Individual-specific levels of CTG•CAG somatic instability are shared across multiple tissues in myotonic dystrophy type 1

Fernando Morales[1*], Eyleen Corrales[1], Melissa Vásquez[1], Baili Zhang[2], Huberth Fernández[3], Fernando Alvarado[3], Sergio Cortés[3], Carolina Santamaría-Ulloa[1], Marigold Myotonic Dystrophy Biomarkers Discovery Initiative-MMDBDI[4], Ralf Krahe[2] & Darren G. Monckton[5]

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

[2]Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, Texas, 77030-4009, USA

[3]Hospital Calderón Guardia/Escuela de Medicina, Universidad de Costa Rica, San José, 2060, Costa Rica

[4]Marigold Myotonic Dystrophy Biomarkers Discovery Initiative (MMDBDI). See Supplementary Table S3 for the full list of investigators.

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

***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: 00 506 2511 2124

Fax: 00 506 2511 5130

Email: fernando.moralesmontero@ucr.ac.cr

Abstract

Myotonic dystrophy type 1 is a complex disease caused by a genetically unstable CTG repeat expansion in the 3'-untranslated region of the *DMPK* gene. Age-dependent, tissue-specific somatic instability has confounded genotype-phenotype associations, but growing evidence suggests that it also contributes directly toward disease progression. Using a well-characterized clinical cohort of DM1 patients from Costa Rica, we quantified somatic instability in blood, buccal cells, skin and skeletal muscle. Whilst skeletal muscle showed the largest expansions, modal allele lengths in skin were also very large and frequently exceeded 2,000 CTG repeats. Similarly, the degree of somatic expansion in blood, muscle and skin were associated with each other. Notably, we found that the degree of somatic expansion in skin was highly predictive of that in skeletal muscle. More importantly, we established that individuals whose repeat expanded more rapidly than expected in one tissue (after correction for progenitor allele length and age), also expanded more rapidly than expected in other tissues. We also provide evidence suggesting that individuals

in whom the repeat expanded more rapidly than expected in skeletal muscle, have an earlier age-at-onset than expected (after correction for the progenitor allele length). Pyrosequencing analyses of the genomic DNA flanking the CTG repeat revealed that the degree of methylation in muscle was well predicted by the muscle modal allele length and age, but that neither methylation of the flanking DNA, nor levels of *DMPK* sense and anti-sense transcripts, could obviously explain individual- or tissue-specific patterns of somatic instability.

Introduction

Myotonic dystrophy type 1 (DM1) (OMIM #160900) is a complex disorder that affects many tissues across the body including the brain, eyes, endocrine system, heart and gastrointestinal tract (1). However, the disorder is so named because of the two most prominent features: myotonia and muscular dystrophy (2). Indeed, DM1 is the most common hereditary muscular dystrophy in adults (2). DM1 is a progressive disease, clinically highly variable (within and between families), with age-at-onset between birth and more than 80 years old. It is transmitted as an autosomal dominant disease with marked anticipation of 20 to 30 years per generation (*i.e.*, decrease in the age-at-onset in successive generations in the same family) (3-5).

The DM1 mutation was identified as an expansion of an unstable CTG repeat localized in the 3'-untranslated region (UTR) of the DM1 protein kinase (*DMPK*) gene and in the promoter region of the SIX homeobox 5 (*SIX5*) gene (OMIM #600963) (6-11). The CTG tract is polymorphic in length in the general population, ranging from 5 – 37 CTG repeats (g.17294_17296(5_37)), while in asymptomatic carriers or affected patients, the number of CTG repeats inherited ranges from approximately 50 to more than 2,000. The number of repeats shows a positive

correlation with the severity of the disease, and an inverse correlation with the age-at-onset. Once the repeat reaches the threshold of expansion (~40 CTG repeats), it becomes highly unstable both in the soma and germline (3, 12-17).

Because of the large size of many of the CTG expansion mutations responsible for DM1, Southern blot hybridization of restriction digested genomic blood DNA has long been the gold standard for the molecular diagnosis of the disease. However, due to the somatic instability of the repeat, the expanded allele is usually detected as a broad smear instead of a discreet band (7, 9, 14, 15, 18). More detailed analysis of somatic instability in the blood DNA of DM1 patients has revealed that somatic instability is allele-length and age-dependent, and highly biased towards expansion (19-21). Further evidence of somatic instability of expanded DM1 alleles is provided by the observation of repeat size variability between tissues in the same individual (22, 23). In particular, skeletal muscle usually shows repeat expansions that are much larger than those observed in blood DNA extracted from peripheral blood leukocytes (PBL) (24-26). It has also been shown that in younger individuals the difference in size between the CTG expansions in muscle and blood are smaller, suggesting that somatic expansion in skeletal muscle is also age-dependent (27, 28). Although the significance of somatic expansion in disease progression remained unclear for many years, by controlling for the confounding effects of somatic instability on genotype to phenotype correlations, we have been able to provide evidence that somatic instability contributes to DM1 disease severity, that is individuals with greater rates of somatic expansion in blood DNA develop symptoms earlier than expected and vice versa (16, 29-32). However, these associations were revealed in blood DNA and it is assumed that somatic expansion in blood is not directly causative, but rather mirrors a parallel process in affected tissues. Here, we sought to test the hypothesis that individual-specific differences in tissue-specific somatic instability are associated with each other and with individual-specific differences in disease severity.

Results

Muscle weakness and muscle pathology directly correlate with age and the size of the estimated progenitor allele length (ePAL)

In order to investigate the drivers of muscle weakness and muscle damage in DM1, we determined muscle weakness using a modified muscular impairment rating scale (MIRS) and assessed muscle damage by histopathological assessment in 35 DM1 patients (Supplementary Table S1). Not surprisingly, muscle weakness and muscle damage were positively associated with each other (Pearson $r^2 = 0.41$, $p < 0.001$, $n = 35$, Supplementary Figure S1). As expected (16, 33), both muscle weakness and muscle damage were associated with age, the size of the estimated progenitor allele length (ePAL – expanded allele size transmitted from the affected parent to their offspring), and an interaction between them (Table 1, for muscle weakness, model 1, $r^2 = 0.30$, $p = 0.005$ and for muscle damage, model 2, $r^2 = 0.23$, $p = 0.018$, $n = 35$). Given the known associations between some progressive DM1 phenotypes and sex (34), we also included sex as an independent variable and used stepwise model selection. These analyses revealed that males had significantly worse muscle damage than females (Table 1, model 3, $r^2 = 0.34$, $p = 0.004$, $n = 35$), although there was no detectable effect of sex on muscle weakness in this cohort (data not shown).

Skeletal muscle presents with the largest expansion size

In order to investigate tissue-specific patterns of somatic instability of the *DMPK* CTG repeat and its association with clinical parameters, we obtained DNA from blood, saliva, skin and skeletal muscle. We were able to size the modal expanded allele length in blood for all 35 DM1 patients, in buccal cells for 20 DM1 patients, in skin for 26 DM1 patients, and in skeletal muscle for 19 DM1 patients (Supplementary Table S1, Figure 1A and 1B). In pairwise comparisons, as expected, skeletal muscle showed the largest CTG expansion size (Supplementary Table S1), significantly larger than in all other tissues (Table 2). The CTG expansion size in skin, was intermediate between, and significantly different from,

muscle and blood/buccal cells, whereas blood and buccal cells showed similar expansion sizes (Table 2). Notably, participants with ePAL < ~ 100 CTG repeats did not show major modal allele differences between tissues (Supplementary Figure S2).

In order to extend these observations and better understand the factors driving somatic expansion in skeletal muscle, we compared these data with ePAL, and blood and skeletal muscle (quadriceps) modal allele lengths available for 17 participants from the *Dystrophia Myotonica* Biomarker Discovery Initiative (DMBDI) (35). There were no statistically significant differences in modal allele sizes between the two different skeletal muscles sources ($p = 0.790$), or in blood modal allele sizes between the two different populations ($p = 0.890$), or between the ePALs ($p = 0.400$) (Table 2). Thus, we combined the Costa Rican and DMBDI blood and skeletal muscle data for the analyses of somatic expansion described below.

Individual-specific somatic expansion measures in skeletal muscle, skin and blood are associated with each other

In order to quantify the degree of somatic expansion in blood DNA, we measured the modal increase in repeat length in blood during the lifetime of the individual (*i.e.*, modal repeat length in blood - ePAL) as previously described (31, 32). We also applied the same approach to quantifying the degree of somatic expansion in skeletal muscle and skin DNA (*i.e.*, modal repeat length in muscle/skin - ePAL). Consistent with the notion that similar factors drive somatic expansion in multiple tissues, the degree of somatic expansion in blood and muscle, blood and skin, and most strikingly, skin and muscle were highly correlated with each other ($p < 0.001$, Figure 2A, B, C, Table 3).

Previous studies using much larger datasets have demonstrated that the degree of somatic expansion in blood DNA is well predicted by the ePAL and age-at-sampling, and an interaction between them (16, 31, 36). Similarly, we found these variables also well predicted the degree of somatic expansion in blood (Table 4, model 1, $r^2 = 0.55$, $p < 0.001$, $n = 50$), skeletal muscle (Table 4, model 2, $r^2 = 0.54$, $p < 0.001$, $n = 35$) and

skin (Table 4, model 3, $r^2 = 0.73$, $p < 0.001$, $n = 26$) in this cohort. Most notably, however, individual-specific residual variation in the degree of somatic expansion in muscle, skin and blood DNA, after correcting both for ePAL, age-at-sampling, and an interaction between them (*i.e.*, the residuals from Table 4, models 1 – 3), were directly correlated with each other (*i.e.*, individuals in who the repeat expands more rapidly than expected in one tissue, also expanded their repeat more rapidly than expected in other tissues) (Figure 2D, E, F, Table 3).

Individual-specific somatic instability in skeletal muscle is associated with variation in age-at-onset of DM1

Previously, we demonstrated that the individual-specific degree of somatic expansion in blood DNA (after correction for ePAL, age-at-sampling, and an interaction between them) was inversely correlated with individual specific variation in age-at-onset (after correction for ePAL) (*i.e.*, individuals in who the repeat expands more rapidly in blood than expected, have an onset earlier than expected) (16, 31). Here we observed a similar trend, albeit not of nominal significance, likely due to the relatively small sample size (Figure 2G, Table 3). Notably, this individual-specific degree of somatic expansion in muscle DNA (after correction for ePAL, age-at-sampling, and an interaction between them) was also inversely correlated with individual specific variation in age-at-onset (after correction for ePAL) and the degree of variance in age-at-onset explained was greater and of nominal statistical significance (Figure 2H, Table 3). Despite the relatively small sample size, a similar trend was observed for residual variation in somatic expansion in skin (Figure 2I, Table 3).

DM1 tissue-specific DNA methylation upstream of *DMPK*

Previous methylation studies in DM1 have shown that altered methylation levels surrounding the CTG repeat expansion are primarily restricted to congenital cases in blood DNA (37-39), although variable levels of altered methylation have also been reported in other somatic tissues and in the male germline of DM1 patients (36, 40, 41). In order to provide further insight into tissue-specific methylation levels in DM1, we performed Pyrosequencing-based Methylation Analysis (PMA) to quantify methylation levels upstream of the CTG repeat expansion, the primary site off differential methylation in DM1 (37, 40). In the 35 Costa-Rican (CR) DM1 patients (only the CR samples were analyzed by PMA), we observed hypermethylation in only three patients (a sample was considered to be hypermethylated when methylation levels were >10% (36, 42)), two congenital cases and one pediatric case (with age-at-onset < 4 years). For these patients, hypermethylation was only observed in blood and buccal cells (except for the pediatric case, which did not reach the hypermethylation threshold in buccal cells). The hypermethylation threshold was not exceeded in skin or skeletal muscle in any of the 35 CR DM1 samples (Supplementary Figures S3 and S4, Supplementary Table S1). However, the median degree of methylation in muscle DNA samples (2.42%) was notably higher than in the blood (0.83%), buccal cells (0.97%) and skin (1.27%) (Friedman rank sum test, chi-squared = 22, $p < 0.001$; Supplementary Figure S4). Whilst the methylation levels observed in blood, buccal cells and skin DNA were nominally well predicted by CTG length (either ePAL or modal repeat length in the relevant tissue), age at sampling, and an interaction between them, these associations were driven exclusively by the three individuals with the longest ePALs (> 600 CTG repeats; data not shown). In skeletal muscle however, even when excluding individuals with the largest ePALs (> 600 CTG repeats), the degree of methylation was well predicted by ePAL, age-at-sampling, and an ePAL-age at sampling interaction (Supplementary Table S2, model 1, $r^2 = 0.30$, $p = 0.005$, $n = 31$), and apparently even more accurately by the modal allele length in muscle (Supplementary Table S2, model 2, $r^2 = 0.64$, $p = 0.0009$, $n = 17$).

***DMPK* sense and anti-sense transcription are higher in muscle**

Previously, it was reported that bidirectional transcription of the *DMPK* gene promotes instability of the CTG•CAG repeat in cultured fibroblasts (43). We therefore sought to quantify the tissue-specific patterns of both sense and anti-sense transcription of the *DMPK* gene. In skin and buccal cells, we were unable to detect the anti-sense *DMPK* transcript, while the sense transcript was detected in only a few patients. In contrast, in both blood and muscle we were able to detect both transcripts in most of the DM1 patients enrolled in this study (Supplementary Figure S5). As expected, the levels of the sense (paired *t*-test, $p < 0.001$, $n = 32$) and anti-sense (paired *t*-test, $p < 0.001$, $n = 32$) transcripts were higher in muscle than in blood. The sense / anti-sense ratio was also higher in muscle than blood (paired *t*-test, $p < 0.001$, $n = 32$). The level of the anti-sense transcript was predicted by the level of the sense transcript in both blood (Table 5, model 1, $r^2 = 0.20$, $p = 0.005$, $n = 34$) and muscle (Table 5, model 2, $r^2 = 0.49$, $p < 0.001$, $n = 34$). However, there were no detectable associations in levels of the sense, anti-sense, or sense / anti-sense ratio between blood and muscle (data not shown). Likewise, there were no detectable associations in levels of the sense, anti-sense, or sense / anti-sense ratio between ePAL, modal CTG or age in either blood or muscle (data not shown). Similarly, there were no detectable associations in levels of the sense, anti-sense, or sense / anti-sense ratio between flanking methylation levels in blood (data not shown). However, there was an apparent weak inverse association between flanking DNA methylation levels in muscle for the anti-sense (Table 5, model 3, $r^2 = 0.088$, $p = 0.055$, $n = 32$), and an apparent weak positive association for sense / anti-sense ratio (Table 5, model 4, $r^2 = 0.067$, $p = 0.083$, $n = 32$). There were no detectable associations for the sense transcript (Table 5, model 5, $r^2 = 0.004$, $p = 0.30$, $n = 32$). And there were no detectable associations between residual somatic instability in muscle or blood, and the sense, anti-sense, or sense / anti-sense ratio (data not shown).

Discussion

Somatic expansion in blood mirrors somatic expansion in skeletal muscle

The modal allele length (measured in blood by Southern blot analysis of restriction digested genomic DNA), typically explains less than 50% of the variation in age-at-onset. Even poorer correlations are usually observed when using the modal allele length estimated from muscle, which is the primarily affected tissue in DM1 patients (3, 14, 15, 27, 28, 44-46). More recently, we have been able to control for the confounding factors of somatic instability and provided more informative genotype-phenotype correlations by using small-pool PCR to estimate the inherited progenitor allele size (ePAL)(16, 30, 31). Additionally, we have established that the ePAL, in an interaction with age, is also the primary driver of the degree of somatic expansion in blood DNA (16, 29-32). Critically, whilst ePAL has been revealed as the primary determinant of disease severity, we have also revealed that individual-specific variation in disease severity (after correction for ePAL and age) is inversely correlated with individual-specific variation in the degree of somatic expansion in blood DNA (*i.e.*, patients in whom the repeat expands more rapidly than expected, have more severe symptoms than expected) (16, 29-31). The major caveat with these studies, is that measures of somatic expansion have been derived from peripheral blood DNA. Given that the hematopoietic system does not obviously contribute to the primary DM1 clinical manifestations, we have inferred it is more likely that individual-specific somatic instability in blood mirrors a parallel pattern of individual-specific somatic instability in primary affected tissues, such as muscle. In order to test this hypothesis, we analyzed tissue-specific somatic instability in DM1 patients.

By comparing the CTG expansion size among different tissues from the same DM1 patient, we observed, as expected, a much larger allele size in skeletal muscle than in blood, similar to what has been reported in other studies (24, 26-28). Skin also showed a very large expansion size,

as has been previously reported (27). Although the expansion size in skin was not as large as it was in skeletal muscle, it was nonetheless very large averaging around 2,500 repeats, some 2,000 repeats larger than observed in blood DNA. In addition, we observed that the modal allele length was very similar in blood and buccal cells, as previously reported (36). The factors that drive these tissue-specific differences on mutational dynamics remain unknown.

Interestingly, we observed, that all of the patients with ePALs <100 CTG repeats showed similar (in tibialis anterior (TA)) or very similar (in quadriceps) modal allele lengths in blood and muscle, which is consistent with the very mild muscle phenotype in most of these DM1 patients. However, all patients with ePALs >100 CTG repeats, showed a much larger allele size in muscle with a more severe muscle phenotype (Supplementary Table S1). These data suggest that tissue-specific mutational dynamics are at least partially modified by inherited repeat size. Additional factors that likely contribute to explain the tissue-specific mutational dynamics are *trans*-acting factors, such as proteins of the DNA mismatch repair system (MMR), known to be involved in the expansion mechanism. Several studies in mice and cell models have been carried out in order to provide insights into the relationships between MMR products (at the RNA or protein levels) and CTG•CAG instability (47-50). But thus far, a clear relationship between MMR levels and instability has not yet been established.

We also demonstrated that the degree of somatic expansion in blood was positively associated with the degree of somatic expansion in skeletal muscle and skin. Even more importantly, however, we demonstrated that after correction for ePAL and age, the individual-specific degree of somatic expansion in blood was positively associated with the individual-specific degree of somatic expansion in skeletal muscle (Figure 2D, $r^2 = 0.46$, $p < 0.001$). These data confirm that the individual-specific degree of somatic expansion in blood does indeed at least partially mirror the individual-specific degree of somatic expansion in skeletal muscle, a key target of the disease pathology. Our data thus provide a rational

explanation for the observed associations between the individual-specific degree of somatic expansion in blood and variation in disease severity previously reported and variation in the *MSH3* MMR gene (16, 29-31, 51).

In further support of these hypotheses, we also provide the first evidence that individual specific variation in age-at-onset (after correction for ePAL) is better predicted by the individual-specific degree of somatic expansion in skeletal muscle DNA (Figure 2H, $r^2 = 0.10$, $p = 0.041$, $n = 32$), than in blood (Figure 2G, $r^2 = 0.03$, $p = 0.131$, $n = 46$). Similar results have been reported in post-mortem cortical DNA samples of Huntington disease (HD), where patients with repeat length distributions skewed towards larger alleles showed an extreme early onset of the disease (52). These results reinforce the concept that somatic expansion is an important modifier of the disease process and therapeutic target in DM1 and other microsatellite repeat expansion disorders (53-55). Interestingly, naphthyridine-azaquinolone, a small molecule compound, was recently shown to induce contractions of the expanded CTG•CAG repeat in HD patient cells and a mouse model (56), an effect that if replicated in DM1 patients would be expected to delay, or potentially even reverse, progression of the disease. An important aspect of a repeat-expansion modulating drug discovery program would be the development of biomarkers capable of reporting on-target engagement and efficacy at the level of somatic instability (55). Ideally, this would be assessed in tissues that are the relevant targets for disease, such as skeletal muscle and brain in DM1, and brain in HD.

Whilst sampling cell-free DNA and/or exosomes from cerebrospinal fluid (CSF) might possibly provide a route toward sampling small amounts of DNA(29), the brain otherwise remains inaccessible to direct sampling of DNA. Although skeletal muscle is nominally more accessible to direct sampling, generating sufficient DNA to size the very large expansions present in this tissue in DM1 still represents a major challenge, especially if using needle biopsies, as is likely in a large clinical trial (57). Blood and/or buccal cell DNA offer possible alternatives, but the repeat dynamics in these tissues do not exhibit the very large somatic expansions observed in skeletal muscle. To this end, our observation here that the

size of somatic expansions in skin are similarly very large, and that they are highly predictive of the degree of somatic expansion in muscle, suggests that skin DNA might represent a more readily accessible surrogate tissue and biomarker for therapies aimed at modifying somatic expansion in DM1 patients.

CTG flanking methylation in muscle is predicted by the ePAL and age

In order to investigate *cis*-acting factors that might contribute toward the tissue-specific somatic instability seen in DM1, we analyzed DNA methylation upstream of the CTG repeat and levels of both sense and anti-sense *DMPK* transcripts. Early studies showed that hypermethylation (in blood DNA) was confined only to congenital cases (38, 39). However, it was subsequently shown that hypermethylation was also present in some DM1 adult-onset cases, in several tissues of the same individual (40). In our study, overt hypermethylation (>10%) was restricted to congenital or pediatric onset cases (three out of 35 DM1 patients) inheriting the largest alleles in our cohort (>600 CTG repeats). In these three patients, overt hypermethylation was confined to blood and buccal cell DNA, although methylation levels were also higher in skin and muscle DNA. Notably, after excluding these three patients, we did not detect any association between the absolute degree of methylation and ePAL and/or age-at-sampling in blood, skin or buccal cells in this cohort (32 DM1 patients). Interestingly, although still remaining relatively low (< 10%), the absolute amount of methylation observed in muscle DNA was higher than the other tissues, and was well predicted by ePAL, age-at-sampling, and an ePAL-age at sampling interaction (Supplementary Table S2, model 1). Moreover, methylation levels in muscle appeared to be even better predicted by the modal CTG length in muscle (Supplementary Table S2, model 2).

These data are consistent with the observations by Lopez Castel *et al.* (40) who found hypermethylation in multiple tissues of three CDM cases, an absence of hypermethylation in the skeletal muscle of adult-onset cases, but detectable hypermethylation in heart, cortex and/or liver of

some patients with very large acquired somatic expansions >4,000 repeats. These data suggest that the primary determinant of methylation levels is the size of the inherited CTG, with a threshold of ~600 CTG repeats above which increased tissue-specific methylation levels are established during early development. These putative early methylation marks appear to be higher in blood and buccal cell lineages, and lower in skin and muscle, possibly as a function of a protective effect of higher levels of *DMPK* expression Despite the fact that in some patients, many cells acquire somatic expansions > 600 in both blood and buccal cell DNA, and that in nearly all patients' modal allele sizes in skin are > 2,000 repeats, methylation levels are not subsequently altered. In muscle however, (and likely in some other issues (40)) where modal sizes frequently exceed 3,000 repeats, additional age and allele length-dependent changes in methylation levels were observed. These data thus suggest that there may be a lower threshold for the induction of methylation during early development (~600 CTG repeats), and a second much higher threshold of >>3,000 repeats that is possibly tissue-dependent (40), for methylation increases acquired postnatally. Overall, our data suggest tissue-specific patterns of repeat instability likely drive subtle tissue-specific patterns of methylation, rather than the other way around. Ideally, these observations should be confirmed by longitudinal analyses in a much larger sample of participants with coverage of additional regions surrounding the *DMPK* locus using more sensitive methods.

Not detectable relationship between CTG somatic expansion and *DMPK* gene expression in this study

Using fibroblasts transfected with a CTG repeat expansion, it was previously demonstrated that somatic instability could be promoted by transcription, and further enhanced by bidirectional transcription of the *DMPK* CTG repeat (43). Notably, the anti-sense *DMPK* transcript (*DM1-AS*) shows a complex pattern of expression with use of multiple alternative transcription starts, splice sites and polyadenylation signals (58). By analyzing a region (amplicon F described in Figure 3A of Gudde et. al. 2017 (58)) that detects the majority of *DM1-AS* transcripts, and in

particular all those that contain the expanded CAG repeat, we carried out a small pilot study to investigate if there were any obvious relationships between *DMPK* and *DM1-AS* expression levels and somatic instability in muscle, skin, blood and buccal cells. Here, we found much higher levels of *DMPK* sense transcripts in muscle relative to blood as previously reported (59). Levels of the anti-sense transcript were also higher in muscle than blood, but more moderately so, such that the relative anti-sense / sense ratio was much higher in blood than muscle. These data could be consistent with an association between absolute levels of either the sense / anti-sense / bidirectional transcription and somatic instability, but suggest that the ratio itself is not a key determinant. However, we detected no obvious associations between residual somatic instability in muscle or blood, and the sense, anti-sense, or sense / anti-sense ratio, albeit with a relatively small sample size. Interestingly, we did detect an apparent inverse association between flanking methylation levels in muscle for the anti-sense (Table 5, model 3, $r^2 = 0.088$, $p = 0.055$, $n = 32$), but not for the sense transcript. Larger sample sizes, a deeper analysis of the different *DM1-AS* isoforms and analysis of more tissues will be required to further evaluate possible intersections between DNA methylation, gene transcription and somatic repeat instability.

In conclusion, our results show that the ePAL can be used as a predictor of the allele size in other tissues such as skin and muscle and that individual-specific patterns of somatic instability are at least partially shared across multiple tissues. We also provide the first evidence that individual-specific skeletal muscle-specific patterns of somatic instability are associated with variation in disease severity in DM1.

Materials and Methods

Patient population. Thirty-five Costa Rican (CR) DM1 patients (17 females and 18 males) were invited to participate in this study. According to their clinical manifestations, patients belong to all different DM1 clinical sub-types: three asymptomatic, two late-onset, 24 classical, four pediatric, and two congenital (CDM) cases. For all DM1 patients enrolled in this study, four tissue samples were simultaneously collected at the

Calderón Guardia Hospital; blood, saliva, skin and skeletal muscle (tibialis anterior, TA). Blood samples (10 ml) were drawn by phlebotomy into EDTA-containing vacutainer tubes. As reported previously (36), saliva DNA enriched for buccal epithelium cells was collected by asking patients to wipe their inner cheeks with their tongue and spit about 5 ml of saliva in to a collection tube. Skeletal muscle and skin samples were obtained by open biopsy. Skin and muscle samples were snap-frozen and stored in liquid nitrogen once they were obtained. Blood and saliva samples were processed immediately after collection. For all samples, we registered the age of the patient at the moment of sampling. The age-at-onset (in the symptomatic patients) and the size of the estimated progenitor allele length (ePAL – expanded allele size inherited by the offspring from the parent) had been previously collected by analysis of blood DNA (16, 29). This study was approved by the Ethics and Scientific Committees of the Calderón Guardia Hospital and the Universidad de Costa Rica. All samples were collected after obtaining informed consent. In addition to these Costa Rican samples, we used some data from an independent study (the Marigold Myotonic Dystrophy Biomarkers Discovery Initiative Study) that collected similar DNA data, but used a different skeletal muscle source (quadriceps from 17 DM1 patients) (35).

Neurologic and pathologic examination. Neurologic examination included measurement of muscle weakness (of the TA muscle) by the ankle dorsiflexion (ADF) test using standardized manual muscle testing. Manual testing was scored on the modified muscular impairment rating scale (MIRS) grades, ranging from 1 to 4 depending on the muscle weakness (*i.e.*, 1 = no weakness, and 4 = severe muscle weakness) (60). In order to avoid operator bias, the same neurologist carried out the test in all patients. For the pathological study of the skeletal muscle, we focused on the presence of ring fibers, adipose tissue, fibrous tissue, internalized nuclei and fiber atrophy. In order to identify these muscle features, once samples were obtained, they were fixed in 10% buffered formalin and transversal and longitudinal cuts of the muscle tissue were made. Subsequently, all samples were imbedded in paraffin and stained with hematoxylin-eosin, periodic acid–Schiff (PAS) and modified trichrome of Gomori, following standard procedures. A Canon EOS Rebel T3i camera, with an adapter for an Olympus microscope, was used to take

photographs of the samples. DM1 patients were classified from 1 to 4 depending on the severity of the muscle damage as ascertained by the histopathological appearance (*i.e.*, 1 = no muscle damage, and 4 = severe muscle damage). Clinical data from each DM1 patient can be found in Supplementary Table S1.

DNA and RNA purification. Prior to DNA and RNA purification from skin and skeletal muscle samples, the tissues (about 20 mg) were macerated on a bullet blender (Next Advance, USA) using zirconium beads (Next Advance, USA). The DNA from all samples was isolated following standard proteinase K/phenol-chloroform extraction procedures. Total RNA from all samples was isolated following a standard TRIzol extraction procedure. Once the RNA was purified, its integrity was verified on an agarose gel stained with ethidium bromide. All of the DNA and RNA samples were quantified by optical density at 260 nm in a NanoDrop spectrophotometer (Thermo Scientific, USA). DNA samples were stored at -20°C and RNA samples at -80°C.

Genomic Southern blot hybridization. Genomic DNA samples (5 µg, where available) were digested overnight with the BglI restriction enzyme (New England Biolabs) using the recommended conditions. Digested products were resolved by electrophoresis at 4°C on 0.8% agarose gels, transferred to a Hybond-N membrane (GE Healthcare, Argentina) and hybridized to the p5B1.4 probe (61) labelled with P^{32}-dCTP using random primers (Prime-a-gene, Promega, USA). The membranes were exposed to X-ray film for five days at -80°C and autoradiographs were developed following standard procedures. The mid-point of the smear, or the point with the highest intensity on the smear, was used to measure the modal allele length. In those samples where the expanded allele did not show a smear but a discreet band, the mid-point of the band was measured. Both the non-disease associated allele (3.4 kb) and the expanded modal allele (>3.4 kb) in each tissue were sized using UVIbandmap software (UVITEC, UK). Data on the modal CTG expansion size for each tissue for each patient can be found in Supplementary Table S1. All blood DNA samples were previously tested for the presence of AciI sensitive CCG/CGG variant repeats (29, 36, 62, 63) and none were detected.

Methylation analysis upstream of the CTG expansion. DNA methylation status and levels upstream of the CTG repeat expansion were determined in all four tissues from all CR DM1 patients by Pyrosequencing-based methylation analysis (PMA) as previously performed (29, 36, 63). In all, we interrogated 11 CpGs upstream of the CTG repeat (seven CpGs located within the CTCF1-binding site) using previously described procedures (29, 36, 63). PyroMark CpG 1.0.11 software (Qiagen) was used to analyze the data: mean methylation percentage (mC/(mC+C)) for each CpG site and the methylation index (MI) (average value of mC/(mC+C) upstream (11 CpGs)). Samples were identified as hypermethylated when mean methylation levels were > 10%. Methylation data upstream of the CTG repeat for each tissue for each DM1 patient can be found in Supplementary Table S1.

RT-PCR for the DMPK sense and anti-sense transcripts. RT-PCR analyses were carried out following the strategy, primers and conditions described by Cho *et al*. (64) and Michel *et al.* (65). Briefly, 250 ng of total RNA were treated with DNAseI (Thermo Scientific, USA) following standard conditions. Three independent RT reactions were carried out using either oligo $(dT)_{18}$ or strand-specific primers for the sense and anti-sense *DMPK* transcripts using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA) following the manufacturer's instructions. cDNA was treated with RNaseA (Thermo Scientific, USA) following standard conditions. Subsequent PCRs to analyze the internal control (beta-2-microglobulin (*B2M*) gene (66, 67)), were performed using the sense and anti-sense *DPMK* transcripts with Taq DNA polymerase (Sigma), using the previously described DM1-S and DM1-AS primers (65) and for *B2M*, we used 5'-TGCTGTCTCCATGTTTGATGTATCT-3' forward and 5'-TCTCTGCTCCCCACCTCTAAGT-3' reverse primers. PCR conditions were optimized for each transcript, the identities of which were confirmed by DNA sequencing. Products were resolved in ethidium bromide agarose gels and densitometric analysis was performed using UVIbandmap software (UVITEC, UK). Signals for the sense and anti-sense *DMPK* transcripts were normalized to *B2M* levels. All analyses were carried out in duplicate. *DMPK* expression data for blood and muscle for each patient can be found in Supplementary Table S1.

Statistical analyses. Continuous characteristics such as muscle weakness, the degree of somatic expansion, or the degree of methylation in muscle DNA were modeled as dependent variables in multivariate or simple regression models. ePAL, age, sex, or the muscle modal allele length were included as independent variables of multivariate or simple regression models. Model selection was based on the Akaike information criterion (AIC) for each model using a back-wards stepwise selection procedure implemented using the step function in R (68). The Wilcoxon, Friedman or the Mann-Whitney U rank sum test were used to compare the median of two or more variables, and Pearson's correlations to study the relationship between two variables. The paired *t*-test was used to compare the means of two variables. Significance level was set at $p = 0.05$. All the statistically significant results obtained throughout the study remained significant after multiple comparison correction (Bonferroni, $p < 0.05$).

Acknowledgements

The authors would like to thank the myotonic dystrophy patients that agreed to participate in this project. We also want to that Mr. William Araya for his technical support. Association Française contre les Myopathies (AFM) (http://www.afm-telethon.com) (AFM18238) and the Universidad de Costa Rica (UCR) (742-B5-300) and the Marigold Foundation funded this research project.

Conflicts of Interest Statement

Within the last five years D.G.M. has been a scientific consultant and/or received an honorarium or stock options from AMO Pharma, Charles River, Vertex Pharmaceuticals, Triplet Therapeutics, and LoQus23, 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. Smith, C.A. and Gutmann, L. (2016) Myotonic Dystrophy Type 1 Management and Therapeutics. *Curr. Treat. Options Neurol.*, **18**, 52.
3. 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.
4. 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.
5. Höweler, C.J., Busch, H.F.M., Geraedts, J.P.M., Niermeijer, M.F. and Staal, A. (1989) Anticipation in myotonic dystrophy: fact or fiction? *Brain*, **112**, 779–797.
6. Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N.G.M., Vooijs, M. *et al.* (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature*, **355**, 548-551.
7. 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.
8. Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B. *et al.* (1992) Detection Of an Unstable Fragment Of DNA Specific to Individuals With Myotonic Dystrophy. *Nature*, **355**, 547-548.
9. 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.
10. 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.
11. Boucher, C.A., King, S.K., Carey, N., Krahe, R., Winchester, C.L., Rahman, S., Creavin, T., Meghji, P., Bailey, M.E.S., Chartier, F.L. *et al.* (1995) A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable $(CTG)_n$ repeat. *Hum. Mol. Genet.*, **4**, 1919–1925.
12. Ashizawa, T., Dunne, P.W., Ward, P.A., Seltzer, W.K. and Richards, C.S. (1994) Effects Of the Sex Of Myotonic Dystrophy Patients On the Unstable Triplet Repeat In Their Affected Offspring. *Neurology*, **44**, 120-122.
13. Brunner, H.G., Bruggenwirth, H.T., Nillesen, W., Jansen, G., Hamel, B.C.J., Hoppe, R.L.E., de Die, C.E.M., Howeler, C.J., van Oost, B.A., Wieringa, B. *et al.* (1993) 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.*, **53**, 1016-1023.
14. 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.
15. 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.

16. 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.
17. 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.
18. Harley, H.G., Rundle, S.A., Reardon, W., Myring, J., Crow, S., Brook, J.D., Harper, P.S. and Shaw, D.J. (1992) Unstable DNA-Sequence In Myotonic Dystrophy. *Lancet*, **339**, 1125-1128.
19. 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.
20. 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.
21. 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.
22. Ishii, S., Nishio, T., Sunohara, N., Yoshihara, T., Takemura, K., Hikiji, K., Tsujino, S. and Sakuragawa, N. (1996) Small increase in triplet repeat length of cerebellum from patients with myotonic dystrophy. *Hum. Genet.*, **98**, 138-140.
23. Jansen, G., Willems, P., Coerwinkel, M., Nillesen, W., Smeets, H., Vits, L., Howeler, C., Brunner, H. and Wieringa, B. (1994) Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)n repeat variation and selection against extreme expansion in sperm. *Am. J. Hum. Genet.*, **54**, 575-585.
24. Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K. and Edstrom, L. (1993) Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum. Mol. Genet.*, **2**, 1397-1400.
25. Ashizawa, T., Dubel, J.R., Tran, L., Harati, Y., Perryman, M.B. and Epstein, H.F. (1993) Muscle DNA shows different triplet expansion size from peripheral blood leukocyte DNA in myotonic dystrophy. *Neurology*, **43**, A 413-A 413.
26. Thornton, C.A., Johnson, K. and Moxley, R.T. (1994) Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.*, **35**, 104-107.
27. Zatz, M., Passos-Bueno, M.R., Cerqueira, A., Marie, S.K., Vainzof, M. and Pavanello, R.C.M. (1995) Analysis of the CTG repeat in skeletal muscle of young and adult myotonic dystrophy patients: when does the expansion occur. *Hum. Mol. Genet.*, **4**, 401-406.
28. Nakamori, M., Sobczak, K., Puwanant, A., Welle, S., Eichinger, K., Pandya, S., Dekdebrun, J., Heatwole, C.R., McDermott, M.P., Chen, T. *et al.* (2013) Splicing biomarkers of disease severity in myotonic dystrophy. *Ann. Neurol.*, **74**, 862-872.
29. 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.
30. 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.

31. 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.
32. Overend, G., Legare, C., Mathieu, J., Bouchard, L., Gagnon, C. and Monckton, D.G. (2019) Allele length of the DMPK CTG repeat is a predictor of progressive myotonic dystrophy type 1 phenotypes. *Hum. Mol. Genet.*, **28**, 2245-2254.
33. 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.
34. Dogan, C., De Antonio, M., Hamroun, D., Varet, H., Fabbro, M., Rougier, F., Amarof, K., Arne Bes, M.C., Bedat-Millet, A.L., Behin, A. *et al.* (2016) Gender as a Modifying Factor Influencing Myotonic Dystrophy Type 1 Phenotype Severity and Mortality: A Nationwide Multiple Databases Cross-Sectional Observational Study. *PLoS One*, **11**, e0148264.
35. Kurkiewicz, A., Cooper, A., McIlwaine, E., Cumming, S.A., Adam, B., Krahe, R., Puymirat, J., Schoser, B., Timchenko, L., Ashizawa, T. *et al.* (2020) Towards development of a statistical framework to evaluate myotonic dystrophy type 1 mRNA biomarkers in the context of a clinical trial. *PLoS One*, **15**, e0231000.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. Yanovsky-Dagan, S., Cohen, E., Megalli, P., Altarescu, G., Schonberger, O., Eldar-Geva, T., Epsztejn-Litman, S. and Eiges, R. (2021) DMPK hypermethylation in sperm cells of myotonic dystrophy type 1 patients. *Eur. J. Hum. Genet.*, https://doi.org/10.1038/s41431-021-00999-3
42. 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.
43. Nakamori, M., Pearson, C.E. and Thornton, C.A. (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Hum. Mol. Genet.*, **20**, 580-588.
44. 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.
45. 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.

46. 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.
47. Du, J., Campau, E., Soragni, E., Jespersen, C. and Gottesfeld, J.M. (2013) Length-dependent CTG.CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. *Hum. Mol. Genet.*, **22**, 5276-5287.
48. Goula, A.V., Stys, A., Chan, J.P., Trottier, Y., Festenstein, R. and Merienne, K. (2012) Transcription elongation and tissue-specific somatic CAG instability. *PLoS Genet*, **8**, e1003051.
49. Mason, A.G., Tome, S., Simard, J.P., Libby, R.T., Bammler, T.K., Beyer, R.P., Morton, A.J., Pearson, C.E. and La Spada, A.R. (2014) Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age. *Hum. Mol. Genet.*, **23**, 1606-1618.
50. Tome, S., Simard, J.P., Slean, M.M., Holt, I., Morris, G.E., Wojciechowicz, K., te Riele, H. and Pearson, C.E. (2013) Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues. *DNA Repair (Amst)*, **12**, 46-52.
51. Flower, M., Lomeikaite, V., Ciosi, M., Cumming, S., Morales, F., Lo, K., Hensman Moss, D., Jones, L., Holmans, P., Investigators, T.-H. *et al.* (2019) MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain*, **142**, 1876-1886.
52. Swami, M., Hendricks, A.E., Gillis, T., Massood, T., Mysore, J., Myers, R.H. and Wheeler, V.C. (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.*, **18**, 3039-3047.
53. Gomes-Pereira, M. and Monckton, D.G. (2006) Chemical modifiers of unstable expanded simple sequence repeats: What goes up, could come down. *Mutat. Res.*, **598**, 15-34.
54. Lopez Castel, A., Cleary, J.D. and Pearson, C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell. Biol.*, **11**, 165-170.
55. Benn, C.L., Gibson, K.R. and Reynolds, D.S. (2021) Drugging DNA Damage Repair Pathways for Trinucleotide Repeat Expansion Diseases. *J. Huntingtons Dis.*, **10**, 203-220.
56. Nakamori, M., Panigrahi, G.B., Lanni, S., Gall-Duncan, T., Hayakawa, H., Tanaka, H., Luo, J., Otabe, T., Li, J., Sakata, A. *et al.* (2020) A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.*, **52**, 146-159.
57. Iachettini, S., Valaperta, R., Marchesi, A., Perfetti, A., Cuomo, G., Fossati, B., Vaienti, L., Costa, E., Meola, G. and Cardani, R. (2015) Tibialis anterior muscle needle biopsy and sensitive biomolecular methods: a useful tool in myotonic dystrophy type 1. *Eur. J. Histochem.*, **59**, 2562.
58. Gudde, A., van Heeringen, S.J., de Oude, A.I., van Kessel, I.D.G., Estabrook, J., Wang, E.T., Wieringa, B. and Wansink, D.G. (2017) Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat. *RNA Biology*, **14**, 1374-1388.
59. Jansen, G., Mahadevan, M., Amemiya, C., Wormskamp, N., Segers, B., Hendriks, W., O'hoy, K., Baird, S., Sabourin, L., Lennon, G. *et al.* (1992) Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. *Nat. Genet.*, **1**, 261-266.
60. Mathieu, J., Boivin, H., Meunier, D., Gaudreault, M. and Begin, P. (2001) Assessment of a disease-specific muscular impairment rating scale in myotonic dystrophy. *Neurology*, **56**, 336-340.
61. Shelbourne, P., Davies, J., Buxton, J., Anvret, M., Blennow, E., Bonduelle, M., Schmedding, E., Glass, I., Lindenbaum, R., Lane, R. *et al.* (1993) Direct diagnosis of myotonic dystrophy with a disease-specific DNA marker. *N. Engl. J. Med.*, **328**, 471-475.

62. Braida, C., Stefanatos, R.K., Adam, B., Mahajan, N., Smeets, H.J., Niel, F., Goizet, C., Arveiler, B., Koenig, M., Lagier-Tourenne, C. *et al.* (2010) 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.*, **19**, 1399-1412.
63. Morales, F., Corrales, E., Zhang, B., Vasquez, M., Santamaria-Ulloa, C., Quesada, H., Sirito, M., Estecio, M.R., Monckton, D.G. and Krahe, R. (2021) Myotonic dystrophy type 1 (DM1) clinical sub-types and CTCF site methylation status flanking the CTG expansion are mutant allele length-dependent. *Hum. Mol. Genet.*, **31**, 262-274.
64. Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N. and Tapscott, S.J. (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell.*, **20**, 483-489.
65. Michel, L., Huguet-Lachon, A. and Gourdon, G. (2015) Sense and Antisense DMPK RNA Foci Accumulate in DM1 Tissues during Development. *PLoS One*, **10**, e0137620.
66. Miyamoto-Mikami, E., Tsuji, K., Horii, N., Hasegawa, N., Fujie, S., Homma, T., Uchida, M., Hamaoka, T., Kanehisa, H., Tabata, I. *et al.* (2018) Gene expression profile of muscle adaptation to high-intensity intermittent exercise training in young men. *Sci. Rep.*, **8**, 16811.
67. Radonic, A., Thulke, S., Mackay, I.M., Landt, O., Siegert, W. and Nitsche, A. (2004) Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun*, **313**, 856-862.
68. Venables, W. and Ripley, B. (2002) *Modern Applied Statistics with S.* Springer, NY.

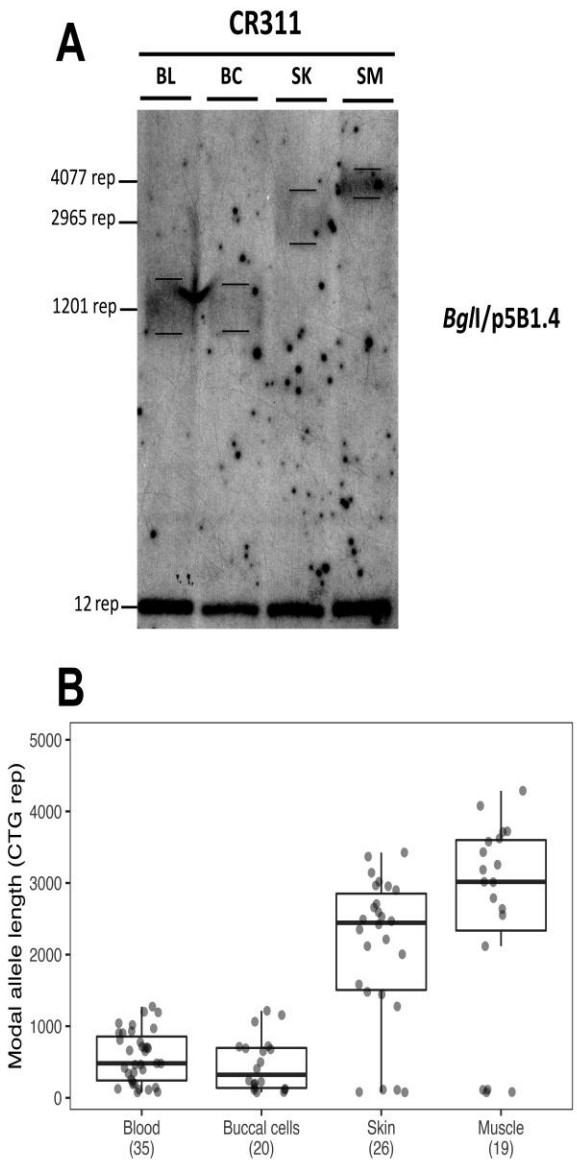

Figure 1. DM1 CTG expansion modal allele size in different tissues. (**A)** Representative autoradiograph (genomic DNA Southern blot hybridization) showing the mutation present in different tissues of the same DM1 patient (CR311, all samples were taken at the same time). The tissues analyzed were: blood (BL), buccal cells (BC), skin (SK) and skeletal muscle (SM). The mutant expanded allele is observed as a smear at the top of the autoradiograph, marked by lines. The non-disease associated allele (12 CTG repeats in this DM1 patient) is shown at the bottom. The expanded modal allele size was measured at the middle of the smear or at the middle of the strongest signal in the smear for each tissue. The molecular weight marker has been transformed to number of repeats. The probe and digestion enzyme used in this assay are also shown. (**B)** Box plots showing the size of the mutation in the tissues analyzed, which is significantly larger in muscle than in the other tissues, and larger in skin than blood and buccal cells. Blood and buccal cells show very similar mutant expanded allele sizes. Box plots show the largest, smallest, median (line), in addition to the 25% and 75% percentiles and the outliers. The number of patients analyzed per tissue for whom we were able to obtain the modal allele size is annotated in parenthesis.

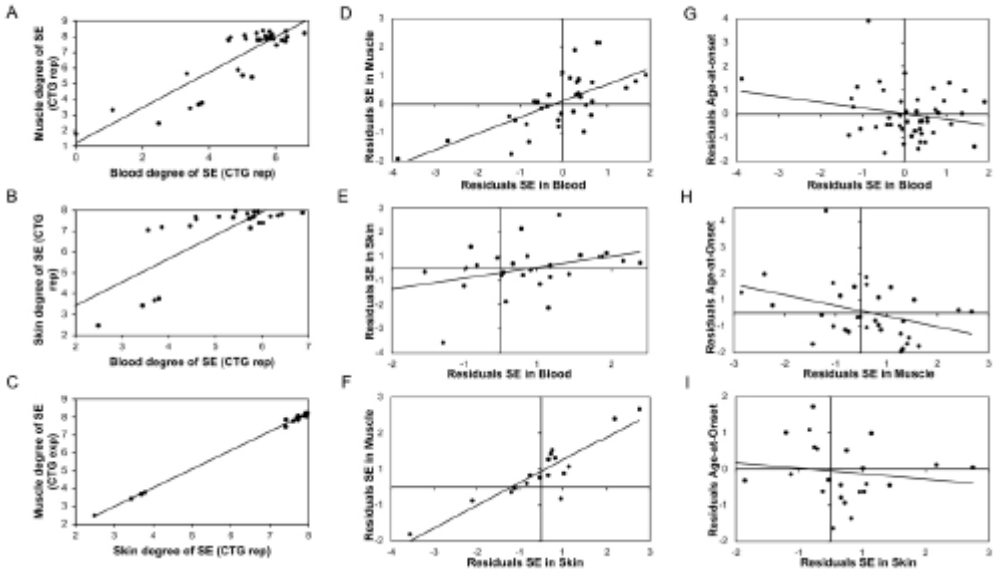

Figure 2. Relationships between the degree of somatic expansion in blood, skeletal muscle and skin, and residual variation among tissues and with the age-at-onset. The plots show the relationship between **(A)** the degree of somatic expansion in muscle and blood, **(B)** skin and blood, and **(C)** skin and muscle. **(D**) Residual variation in somatic expansion between muscle and blood, **(E)** skin and blood, **(F)** muscle and skin; and residual variation **(G)** in age at onset, and residual variation in somatic expansion in blood **(G)**, muscle **(H)**, and skin **(I)**.

Table 1. Multi-variate regression analysis for muscle weakness and damage

Model	adjusted r^2	p	parameter		coefficient	standard error	z-statistic	p
model 1: MW = β_0 + β_1 ln(ePAL) + β_2 ln(ePAL)^2 + β_3 ln(age$_s$) + β_4 Interaction n = 35 DM1 patients	0.297	= 0.005	Intercept	β_0	-116.16	36.43	-3.2	0.003
			ln(ePAL)	β_1	30.77	9.49	3.2	0.003
			ln(ePAL)^2	β_2	-1-87	0.56	-3.3	0.002
			ln(age$_s$)	β_3	15.82	5.74	2.8	0.010
			Interaction	β_4	-2.56	1.01	-2.5	0.017
model 2: MD = β_0 + β_1 ln(ePAL) + β_2 ln(ePAL)^2 + β_3 ln(age$_s$) + β_4 Interaction n = 35 DM1 patients	0.228	= 0.018	Intercept	β_0	-55.79	27.78	-2.0	0.054
			ln(ePAL)	β_1	14.12	7.23	2.0	0.060
			ln(ePAL)^2	β_2	-0.83	0.43	-1.9	0.063
			ln(age$_s$)	β_3	8.06	4.38	1.8	0.076
			Interaction	β_4	-1.17	0.77	-1.5	0.142
model 3: MD = β_0 + β_1 ln(ePAL) + β_2 ln(ePAL)^2 + β_3 ln(age$_s$) + β_4 Sex + β_5 Interaction n = 35 DM1 patients	0.336	= 0.004	Intercept	β_0	-60.05	25.83	-2.3	0.027
			ln(ePAL)	β_1	15.86	6.75	2.4	0.026
			ln(ePAL)^2	β_2	-0.96	0.40	-2.4	0.023
			ln(age$_s$)	β_3	8.44	4.06	2.1	0.047
			Sex = male	β_4	-0.66	0.27	-2.4	0.022
			Interaction	β_5	-1.24	0.72	-1.7	0.094

Table 1. Multi-variate regression models of the relationship between the estimated progenitor allele length (ePAL), age-at-sampling (age$_s$), their interaction and muscle weakness (MW, model 1) and muscle damage (MD, model 2), in addition to sex and muscle damage (MD, model 3). The table shows the squared coefficient of correlation (r^2) and statistical significance (p) for each model, and the coefficient, standard error, z-statistic and statistical significance (p), associated with each parameter in the model. The use of natural logarithm (ln) is indicated as appropriated. The number of individuals used in each analysis is indicated (n).

Table 2. Comparison of tissue-specific patterns of somatic instability of the *DMPK* CTG repeat in four different tissues from DM1 patients

Model	Variables (CTG repeat length)	Median (CTG rep)	N	*p*

WT	Blood vs Buccal cells	325 vs 322	20	= 0.250
WT	Skeletal Muscle vs Blood	3016 vs 691	19	< 0.001
WT	Skeletal Muscle vs Buccal cells	2987 vs 451	14	= 0.005
WT	Skeletal Muscle vs Skin	3185 vs 2491	17	= 0.002
WT	Skin vs Blood	2445 vs 566	26	< 0.001
WT	Skin vs Buccal cells	2350 vs 240	19	< 0.001
MWU	Modal allele length TA vs Quad	3016 vs 2993	19 vs 17	= 0.790
MWU	Blood modal allele length CR vs DMBDI	691 vs 588	35 vs 17	= 0.890
MWU	ePAL CR vs DMBDI	245 vs 318	35 vs 17	= 0.400

WT = Wilcoxon paired signed-rank test; MWU = Mann-Whitney U rank-sum test; TA = tibialis anterior
Quad = quadriceps; CR = Costa Rica; DMBDI = *Dystrophia Myotonica* Biomarker Discovery Initiative

Table 3. Correlations between degree of somatic variation and its associated residuals among the tissues investigated in DM1 patients

Model	Variables (CTG repeat length)	r^2	N	*p*	Figure
Correlation	DSE Blood and Skeletal Muscle	0.75	36	< 0.001	2A
Correlation	DSE Blood and Skin	0.59	26	< 0.001	2B
Correlation	DSE Skin and Skeletal Muscle	0.99	17	< 0.001	2C
Correlation	Res. Blood and Skeletal Muscle	0.46	34	< 0.001	2D
Correlation	Res. Blood and Skin	0.07	26	= 0.095	2E
Correlation	Res. Skin vs Skeletal Muscle	0.79	17	< 0.001	2F
Correlation	Res. DSE in Blood and Res. age-at-onset	0.03	46	= 0.131	2G
Correlation	Res. DSE in Skeletal Muscle and Res. age-at-onset	0.10	32	= 0.041	2H
Correlation	Res. DSE in Skin and Res. age-at-onset	~ 0	23	= 0.53	2I

DSE = degree of somatic expansion; Res. = residuals

Table 4. Multi-variate regression analysis for the allele size in DM1 muscle and the levels of somatic expansion in DM1 muscle and blood

Model	adjusted r^2	*p*	parameter	coefficient	standard	*z*-statistic	*p*

						error		
model 1: $\ln(SE_{BL}) = \beta_0 + \beta_1 \ln(ePAL) + \beta_2 Age_s + \beta_3$ Interaction n = 50 DM1 patients	0.545	< 0.001	Intercept	β_0	-0.22	3.06	-0.1	0.942
			ln(ePAL)	β_1	0.86	0.52	1.6	0.107
			Age_s	β_2	-0.06	0.05	-1.2	0.254
			Interaction	β_3	0.01	0.01	1.4	0.173
model 2: $\ln(SE_{SM}) = \beta_0 + \beta_1 \ln(ePAL) + \beta_2 Age_s + \beta_3$ Interaction n = 35 DM1 patients	0.537	< 0.001	Intercept	β_0	3.47	5.91	0.6	0.561
			ln(ePAL)	β_1	0.67	1.00	0.7	0.508
			Age_s	β_2	-0.13	0.10	-1.3	0.205
			Interaction	β_3	0.02	0.02	1.2	0.224
model 3: $\ln(SE_{SK}) = \beta_0 + \beta_1 \ln(ePAL) + \beta_2 Age_s + \beta 3$ Interaction n = 26 DM1 patients	0.716	< 0.001	Intercept	β_0	12.28	3.85	3.2	0.004
			ln(ePAL)	β_1	-0.94	0.65	-1.4	0.171
			Age_s	β_2	-0.27	0.06	-4.3	< 0.001
			Interaction	β_3	0.05	0.01	4.3	< 0.001

Table 4. Multi-variate regression models of the relationship between the estimated progenitor allele length (ePAL), age-at-sampling (Age_s), their interaction and the levels of somatic expansion (SE) in blood (model 1), skeletal muscle (model 2) and in skin (model 3). The table shows the squared coefficient of correlation (r^2) and statistical significance (p) for each model, and the coefficient, standard error, z-statistic and statistical significance (p), associated with each parameter in the model. The use of natural logarithm (ln) is indicated as appropriated. The number of individuals used in each analysis is indicated (n).

Table 5. Regression analysis for the levels of *DMPK* expression in DM1 muscle and blood and its relationship with methylation flanking the *DMPK* CTG repeat expansion.

Model	adjusted r^2	p	parameter		coefficient	standard error	z-statistic	p
model 1: ln(AS_*DMPK*_BL) = $\beta_0 + \beta_1$ ln(S_*DMPK*_BL)	0.198	= 0.005	Intercept	β_0	1.17	0.73	1.6	0.120
n = 34 DM1 patients			ln(S_*DMPK*_BL)	β_1	0.57	0.19	3.0	0.005
model 2: ln(AS_*DMPK*_SM) = $\beta_0 + \beta_1$ ln(S_*DMPK*_SM)	0.486	< 0.001	Intercept	β_0	-8.82	2.27	-3.9	< 0.001
n = 34 DM1 patients			ln(S_*DMPK*_SM)	β_1	1.60	0.28	5.7	< 0.001
model 3: AS_*DMPK*_SM = $\beta_0 + \beta_1$ CTCF	0.088	= 0.055	Intercept	β_0	74.03	8.04	9.2	< 0.001
n = 32 DM1 patients			CTCF	β_1	-4.38	2.20	-2.0	0.055
model 4: AS/S_*DMPK*_SM = $\beta_0 + \beta_1$ CTCF	0.067	= 0.083	Intercept	β_0	50.77	6.45	7.9	< 0.001
n = 32 DM1 patients			CTCF	β_1	3.16	1.76	1.8	0.083
model 5: S_*DMPK*_SM = $\beta_0 + \beta_1$ CTCF	0.004	= 0.296	Intercept	β_0	3352	199	16.8	< 0.001
n = 32 DM1 patients			CTCF	β_1	-58.05	54.59	-1.6	0.296

Table 5. Simple regression models of the relationship between the levels of the sense *DMPK* transcript and the levels of antisense *DMPK* transcript in blood (model 1) and in muscle (model 2); or between methylation flanking the *DMPK* CTG repeat expansion (CTCF) and the levels of antisense (model 3), sense/antisense ratio (model 4) and sense (model 5) *DMPK* transcripts. The table shows the squared coefficient of correlation (r^2) and statistical significance (p) for each model, and the coefficient, standard error, z-statistic and statistical significance (p), associated with each parameter in the model. The use of natural logarithm (ln) is indicated as appropriated. The number of individuals used in each analysis is indicated (n).