, 1871-1877.
14. Tsilfidis, C., Mackenzie, A.E., Mettler, G., Barcelo, J. and Korneluk, R.G. (1992) Correlation Between CTG Trinucleotide Repeat Length and Frequency Of Severe Congenital Myotonic Dystrophy. *Nat. Genet.*, **1**, 192-195.
15. Bergoffen, J., Kant, J., Sladky, J., McDonald-McGinn, D., Zackai, E.H. and Fischbeck, K.H. (1994) Paternal transmission of congenital myotonic dystrophy. *J. Med. Genet.*, **31**, 518-520.
16. de Die-Smulders, C.E., Smeets, H.J., Loots, W., Anten, H.B., Mirandolle, J.F., Geraedts, J.P. and Howeler, C.J. (1997) Paternal transmission of congenital myotonic dystrophy. *J. Med. Genet.*, **34**, 930-933.
17. Di Costanzo, A., de Cristofaro, M., Di Iorio, G., Daniele, A., Bonavita, S. and Tedeschi, G. (2009) Paternally inherited case of congenital DM1: brain MRI and review of literature. *Brain Dev.*, **31**, 79-82.
18. Ohya, K., Tachi, N., Chiba, S., Sato, T., Kon, S., Kikuchi, K., Imamura, S., Yamagata, H. and Miki, T. (1994) Congenital Myotonic Dystrophy Transmitted From an Asymptomatic Father With a DM-Specific Gene. *Neurology*, **44**, 1958-1960.
19. Tanaka, Y., Suzuki, Y., Shimozawa, N., Nanba, E. and Kondo, N. (2000) Congenital myotonic dystrophy: report of paternal transmission. *Brain Dev.*, **22**, 132-134.
20. Zeesman, S., Carson, N. and Whelan, D.T. (2002) Paternal transmission of the congenital form of myotonic dystrophy type 1: a new case and review of the literature. *Am. J. Med. Genet.*, **107**, 222-226.
21. Barcelo, J.M., Pluscauskas, M., MacKenzie, A.E., Tsilfidis, C., Narang, M. and Korneluk, R.G. (1994) Additive influence of maternal and offspring DM-kinase gene CTG repeat lengths in the genesis of congenital myotonic dystrophy. *Am. J. Hum. Genet.*, **54**, 1124-1125.
22. Clark, C., Petty, R.K. and Strong, A.M. (1998) Late presentation of myotonic dystrophy. *Clin. Exp. Dermatol.*, **23**, 47-48.
23. Cobo, A.M., Baiget, M., Demunain, A.L., Poza, J.J., Emparanza, J.I. and Johnson, K. (1993) Sex-Related Difference In Intergenerational Expansion Of Myotonic Dystrophy Gene. *Lancet*, **341**, 1159-1160.

24. Steinbach, P., Glaser, D., Vogel, W., Wolf, M. and Schwemmle, S. (1998) The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *Am. J. Hum. Genet.*, **62**, 278-285.
25. Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V. and Tapscott, S.J. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation- sensitive insulator at the DM1 locus. *Nat. Genet.*, **28**, 335-343.
26. Lanni, S. and Pearson, C.E. (2019) Molecular genetics of congenital myotonic dystrophy. *Neurobiol. Dis.*, **132**, 104533.
27. Brouwer, J.R., Huguet, A., Nicole, A., Munnich, A. and Gourdon, G. (2013) Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus. *J. Nucleic Acids*, **2013**, 567435.
28. Yanovsky-Dagan, S., Mor-Shaked, H. and Eiges, R. (2015) Modeling diseases of noncoding unstable repeat expansions using mutant pluripotent stem cells. *World J. Stem Cells*, **7**, 823-838.
29. Ruiz Buendia, G.A., Leleu, M., Marzetta, F., Vanzan, L., Tan, J.Y., Ythier, V., Randall, E.L., Marques, A.C., Baubec, T., Murr, R. *et al.* (2020) Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion. *Sci. Adv.*, **6**, eaaz4012.
30. Lopez Castel, A., Nakamori, M., Tome, S., Chitayat, D., Gourdon, G., Thornton, C.A. and Pearson, C.E. (2011) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum. Mol. Genet.*, **20**, 1-15.
31. Yanovsky-Dagan, S., Avitzour, M., Altarescu, G., Renbaum, P., Eldar-Geva, T., Schonberger, O., Mitrani-Rosenbaum, S., Levy-Lahad, E., Birnbaum, R.Y., Gepstein, L. *et al.* (2015) Uncovering the Role of Hypermethylation by CTG Expansion in Myotonic Dystrophy Type 1 Using Mutant Human Embryonic Stem Cells. *Stem Cell Reports*, **5**, 221-231.
32. Barbe, L., Lanni, S., Lopez-Castel, A., Franck, S., Spits, C., Keymolen, K., Seneca, S., Tome, S., Miron, I., Letourneau, J. *et al.* (2017) CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy. *Am. J. Hum. Genet.*, **100**, 488-505.
33. Legare, C., Overend, G., Guay, S.P., Monckton, D.G., Mathieu, J., Gagnon, C. and Bouchard, L. (2019) DMPK gene DNA methylation levels are associated with muscular and respiratory profiles in DM1. *Neurol. Genet.*, **5**, e338.
34. Breton, E., Legare, C., Overend, G., Guay, S.P., Monckton, D., Mathieu, J., Gagnon, C., Richer, L., Gallais, B. and Bouchard, L. (2020) DNA methylation at the DMPK gene locus is associated with cognitive functions in myotonic dystrophy type 1. *Epigenomics*, **12**, 2051-2064.
35. Coppede, F. (2020) Epigenetics of neuromuscular disorders. *Epigenomics*, **12**, 2125-2139.
36. Morales, F., Vasquez, M., Cuenca, P., Campos, D., Santamaria, C., Del Valle, G., Brian, R., Sittenfeld, M. and Monckton, D.G. (2015) Parental age effects, but no evidence for an intrauterine effect in the transmission of myotonic dystrophy type 1. *Eur. J. Hum. Genet.*, **23**, 646-653.
37. Saito, T. and Rehmsmeier, M. (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, **10**, e0118432.
38. Harper, P.S. (2001) *Myotonic Dystrophy*. Harcourt Publisers Ltd, London.
39. Geifman-Holtzman, O. and Fay, K. (1998) Prenatal diagnosis of congenital myotonic dystrophy and counseling of the pregnant mother: case report and literature review. *Am. J. Med. Genet.*, **78**, 250-253.
40. Novelli, G., Gennarelli, M., Menegazzo, E., Angelini, C. and Dallapiccola, B. (1995) Discordant Clinical Outcome In Myotonic Dystrophy Relatives Showing (CTG)(N) Greater-Than-700 Repeats. *Neuromuscul. Disord.* **5**, 157-159.
41. Redman, J.B., Fenwick, R.G., Fu, Y.H., Pizzuti, A. and Caskey, C.T. (1993) Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *JAMA*, **269**, 1960-1965.
42. Martorell, L., Johnson, K., Boucher, C.A. and Baiget, M. (1997) Somatic instability of the myotonic dystrophy (CTG)n repeat during human fetal development. *Hum. Mol. Genet.*, **6**, 877-880.

43. Martorell, L., Martinez, J.M., Carey, N., Johnson, K. and Baiget, M. (1995) Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J. Med. Genet.*, **32**, 593-596.
44. Martorell, L., Monckton, D.G., Gamez, J., Johnson, K.J., Gich, I., de Munain, A.L. and Baiget, M. (1998) Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum. Mol. Genet.*, **7**, 307-312.
45. Monckton, D.G., Wong, L.J.C., Ashizawa, T. and Caskey, C.T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.*, **4**, 1-8.
46. Wong, L.-J.C., Ashizawa, T., Monckton, D.G., Caskey, C.T. and Richards, C.S. (1995) Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.*, **56**, 114-122.
47. Jaspert, A., Fahsold, R., Grehl, H. and Claus, D. (1995) Myotonic Dystrophy - Correlation Of Clinical Symptoms With the Size Of the CTG Trinucleotide Repeat. *J. Neurol.*, **242**, 99-104.
48. Marchini, C., Lonigro, R., Verriello, L., Pellizzari, L., Bergonzi, P. and Damante, G. (2000) Correlations between individual clinical manifestations and CTG repeat amplification in myotonic dystrophy. *Clin. Genet.*, **57**, 74-82.
49. Cumming, S.A., Jimenez-Moreno, C., Okkersen, K., Wenninger, S., Daidj, F., Hogarth, F., Littleford, R., Gorman, G., Bassez, G., Schoser, B. *et al.* (2019) Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort. *Neurology*, **93**, e995-e1009.
50. Hildonen, M., Knak, K.L., Duno, M., Vissing, J. and Tumer, Z. (2020) Stable Longitudinal Methylation Levels at the CpG Sites Flanking the CTG Repeat of DMPK in Patients with Myotonic Dystrophy Type 1. *Genes (Basel)*, **11,** 936.
51. Morales, F., Vasquez, M., Corrales, E., Vindas-Smith, R., Santamaria-Ulloa, C., Zhang, B., Sirito, M., Estecio, M.R., Krahe, R. and Monckton, D.G. (2020) 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.*, **29**, 2496-2507.
52. Ashizawa, T., Dubel, J.R. and Harati, Y. (1993) Somatic Instability Of CTG Repeat In Myotonic Dystrophy. *Neurology*, **43**, 2674-2678.
53. Higham, C.F., Morales, F., Cobbold, C.A., Haydon, D.T. and Monckton, D.G. (2012) High levels of somatic DNA diversity at the myotonic dystrophy type 1 locus are driven by ultra-frequent expansion and contraction mutations. *Hum. Mol. Genet.*, **21**, 2450-2463.
54. Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Aspelund, T., Cui, H., Yu, W., Rongione, M.A., Ekstrom, T.J., Harris, T.B. *et al.* (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA*, **299**, 2877-2883.
55. Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J. *et al.* (2012) Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. U S A*, **109**, 10522-10527.
56. Jung, M. and Pfeifer, G.P. (2015) Aging and DNA methylation. *BMC Biol.*, **13**, 7.
57. Brykczynska, U., Pecho-Vrieseling, E., Thiemeyer, A., Klein, J., Fruh, I., Doll, T., Manneville, C., Fuchs, S., Iazeolla, M., Beibel, M. *et al.* (2016) CGG Repeat-Induced FMR1 Silencing Depends on the Expansion Size in Human iPSCs and Neurons Carrying Unmethylated Full Mutations. *Stem Cell Reports*, **7**, 1059-1071.
58. Greene, E., Mahishi, L., Entezam, A., Kumari, D. and Usdin, K. (2007) Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.*, **35**, 3383-3390.
59. Liu, H., Tang, T.S. and Guo, C. (2018) Epigenetic profiles in polyglutamine disorders. *Epigenomics*, **10**, 9-25.
60. Nageshwaran, S. and Festenstein, R. (2015) Epigenetics and Triplet-Repeat Neurological Diseases. *Front. Neurol.*, **6**, 262.
61. Koch, M.C., Grimm, T., Harley, H.G. and Harper, P.S. (1992) Genetic Risks For Children Of Women With Myotonic Dystrophy - Reply. *Am. J. Hum. Genet.*, **50**, 1341-1342.

62. Corrales, E., Vasquez, M., Zhang, B., Santamaria-Ulloa, C., Cuenca, P., Krahe, R., Monckton, D.G. and Morales, F. (2019) Analysis of mutational dynamics at the DMPK (CTG)n locus identifies saliva as a suitable DNA sample source for genetic analysis in myotonic dystrophy type 1. *PLoS One*, **14**, e0216407.
63. Colella, S., Shen, L., Baggerly, K.A., Issa, J.P. and Krahe, R. (2003) Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques*, **35**, 146-150.
64. Allison, P.D. (2012) *Logistic Regression Using SAS: Theory and Application*. SAS Institute, Cary, NC.

Figure 1. Distribution of the ePAL among DM1 clinical sub-types. Box-plot showing the ePAL distribution between four different DM1 clinical sub-types (late onset, bottom; classical, one step up; pediatric, one step up, and congenital, top). Some overlap occurs between clinical sub-types, which is more evident between the classical and pediatric than between the classical or pediatric and congenital sub-types. Legend on the right indicates all information contained on each box-plot in the figure.

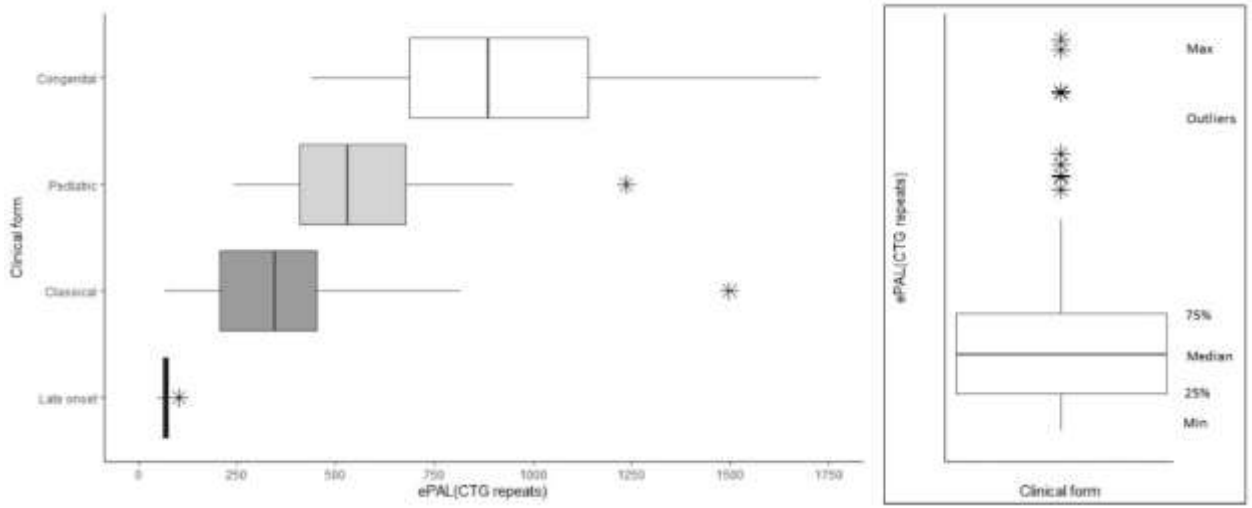

Figure 2. Heatmap of DNA methylation levels upstream and downstream of the DM1 CTG repeat expansion. The heatmap shows that methylation is present, preferentially, in DM1 patients with large CTG repeat expansions. Methylation levels correlate positively with expansion size and negatively with age. Methylation levels measured in 219 different DM1 patients are displayed in a white to red scale (0 to 60%). By using Pyrosequencing-based methylation analysis (PMA), methylation was interrogated in 11 and 6 CpG sites for the CTCF-1 and CTCF-2 binding sites (indicated at both sides of the figure), respectively. Each column of the heatmap represents the data from a single different DM1 patient. Grey, no methylation data determined. Middle-row, color-encoded clinical sub-type and sex of each patient; bottom-row, size of the ePAL (46 – 1,727 CTG repeats) and age (0.5 – 77 years old).

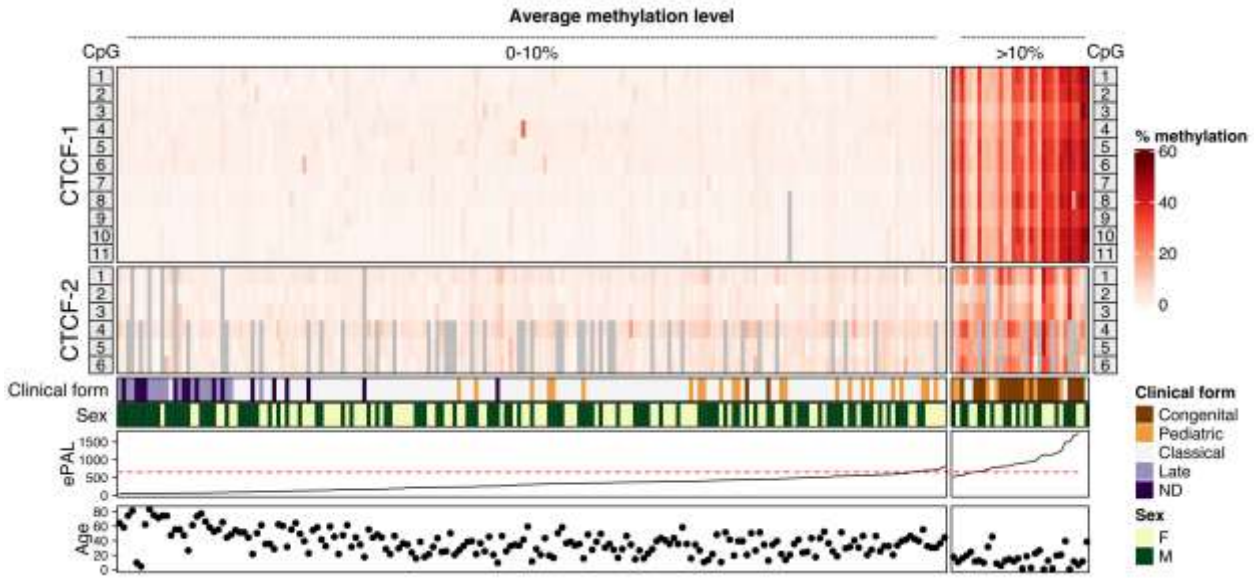

Figure 3. Correlation between methylation levels flanking the CTG repeat expansion and the ePAL and age-at-sampling. Plot A shows a positive relationship between the ePAL and methylation levels, being significant for MetUS (solid line) but not for MetDS (dotted line) (see text), although the data are trending. Plot B shows a negative relationship between the age-at-sampling and methylation levels, being significant for MetUS (solid line) but not for MetDS (dotted line) (see text), although the data are trending.

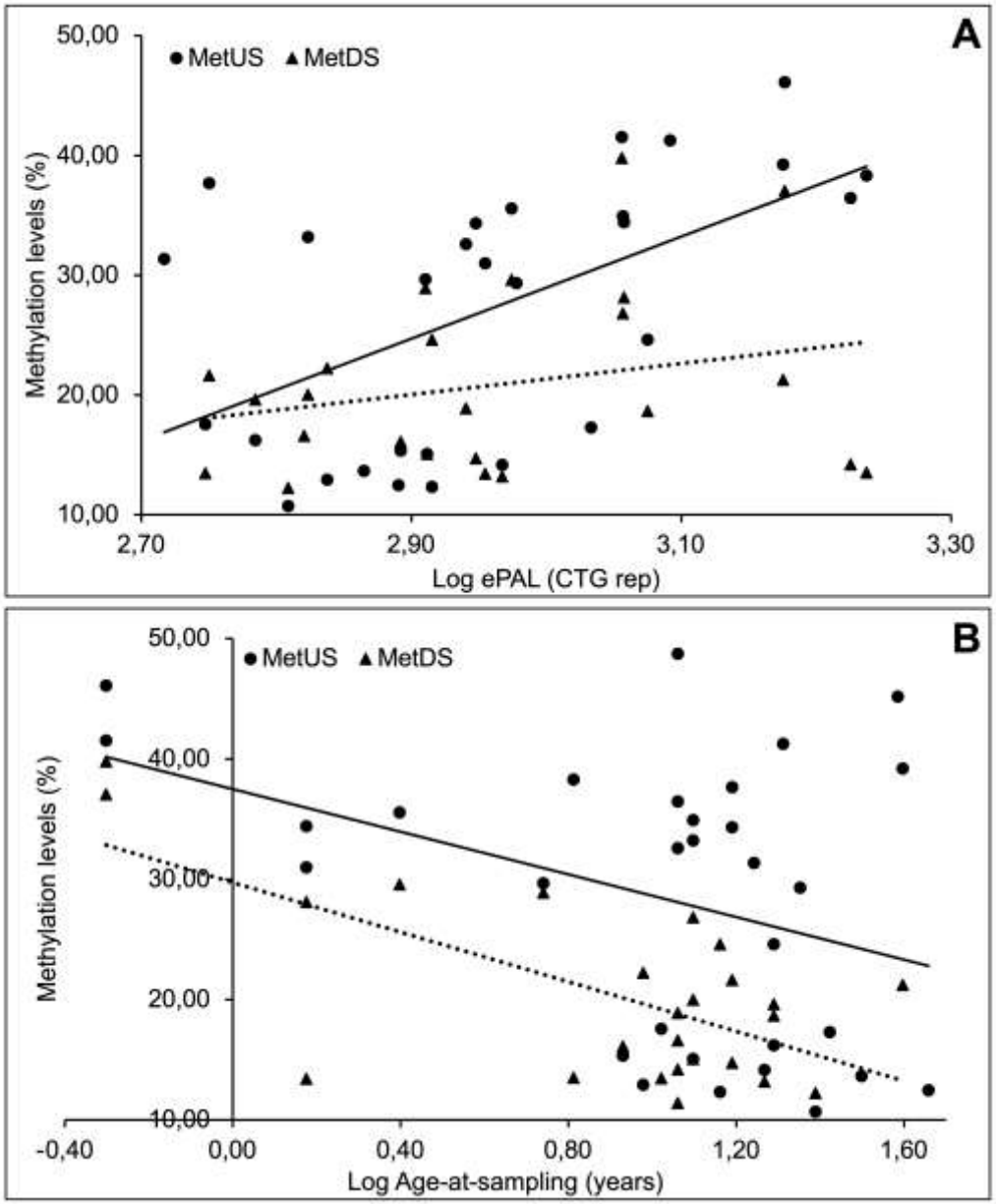

Figure 4. Pedigree of a Costa Rican DM1 family with rare paternal transmission of abnormal methylation flanking the CTG repeat expansion. The pedigree shows affecteds/carriers of the DM1 mutation (filled symbols; molecularly confirmed, or based on family history and clinical information) or non-affected family members (open symbols). Below each symbol appears the patient ID, the size of the ePAL and the coded methylation status: CTCF1 and CTCF2, unmethylated, ***CTCF1*** and ***CTCF2,*** methylated. In this family, only CR314 shows methylation flanking the CTG-repeat expansion, which was inherited from his affected father.

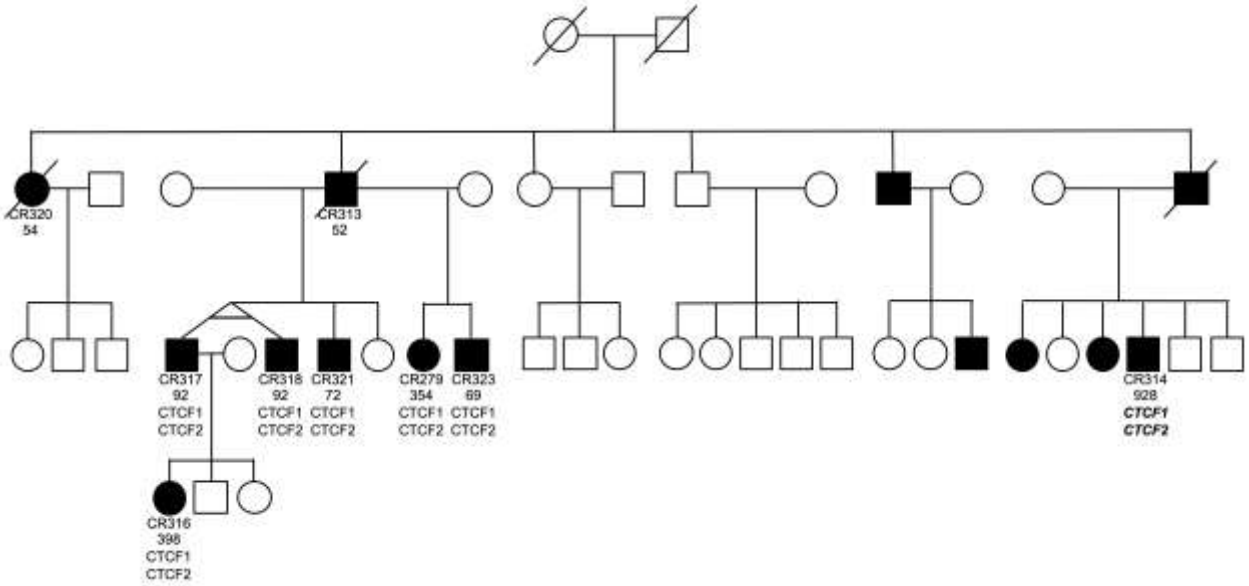

Table 1. Summary of the Costa Rican DM1 population analyzed.

Clinical sub-type	n Patients	Age range (years)	ePAL range (CTG repeats)	ePAL average (CTG repeats)	Percentile ePAL range (CTG repeats)*	% Methylation upstream	% Methylation downstream	Inheritance
Asymptomatic	16	4 – 73	50 – 261	99		0 (0/16)	0 (0/13)	5P, 3M, 8U
Late Onset	17	44 – 83	46 – 103	69	62 – 75	0 (0/17)	0 (0/14)	2P, 1M, 14U
Classical	140	13 – 64	66 – 1498	351	206 – 452	3.57 (5/140)	2.16 (3/139)	55P, 37M, 48U
Pediatric	30	8 – 55	241 – 1236	551	407 – 676	23.33 (7/30)	10.34 (3/29)	15P, 14M, 1U
Congenital	22	0.5 – 26	440 – 1727	959	688 – 1140	86.36 (19/22)	86.36 (19/22)	22M
Total	225	0.5 – 83	46 – 1727	406		13.78 (31/225)	11.52 (25/217)	77P, 77M, 71U

Inheritance: P, paternal; M, maternal; U, unknown.
Methylation downstream could not be measured in 3 asymptomatic, 3 late-onset, 1 classical and 1 pediatric case.
*Percentile shown are the 25th and 75th of the ePAL range for late-onset, classical, pediatric and congenital DM1 clinical sub-types.
Note, ePAL could not be measured in one classical and one congenital case.

Table 2. Logistic regression analyses carried out in this study to identify the modifiers of the DM1 clinical sub-types and CpG methylation flanking the CTG repeat

Model	Variables		OR	CI	p	N
LR1	ePAL vs. CDM/non-CDM		1.007	1.004 – 1.009	< 0.0001	223
LR2	ePAL vs. MetUS		1.015	1.009 – 1.021	< 0.0001	223
LR3	ePAL vs. MetDS		1.009	1.006 – 1.012	< 0.0001	216
LR4	IE/ePAL vs. MetUS	IE	0.990	0.982 – 0.999	= 0.028	46
		ePAL	1.029	1.010 – 1.050	= 0.004	46
LR5	IE/ePAL vs. MetDS	IE	0.996	0.991 – 1.001	= 0.109	44
		ePAL	1.009	1.003 – 1.015	= 0.002	44
LR6	Inheritance/ePAL vs. MetUS	Inheritance	0.013	0.000 – 0.438	= 0.0150	152
		ePAL	1.018	1.009 – 1.026	< 0.0001	152
LR7	Inheritance/ePAL vs. MetDS	Inheritance	0.089	0.011 – 0.743	= 0.0250	150
		ePAL	1.008	1.004 – 1.011	< 0.0001	150

LR = logistic regression; OR = odds ratio; CI = confidence intervals; N = samples size; ePAL = estimated progenitor allele length; CDM = congenital DM1; non-CDM = no congenital DM1; MetUS = methylation upstream; MetDS = methylation downstream; IE = intergenerational expansion.

Table 3. Multinomial regression (MNR) analysis in DM1 to determine differences on the ePAL between all different DM1 clinical sub-types.

Model	Clinical Sub-type	Reference	Variables	OR	CI	*p*-value	N
MNR1	Classical vs.	Late-Onset	Sex	10.7333	1.1955 – 96.3640	= 0.0341	207
			ePAL	1.1661	1.1350 – 1.1980	< 0.0001	207
	Pediatric vs.		Sex	23.7746	2.3319 – 242.3889	= 0.0075	207
			ePAL	1.1718	1.1410 – 1.2034	< 0.0001	207
	Congenital vs.		Sex	25.6823	2.1360 – 308.7880	= 0.0105	207
			ePAL	1.1767	1.1461 – 1.2081	< 0.0001	207
MNR2	Late-Onset vs.	Classical	Sex	0.0932	0.0105 – 0.8270	= 0.0331	207
			ePAL	0.8576	0.8363 – 0.8794	< 0.0001	207
	Pediatric vs.		Sex	2.2150	0.9145 – 5.3653	= 0.0781	207
			ePAL	1.0049	1.0027 – 1.0071	< 0.0001	207
	Congenital vs.		Sex	2.3927	0.6413 – 8.9281	= 0.1941	207
			ePAL	1.0091	1.0061 – 1.0121	< 0.0001	207
MNR3	Late-Onset vs.	Pediatric	Sex	0.0420	0.0043 – 0.4154	= 0.0067	207
			ePAL	0.8534	0.8300 – 0.8774	< 0.0001	207
	Classical vs.		Sex	0.4515	0.1869 – 1.0904	= 0.0771	207
			ePAL	0.9951	0.9930 – 0.9973	< 0.0001	207
	Congenital vs.		Sex	1.0803	0.2808 – 4.1566	= 0.9106	207
			ePAL	1.0042	1.0017 – 1.0067	= 0.0012	207
MNR4	Late-Onset vs.	Congenital	Sex	0.0389	0.0035 – 0.4393	= 0.0087	207
			ePAL	0.8498	0.8220 – 0.8786	= 0.0001	207
	Classical vs.		Sex	0.4179	0.1135 – 1.5387	= 0.1895	207
			ePAL	0.9910	0.9881 – 0.9939	< 0.0001	207
	Pediatric vs.		Sex	0.9257	0.2427 – 3.5311	= 0.9100	207
			ePAL	0.9958	0.9933 – 0.9983	< 0.0001	207

Each model compares all DM1 clinical sub-types against a reference. Therefore, four models (MNR 1-4) were tested. Each of three categories of clinical sub-type in the second column is compared with a reference clinical sub-type category in the third column. OR > 1 for ePAL in a clinical sub-type from the second column is an indicator of a greater ePAL in that clinical sub-type as compared to the clinical sub-type in the third column. Similarly, OR < 1 is an indicator of a smaller ePAL in the clinical sub-type of the second column as compared to the reference clinical sub-type.

Table 4. Multi-variate regression models of the relationship between the estimated progenitor allele length, age-at-sampling, sex, disease and upstream and downstream methylation.

Model	adjusted r^2	*p*-value	parameter		coefficient	standard error	*z*-statistic	*p*-value
model 1: MetUS = $\beta_0 + \beta_1$ Log(ePAL) + β_2 Log(age_s) + β_3 sex + β_4 Disease	0.606	< 0.001	intercept	β_0	0.38	4.76	0.1	= 0.936
			Log(ePAL)	β_1	4.90	1.32	3.7	< 0.001
n = 223 DM1 patients			Log(age_s)	β_2	-6.40	1.66	-3.9	< 0.001
			Sex	β_3	-0.22	0.80	-0.3	= 0.781
			Disease	β_4	16.69	1.77	9.5	< 0.001
model 2: MetDS = $\beta_0 + \beta_1$ Log(ePAL) + β_2 Log(age_s) + β_3 sex + β_4 Disease	0.677	< 0.001	intercept	β_0	8.71	2.80	3.1	= 0.002
			Log(ePAL)	β_1	2.31	0.79	2.9	= 0.004
N = 216 DM1 patients			Log(age_s)	β_2	-6.98	0.97	-7.2	< 0.001
			Sex	β_3	-0.27	0.47	-0.6	= 0.562
			Disease	β_4	9.29	1.03	9.0	< 0.001

The table shows the squared coefficient of correlation (r^2) and statistical significance (*p*-value) for each model, and the coefficient, standard error, *z*-statistic and statistical significance (*p*-value), associated with each parameter in the model. The number of individuals used in each analysis is indicated (N).