DNA Repair 40 (2016) 57–66

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

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

Fernando Morales[a,b,c,*], Melissa Vásquez[a,c], Carolina Santamaría[a,d], Patricia Cuenca[a,b,c], Eyleen Corrales[a], Darren G. Monckton[e]

[a] Instituto de Investigaciones en Salud (INISA), Universidad de Costa Rica, San José, Costa Rica
[b] Escuela de Medicina, Universidad de Costa Rica, San José, Costa Rica
[c] Centro de Investigación en Neurociencias, Universidad de Costa Rica, San José, Costa Rica
[d] Escuela de Nutrición, Universidad de Costa Rica, San José, Costa Rica
[e] Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom

ARTICLE INFO

Article history:
Received 3 August 2015
Received in revised form 28 January 2016
Accepted 29 January 2016
Available online 8 March 2016

Keywords:
Myotonic dystrophy
Somatic mosaicism
Modifier gene
DNA mismatch repair
Simple sequence repeat
Trinucleotide repeat

ABSTRACT

Somatic mosaicism of the expanded CTG repeat in myotonic dystrophy type 1 is age-dependent, tissue-specific and expansion-biased, contributing toward the tissue-specificity and progressive nature of the symptoms. Previously, using regression modelling of repeat instability we showed that variation in the rate of somatic expansion in blood DNA contributes toward variation in age of onset, directly implicating somatic expansion in the disease pathway. Here, we confirm these results using a larger more genetically homogenous Costa Rican DM1 cohort ($p < 0.001$). Interestingly, we also provide evidence that supports subtle sex-dependent differences in repeat length-dependent age at onset and somatic mutational dynamics. Previously, we demonstrated that variation in the rate of somatic expansion was a heritable quantitative trait. Given the important role that DNA mismatch repair genes play in mediating expansions in mouse models, we tested for modifier gene effects with 13 DNA mismatch gene polymorphisms (one each in *MSH2*, *PMS2*, *MSH6* and *MLH1*; and nine in *MSH3*). After correcting for allele length and age effects, we identified three polymorphisms in *MSH3* that were associated with variation in somatic instability: Rs26279 ($p = 0.003$); Rs1677658 ($p = 0.009$); and Rs10168 ($p = 0.031$). However, only the association with Rs26279 remained significant after multiple testing correction. Although we revealed a statistically significant association between Rs26279 and somatic instability, we did not detect an association with the age at onset. Individuals with the A/A genotype for Rs26279 tended to show a greater propensity to expand the CTG repeat than other genotypes. Interestingly, this SNP results in an amino acid change in the critical ATPase domain of MSH3 and is potentially functionally dimorphic. These data suggest that *MSH3* is a key player in generating somatic variation in DM1 patients and further highlight *MSH3* as a potential therapeutic target.

1. Introduction

Myotonic dystrophy type 1 (DM1) is an extremely variable multisystem disease that affects individuals of both sexes and all ages. It is characterized by the presence of myotonia, progressive muscle weakness and wasting, cardiac conduction defects and neurological manifestations [1]. It is inherited in an autosomal dominant fashion, and is characterised by extreme anticipation with age at onset typically decreasing by 20–30 years per generation [2]. The mutation responsible for DM1 is the expansion of an unstable CTG trinucleotide repeat located in the 3′-untranslated region of the *dystrophia myotonica* protein kinase (*DMPK*) gene located at chromosome 19q13.3 [3–8]. Unaffected individuals in the general population typically inherit between 5 and 37CTG repeats. Alleles in the range 38–49 repeats are described as premutations. Such alleles are not normally associated with symptoms, but have high potential to expand into the disease associated range during germ line transmission [9,10]. Individuals inheriting so called protomuta-

* Corresponding author at: Cuidad Universitaria Rodrigo Facio, Instituto de Investigaciones en Salud, Universidad de Costa Rica, 2060 San Pedro de Montes de Oca, San José, Costa Rica.
E-mail address: fernando.moralesmontero@ucr.ac.cr (F. Morales).

http://dx.doi.org/10.1016/j.dnarep.2016.01.001
1568-7864/

tions alleles, 50–80 repeats, typically develop only mild symptoms [11]. Full mutations >80 repeats are associated with more severe adult forms of the disease, increasing up to several thousand repeats in congenitally affected children [4,6,12–14]. Once into the expanded range, the repeat becomes highly genetically unstable in both the germ line and soma. Sex-dependent germ line expansions account for anticipation, the almost exclusive maternal transmission of congenital DM1 and the excess of transmitting grandfathers [12,14–18]. Somatic instability is expansion-biased, age-dependent and tissue-specific [19–22]. Notably, the repeat length in affected tissues such as skeletal muscle is typically many thousands of repeats larger in affected adults [22–25]. Given the positive correlation between repeat length and disease severity, it seems likely that somatic mosaicism contributes toward both the tissue specificity and progressive nature of the symptoms.

Age-dependent, expansion-biased somatic mosaicism results in increases in modal allele length in blood DNA over time [19–21]. This effect has the potential to confound genotype-phenotype correlations and the interpretation of intergenerational transmissions [21,26–28]. One way in which the confounding effects of somatic mosaicism can be addressed is by estimating the progenitor allele length (ePAL) [21]. PAL is defined as the single allele transmitted in the egg or sperm from the affected parent. Using small pool-PCR (SP-PCR) to resolve the high levels of variation observed in blood DNA into discrete length alleles derived from individual cells it is possible to gain an insight into the degree of somatic variation present in an individual at a given time [21]. In relatively young individuals, or individuals inheriting relatively small expanded alleles, a highly positively skewed distribution with a relatively sharp lower boundary is often observed in the expanded allele in blood DNA. Given the expansion-biased nature of somatic mosaicism, this observed lower boundary represents the best available estimate for the PAL. Although in older individuals and/or individuals inheriting larger alleles, such a sharp lower boundary is often not observed and some large contractions are observed in some cells, it is still possible to estimate PAL as the lower bound of a more normally distributed spread of alleles [21]. Using this approach to determine ePAL we were able to correct the inherent age at sampling bias that is further inflated by anticipation in DM1 [29]. Using simple regression modeling we were able to determine that ePAL accounts for ~70% of the variation in age of onset, establishing that allele length is the major modifier of disease severity in DM1. In the same study, using single molecule PCR approaches we also obtained a detailed quantitative measure of the degree of somatic instability in the blood DNA of each individual. These data enabled us to quantify the effects of ePAL and age at sampling, which interact synergistically to account for ~89% of the variation in somatic instability. Critically, we also established that residual variation in somatic instability (*i.e.* somatic instability not accounted for by age at sampling and ePAL) and residual variation in age at onset (*i.e.* variation in age at onset not accounted for by ePAL) are inversely correlated with each other (*i.e.* individuals with greater rates of somatic expansion develop symptoms earlier than expected and *vice versa*), providing direct evidence that somatic instability contributes toward the progressive nature of DM1. Moreover, we also established that individual-specific variation in residual somatic instability is a heritable quantitative trait, suggesting that individual specific differences in somatic instability are mediated, at least in part, by genetic modifiers [29].

Somatic mosaicism of the CTG repeat accumulates throughout the life time of an individual as the net product of multiple small expansions and contractions [30]. Although the precise molecular mechanisms that generate repeat length changes have not been established, data from various expanded CTG•CAG repeat mouse models have revealed an essential role for components of the DNA mismatch repair (MMR) machinery. Specifically, engineered null alleles have revealed critical roles for *Msh2*, *Msh3*, *Pms2*, *Mlh1* and *Mlh3* in generating expansions [31–34]. In contrast, mice homozygous null for *Msh6* had higher levels of somatic instability than wild type mice [32]. This effect is thought to be mediated by competition between MSH6 and MSH3 for the obligate partner MSH2, with *Msh6* null freeing more MSH2 to form expansion promoting MSH2/3 MutS beta complexes [32]. More recently, natural polymorphisms in *Msh3* and *Mlh1* have been revealed as mediators of mouse strain specific differences in CTG•CAG repeat instability [34,35]. It has been similarly hypothesised that polymorphisms in human DNA mismatch repair genes might mediate individual-specific differences in CTG•CAG repeat instability in DM1 patients [27]. Establishing a modifier role for any MMR gene variants in DM1 patients would have implications for establishing the mechanism of expansion and may also have prognostic value. Here, we have sought to test the hypothesis that mismatch repair gene variants are associated with individual specific variation in the degree of somatic variation by quantifying somatic mosaicism and genotyping mismatch repair gene polymorphisms in a large cohort of DM1 patients.

2. Material and methods

2.1. DM1 population

Costa Rican individuals with a molecular diagnosis of DM1 were recruited to the study and information on age at sampling and age of onset of DM1 symptoms were collected. Age of onset was based on the detection of physical myotonia (grip myotonia), muscle weakness and/or the presence of cataracts. Age of onset was recorded after clinical evaluation by one of four different neurologists, or after an interview by the same neurologists or by one of two different geneticists. In total we recruited 199 individuals to the study (114 males, 85 females) (see Supplementary Table S1). Of the 199 individuals recruited, 178 presented with DM1 symptoms (101 males, 77 females) and 21 remained asymptomatic (13 males, 8 females) at the time of the sampling. Participants in the study were recruited with informed consent, in accordance with the protocols approved by the Ethical Scientific Committee of the University of Costa Rica. After obtaining the informed consent on each participant, a blood sample was obtained and leukocyte DNA was purified using standard procedures.

2.2. Estimation of PAL and quantification of somatic instability

We used small pool PCR analyses to estimate PAL (ePAL) in all 199 individuals using the previously described methods [21,29]. Briefly, five replicate reactions with 180–300 pg of genomic DNA (~30–50 cellular equivalents) were analysed and the progenitor allele estimated from the lower boundary of the distribution [21,29] (Fig. 1). ePAL was estimated blind relative to any other data. We also used small pool PCR analyses to measure the degree of somatic instability of the expanded *DMPK* CTG repeat in 174 individuals (99 males, 75 females) using the previously described methods [21,29]. For detailed quantification of the degree of somatic variation, samples were amplified in multiple reactions with ~10–70 pg DNA per reaction such that alleles amplified from single input molecules could be sized (Supplementary Fig. S1). For each individual, the degree of somatic instability was defined as the range between the 10th and 90th percentile of the allele length frequency distribution thus derived [29].

2.3. Selection of DNA MMR gene polymorphisms

Polymorphisms in the *MSH3*, *MSH2*, *MSH6*, *PMS2* and *MLH1* MMR candidate modifier genes were identified in the human genome