Hum Genet (2014) 133:1311–1318
DOI 10.1007/s00439-014-1467-8

ORIGINAL INVESTIGATION

Modifiers of $(CAG)_n$ instability in Machado–Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes

Sandra Martins · Christopher E. Pearson · Paula Coutinho · Sylvie Provost · António Amorim · Marie-Pierre Dubé · Jorge Sequeiros · Guy A. Rouleau

Received: 24 March 2014 / Accepted: 3 July 2014 / Published online: 16 July 2014

Abstract Twelve neurological disorders are caused by gene-specific CAG/CTG repeat expansions that are highly unstable upon transmission to offspring. This intergenerational repeat instability is clinically relevant since disease onset, progression and severity are associated with repeat size. Studies of model organisms revealed the involvement of some DNA replication and repair genes in the process of repeat instability, however, little is known about their role in patients. Here, we used an association study to search for genetic modifiers of $(CAG)_n$ instability in 137 parent–child transmissions in Machado–Joseph disease (MJD/SCA3). With the hypothesis that variants in genes involved in DNA replication, repair or recombination might alter the MJD CAG instability patterns, we screened 768 SNPs from 93 of these genes. We found a variant in *ERCC6* (rs2228528) associated with an expansion bias of MJD alleles. When using a gene–gene interaction model, the allele combination G–A (rs4140804–rs2972388) of *RPA3–CDK7* is also associated with MJD instability in a direction-dependent manner. Interestingly, the transcription-coupled repair factor ERCC6 (aka CSB), the single-strand binding protein RPA, and the CDK7 kinase part of the TFIIH transcription repair complex, have all been linked to transcription-coupled repair. This is the first study performed in patient samples to implicate specific modifiers of CAG instability in humans. In summary, we found variants in three transcription-coupled repair genes associated with the MJD mutation that points to distinct mechanisms of $(CAG)_n$ instability.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-014-1467-8) contains supplementary material, which is available to authorized users.

S. Martins · G. A. Rouleau (✉)
Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, 3801 rue University, Montreal, QC H3A2B4, Canada
e-mail: guy.rouleau@mcgill.ca

S. Martins · A. Amorim
IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, 4200-465 Porto, Portugal
e-mail: smartins@ipatimup.pt

C. E. Pearson (✉)
Genetics and Genome Biology, The Hospital for Sick Children, 686 Bay Street, Toronto, ON M5G 0A4, Canada
e-mail: cepearson.sickkids@gmail.com

C. E. Pearson
Program of Molecular Genetics, University of Toronto, Toronto, ON M5S1A1, Canada

P. Coutinho
Serviço Neurologia, Hospital São Sebastião, 4520-211 Feira, Portugal

S. Provost · M.-P. Dubé
The Montreal Heart Institute Research Centre, Montreal, QC H1T1C8, Canada

A. Amorim
Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal

M.-P. Dubé
Université de Montréal, Montreal, QC H3T1J4, Canada

J. Sequeiros
UnIGENe, IBMC, 4150-180 Porto, Portugal

J. Sequeiros
ICBAS, Universidade do Porto, 4050-313 Porto, Portugal

Introduction

Disease-associated repeats are commonly called "dynamic mutations" due to the recurrent changes in size of pathogenic/expanded alleles observed upon parent–offspring transmission (Richards and Sutherland 1992). Intergenerational instability is a common trait to most of the over 30 described human neurological and neuromuscular inherited repeat-associated diseases. Depending on the *locus*, the tendency towards expansion or contraction and magnitude achieved in male *versus* female lineage varies considerably (Pearson et al. 2005). Together with the biological novelty in dissecting the mechanisms of intergenerational instability, much attention has been given to this topic due to the clinical consequence of allele sizes: repeat disorders are commonly characterized by the correlation between size of expanded alleles and severity of the phenotype. Therefore, repeat instability causes devastating effects, particularly when an expansion bias is observed (e.g., in Huntington disease (HD), myotonic dystrophy (DM) and most spinocerebellar ataxia's (SCAs)) since it underlies the clinical phenomenon of *anticipation*, i.e., an earlier age-of-onset and increased disease severity in successive generations (Pearson et al. 2005).

The high mutability of expanded alleles implies that inherited mutations occur during mitotic, meiotic and post-meiotic development of germ cells, and/or postzygotically; thus, all stage-specific DNA metabolic activities of replication, repair, and recombination may contribute to repeat instability. Although sometimes contradictory and controversial, several in vitro and in vivo studies have shed some light into these complex and interdependent processes that involve both *cis*-elements and *trans*-acting factors (Kovtun and McMurray 2008; Lopez Castel et al. 2010). Indeed, there is evidence of somatic CAG instability in MJD patient tissues (Cancel et al. 1998; Tanaka et al. 1996, 1999) and support for both proliferation and non-proliferative processes of CAG instability in the central nervous system (Lopes-Cendes et al. 1996; Takano et al. 1996). Replication and instability can be mechanistically linked through nucleases involved in Okazaki processing during lagging-strand synthesis [as shown in yeast rad27/fen 1 mutants, although not reproduced in mouse (Larsen et al. 2003; van den Broek et al. 2006) or human studies (Moe et al. 2008)], replication stalling (Krasilnikova and Mirkin 2004) and impairment during replication fork progression (Gan et al. 2011). In addition, other replication-associated proteins like helicases (Chan et al. 2012), DNA polymerases (Shah et al. 2012), and ligases [as DNA ligase 1, LIG1 (Lopez Castel et al. 2009)] may participate in repeat instability, particularly in the process of repairing aberrant tracts. Repair-associated mechanisms include mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), and homologous-recombination repair. The MMR system is responsible for the elimination of base–base mismatches as well as insertion–deletion loops, which arise as a consequence of DNA polymerase slippage during replication and recombination. Recent evidence revealed slipped-DNAs at the DM1 locus in patient tissues (Axford et al. 2013). In the context of repeat instability, several mouse models indicate that, contrarily to expected, MMR proteins do not play a corrective role in CAG transgenes, but need instead to be functional to cause expansion mutations (Kovtun and McMurray 2001). The most studied MSH2, MSH3, PMS2, MLH1, and MLH3 are definitively involved in repeat instability, but the exact mechanism of their action remains unclear as does their relevance to humans (Ezzatizadeh et al. 2012; Pinto et al. 2013). BER is specialized in the detection and removal of damaged DNA bases. BER enzymes have been shown to affect, particularly, somatic instability: OGG1 (the major DNA glycosylase in the brain involved in removing endogenous lesions from the genome) (Kovtun et al. 2007), and APE1 (the major endonuclease that cleaves apurinic/apyrimidinic sites) are unable to process lesions within secondary DNA structures, typical of CAG repeats (Goula et al. 2009). Transcription-coupled nucleotide excision repair (TC-NER) is the example of a link between different pathways (Lin and Wilson 2007). Indeed, in human cells, this mechanism requires coordination of not only NER proteins (CSB, XPA, ERCC1, and XPG) and a transcription factor (TFIIS), but also MMR components (MSH2/MSH3), and an E3 ubiquitin ligase (BRCA1/BARD1) (Lin and Wilson 2007). Finally, homologous-recombination events can be either reciprocal (crossing over) or nonreciprocal (gene conversion) depending on whether it occurs with or without the exchange of flanks. Both imply heteroduplex intermediates formed by complementary base pairing of single strands that may be corrected by the MMR machinery. In pedigree analyses, recombination-mediated instability is more easily noticed when repeat tracts show an interrupted cassette-like configuration (Brown and Brown 2004); nevertheless, gene conversion has also been claimed to underlie repeat instability of some alleles in HD (Pritchard et al. 1992), DM (O'Hoy et al. 1993), and MJD (Mittal et al. 2005).

In this study, we searched for modifiers of instability in the most common dominant ataxia worldwide: Machado–Joseph disease (MJD) or spinocerebellar ataxia type 3 (SCA3) (Sequeiros et al. 2011). The mutation responsible for MJD is a CAG repeat expansion in exon 10 of the *ATXN3* gene (14q32.12) (Kawaguchi et al. 1994) (SCAbase; www.scabase.eu). MJD has a variable clinical presentation, with four sub-types described according to age-of-onset and phenotype severity (Coutinho and Andrade 1978; Rosenberg 1983). Although this is typically

a late-onset disease, patients have been described with onset between 4 and 70 years old (Carvalho et al. 2008; Coutinho 1992); this variability is mainly (50–80 %) explained by the $(CAG)_{exp}$ size (Maciel et al. 1995, 2001; Maruyama et al. 1995; van de Warrenburg et al. 2005). Expanded alleles, usually over 61 repeats, display a mutation bias towards expansion (mainly in the paternal lineage), which reflects the importance of the intergenerational instability at the clinical level.

Here, to search for *trans*-acting modifiers of MJD intergenerational instability, we performed an association study to evaluate whether DNA variants in replication, repair and recombination genes modify $(CAG)_n$ instability in MJD families.

Materials and methods

Subjects

We analyzed 137 patients from 61 MJD families, mostly of European (Portuguese) origin. DNA samples, previously extracted from peripheral blood, were quantified by a PicoGreen reaction done in an ABI Prism 7900HT Real-Time PCR System (Applied Biosystems®). In case of insufficient DNA for the SNP genotyping (39 samples), we performed genome amplification via multiple displacement amplification (MDA) using the REPLI-g Midi Kit (Qiagen®). Mutational errors introduced by MDA have been reported as negligible in the analysis of biallelic markers (Jiang et al. 2005).

Analysis of the repeat instability

To uniformly assess the expanded CAG repeat size in blood samples of our 137 parent–child pairs, we performed capillary electrophoresis of amplified samples in an ABIPrism 3730 sequencer (Applied Biosystems®) followed by GeneMapper analysis. For each transmission of the expanded allele, we determined the intergenerational repeat instability by comparing the absolute changes in $(CAG)_n$ size between patient and transmitting parent.

Selection of candidate genes and SNP variants

We selected a total of 93 genes known to be involved in DNA repair (MMR, BER, and NER), replication and recombination (Supplementary Table S1). Diversity in these genes was assessed through SNP genotyping. We started by selecting non-synonymous SNPs described in ENSEMBL database (www.ensemble.org), and other non-synonymous variants that although rare could affect the interaction with the *ATXN3* locus; these included SNPs (1) within the interaction region with other repair/replication proteins, (2) within crucial protein domains, and/or (3) associated to repair efficiency. In addition, some intronic SNPs were included when located (4) in splice sites, (5) in transcription factor binding sites, (6) within triplex-forming sequences, or (7) due to high minor allele frequencies in control populations to serve as tagSNPs (www.HapMap.org) (Supplementary Table S2). A total of 768 SNPs were genotyped through the Illumina GoldenGate platform (McGill University and Genome Québec Innovation Centre). Among the 768 selected SNPs spread through the 93 candidate genes, in the offspring, 449 were not polymorphic in our cohort, and 83 SNPs were removed because of a completion rate lower than 80 %. There remained 236 SNPs in 78 candidate genes for association testing (Supplementary Table S2).

Statistical analyses

We performed a genetic association study between the SNPs and MJD instability using pedigree-based association testing. The tested endpoint was repeat size change of the deleterious allele upon transmission, which was assessed: (1) as a continuous variable, representing the difference in size to capture the full distribution of instability; as a quantitative trait, recoding (2a) the direction (contraction, no change, expansion) and (2b) the amplitude of $(CAG)_n$ length variation (large contraction, small contraction, no change, small expansion, large expansion; small instability defined as repeat size changes lower than 2 CAGs); and (3) as a dichotomous variable (contraction *versus* no change; expansion *versus* no change). Genetic association analysis in nuclear families for SNPs, haplotypes and gene–gene interactions was performed using UNPHASED v3.1.3 (Dudbridge 2008).

To evaluate the robustness of the findings, we tested results with the addition of covariates. We have performed univariate and multivariate analyses of $(CAG)_n$ instability in function of (1) the flanking SNP-based haplotype of MJD alleles, (2) the gender of the transmitting parent, and (3) the initial repeat size (present in the blood of the transmitting parent). We first focused our approach on each variant individually, and next analyzed combinations of multiple DNA variants (both at the intragenic and intergenic levels) while considering modifier effects on instability modulated by a combination of variants on candidate genes. To determine the number of independent statistical tests and define the significance threshold for each analysis, we used the method of Gao et al. (Gao et al. 2008) to compute the M_{eff} followed by Bonferroni correction. No adjustment was made for the different expansion coding models. The significance threshold was set to $p = 0.0005$. Descriptive statistics and evaluation of covariates to include in the model were carried out using the SAS 9.2 software.

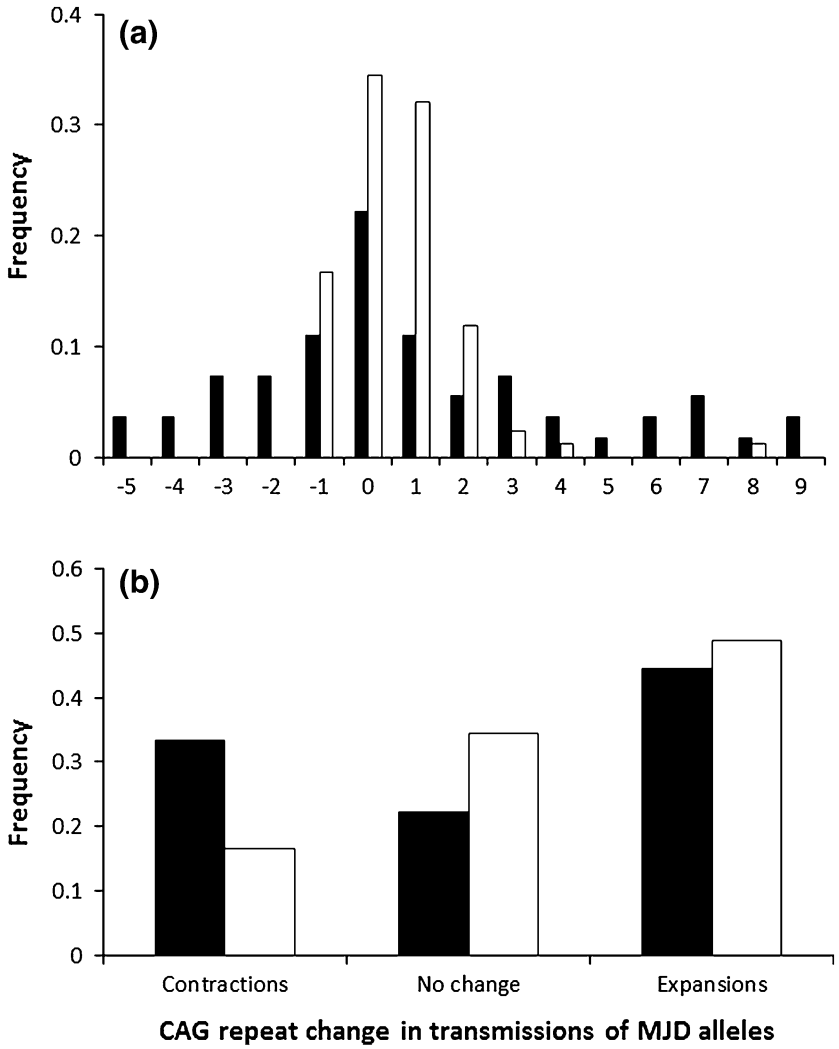

Fig. 1 Intergenerational instability in the $(CAG)_n$ size of MJD alleles, showing the full distribution (**a**) and instability direction (**b**), in paternal *(black bars*; $n = 54$) and maternal (*white bars*; $n = 84$) transmissions

We assessed the MJD haplotype (Machado vs. Joseph), but in the cohort of 137 patients analyzed, we did not find a significant effect on instability; thus, this covariate was not included in further association analyses.

Results

In our 137 MJD patient–transmitting parent pairs, intergenerational instability ranged between −5 and +9 repeats. Considering large instability when repeat sizes change two or more CAGs during transmission, we found 8.7 % of large and 14.5 % of small contractions; and 23.9 % of small and 23.2 % of large expansions; repeat size was stable in 29.7 % of the transmissions (Fig. 1).

We tested 236 SNPs in 78 DNA metabolizing genes of the offspring for association with transmitted MJD CAG repeat size changes. We found association with the *ERCC6* gene (*excision repair cross-complementing rodent repair deficiency, complementation group 6*) involved in the transmitted instability of MJD alleles in patient blood. The variant rs2228528 (NP_000115.1:p.Gly399Asp) is associated with an expansion bias (p value <0.0005); patients with the GG genotype more frequently had expansions of the CAG repeat than stable transmissions ($n = 53$ vs. $n = 26$). When covariates (gender of the transmitting parent and mutational origin) were included in the binary model, the association was maintained ($p = 0.0009$). Functional prediction of this variant on the ERCC6 protein indicated that this polymorphism is likely to be tolerated, according to SIFT (score = 1) and PolyPhen (score = 0.001); also the affected amino acid has not been identified in a critical domain of this DNA-binding protein. This result was not surprising since rather than a general repair-pathway defect (which would cause genome-wide instability), we hypothesize that specific alterations on the interaction of repair proteins with *ATXN3* may result in the *locus*-specific repeat instability observed in MJD families. Alternatively, this SNP can be in LD with other variant within *ERCC6* resulting in the observed expansion bias. Therefore, we tested

Table 1 Haplotypes in the *ERCC6* gene of MJD patients associated with an expansion bias of inherited $(CAG)_n$ alleles (significance threshold $p < 0.0005$)

SNP IDs					*p* value
rs2228529	rs2228527	rs4253046	rs2228528	rs4253013	$p < 0.0008$
rs2228529	rs2228527	rs4253046	rs2228528	x	$p < 0.0005$
x	rs2228527	rs4253046	rs2228528	x	$p < 0.0005$
rs2228529	x	rs4253046	rs2228528	x	$p < 0.0005$
rs2228529	rs2228527	x	rs2228528	x	$p < 0.0005$
rs2228529	x	x	rs2228528	x	$p < 0.0005$
x	rs2228527	x	rs2228528	x	$p < 0.0005$
x	x	rs4253046	rs2228528	x	$p < 0.0005$

whether other variants from the same gene resulted in a similar instability pattern or not. By analyzing *ERCC6* haplotypes with all five SNPs genotyped, we found a significant association when non-synonymous SNPs are included, but not in the presence of rs4253013 (the single synonymous variant genotyped). Indeed, significant bias towards expansion was observed for haplotypes with rs2228528 and any combination of one or two of the other synonymous SNPs (Table 1). Regarding the recombination hotspots and haplotype structure of *ERCC6*, assessed for the European population by the HapMap consortium, one can see that analyzed SNPs are spread through different haplotype blocks. This suggests that *ERCC6*-dependent MJD instability is mainly associated with its larger LD block of 32 kb, although additional non-synonymous SNPs (independently segregated from rs2228528 in healthy European controls) may also contribute to this effect. The fact that these variants per se do not show significant associations with instability may be explained by their weaker effect when compared to rs2228528.

Next, to test the hypothesis of more than one modifier contributing in concert to CAG instability, we analyzed SNPs of different genes with a pedigree-based association test for gene–gene interaction. We observed the genetic interaction between *RPA3* and *CDK7* genes (variants rs4140804 and rs2972388, respectively) affecting again the direction of instability during transmission of expanded MJD alleles ($p = 9.4 \times 10^{-6}$). The allelic combination G–A (*RPA3-CDK7*) is associated with an expansion bias (AddVal = 169.7) when the G–G combination (that shows the lowest correlation with expansion) has been set as reference.

The replication protein A3 (RPA3) is a trimeric single-strand binding protein, involved in DNA replication, various processes of DNA repair and recombination (Binz et al. 2004; Oakley and Patrick 2010; Wold 1997). In *RPA3*, rs4140804 is an intronic SNP. The rs2972388 of *cyclin-dependent kinase 7*, *CDK7* gene, is in high LD with the remaining SNPs in this gene that are all embedded in a single haplotype block in the European population. Other SNP variants of *CDK7* and *RPA3* may, therefore, be affecting the CDK7–RPA3 interaction that contributes to MJD instability. In the String database (http://string-db.org), evidence of a functional link between CDK7 (ENSP00000256443) and RPA3 (ENSP00000223129) has been suggested, although association predicted in curated databases for this interaction harbors only a medium level of confidence (score 0.900).

Discussion

In repeat-associated disorders, the allele size is the major determinant of disease age-of-onset, progression and severity. Due to the repetitive nature of the repeat tract, the formation of unusual DNA structures during DNA replication, repair and recombination is thought to lead to repeat length mutations. Here, we found variants in genes that code for proteins involved in nucleotide excision repair (*ERCC6*), in transcription initiation and DNA repair (*CDK7*) and various DNA metabolic pathways (*RPA3*) associated with different instability patterns of intergenerationally transmitted expanded CAG alleles in MJD families.

Many DNA repair proteins are involved in cancer biology, with several mutations identified as the cause for cancer predisposition (Curtin 2012); in this case, sequence alterations prevent their ability to repair damaged DNA and, as a consequence, high levels of mutagenesis and tumorigenesis occur. In our study, we did not expect to find variants altering the DNA repair activity in general; otherwise, we would observe a higher risk for cancer susceptibility in patients with more MJD instability. Instead, our hypothesis is that some of the analyzed variations may alter the repair protein interaction with the *ATXN3* locus, resulting in different instability patterns.

Evidence for distinct mechanisms underlying the direction of instability in MJD comes from our analyses since both the rs2228528 variant within *ERCC6* and the *CDK7–RPA3* allele combination in the gene–gene interaction model have been shown to affect MJD instability in a direction-dependent manner. Several studies with bacteria and yeast suggest different mechanisms for contraction and expansion mechanisms, although unicellular

organisms differ from mammals in their propensity to contract or expand. In humans, different mechanisms underlying small and large expansions have been suggested in HD, where many small changes were detected early in paternal gametogenesis, whereas large expansions occurred during or after meiosis (Yoon et al. 2003). In MJD, the timing of instability remains unknown; our findings gave us a hint on the pathways that may be involved in postzygotic MJD instability since we analyzed genotypes of patients inheriting the expansion. In a previous clinical study with MJD families, evidence for distinct mechanisms of contraction *versus* expansion came from the observation of the two MJD haplotype backgrounds preferentially associated to one or the other direction of repeat change: 75 % of contractions were observed in families carrying the Machado lineage, whereas 72 % of expansions in the worldwide spread Joseph haplotype (Martins et al. 2008). Interestingly, the two mutational Machado and Joseph MJD lineages are *yin-yang* stable SNP-backgrounds in which independent-origin expansions occurred. The fact that these two very distinct haplotype backgrounds of the CAG expansion are differently associated to one or other direction of MJD instability supports our hypothesis that different instability patterns may result from alterations in the repair protein interaction with the *ATXN3* locus. In a transgenic mouse model of MJD, different tendencies towards contraction and expansion have been found for transmissions through female and male progenitors, respectively (Silva-Fernandes et al. 2010).

Polymorphisms in *CDK7* (Angstadt et al. 2012; Barry et al. 2012; Cunningham et al. 2009; George Priya Doss et al. 2013; Jeon et al. 2010; Li et al. 2007; Ma et al. 2011), *ERCC6* (Abbasi et al. 2009; Baas et al. 2010; Chang et al. 2009; Chiu et al. 2008; Laugel et al. 2010; Lin et al. 2008; Ma et al. 2009; Tuo et al. 2006) and *RPA* (Jin et al. 2013; Michiels et al. 2007) have been shown to significantly associate with altered susceptibility to certain cancers, through unknown mechanisms. The functional contribution of any DNA repair gene polymorphism upon either cancer or CAG repeat instability is an area of interest, but will require considerable efforts in future research.

It is interesting that we detected significant associations of polymorphic variants of three factors each of which is involved in the same DNA repair pathway: transcription-coupled repair. CDK7, a part of the TFIIH complex, has been implicated in repair. Interestingly, RPA has been shown to interact with TFIIH (Compe and Egly 2012; Mu et al. 1997), providing further support for a functional implication in our observed significance association of RPA variants with CDK7 variants and MJD CAG instability. Some evidence for a role of transcription in CAG instability has already been provided from model systems (Nakamori et al. 2011), as well as the involvement of several factors including CSB/ERCC6. For example in a cell model harboring an integrated CAG tract, transcription through CAG repeats induces their instability and knockdown of CSB/ERCC6 (a component specifically required for transcription-coupled NER) stabilizes CAG repeats against transcription-induced CAG contractions (Lin and Wilson 2007). Some evidence suggests that CSB/ERCC6 can protect against intergenerational and somatic expansion of CAG tracts in HD mice (Kovtun et al. 2011). Furthermore, recent findings in a wild-type and an HD transgenic mouse model revealed differential expression of—Pcna, Rpa1, Msh6, Fen1 and Lig1, between the striatum and the cerebellum; tissues that are susceptible and spared to neurodegeneration and show differential levels of repeat expansions (Mason et al. 2014). The elevated expression of these factors in the cerebellum over the striatum suggested that they may protect against the higher CAG expansions observed in the striatum. Curiously, we did not detect significant associations of polymorphisms of either *MSH2* or *MSH3* in MJD CAG instability, which might have been expected since several mouse models of CAG instability have shown a dependence upon *MSH2* and *MSH3* genes (Tome et al. 2013). However, there has been no study using affected patients to implicate a genetic pathway.

Limitations

We have presumed that the CAG size differences between the blood of the parent and that of the offspring are a reflection of the level of somatic instability ongoing in specific tissues. While this may not be the case, a detailed understanding of the somatic instability in any tissue has not been performed. Furthermore, we have assessed a potential contribution of SNPs in the offspring, which can represent influences upon postzygotic instability. The full set of diploid SNPs of the paternal and maternal genomes must be considered for contribution to germline instability (Pearson 2003), an analysis that has not been performed herein. The relative contribution of DNA repair gene variations in the parental germline versus that of the offspring is not easily assessed (Tome et al. 2013). Some DNA metabolizing proteins like MSH2/MSH3 affect repeat instability in all tissues equally (Manley et al. 1999; Tome et al. 2013), some uniquely affect instability in the germline, like LIG1 (Tome et al. 2013), and some affect only the neural tissues, like XPA (Hubert et al. 2011). Our study is a step towards understanding the pathways underlying these differences, which may be further explored using animal models to answer the remaining questions. As for all genetic association studies, a replication study in a similar or ethnically distinct population would add valuable evidence to support the genetic associations reported in this study.

Acknowledgments This work was supported by the Canadian Institutes of Health Research; and the Levesque Chair for research in Neurogenetics [to G.A.R.]; and the Portuguese Foundation for Science and Technology [SFRH/BPD/77969/2011 to S.M].

Conflict of interest The authors have declared that no competing interests exist.

Ethical standards All experiments here described comply with the current laws of the countries in which they were performed.

References

Abbasi R, Ramroth H, Becher H, Dietz A, Schmezer P, Popanda O (2009) Laryngeal cancer risk associated with smoking and alcohol consumption is modified by genetic polymorphisms in ERCC5, ERCC6 and RAD23B but not by polymorphisms in five other nucleotide excision repair genes. Int J Cancer 125:1431–1439

Angstadt AY, Thayanithy V, Subramanian S, Modiano JF, Breen M (2012) A genome-wide approach to comparative oncology: high-resolution oligonucleotide aCGH of canine and human osteosarcoma pinpoints shared microaberrations. Cancer Genet 205:572–587

Axford MM, Wang YH, Nakamori M, Zannis-Hadjopoulos M, Thornton CA, Pearson CE (2013) Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. PLoS Genet 9:e1003866

Baas DC, Despriet DD, Gorgels TG et al (2010) The ERCC6 gene and age-related macular degeneration. PLoS One 5:e13786

Barry KH, Koutros S, Andreotti G et al (2012) Genetic variation in nucleotide excision repair pathway genes, pesticide exposure and prostate cancer risk. Carcinogenesis 33:331–337

Binz SK, Sheehan AM, Wold MS (2004) Replication protein a phosphorylation and the cellular response to DNA damage. DNA Repair (Amst) 3:1015–1024

Brown LY, Brown SA (2004) Alanine tracts: the expanding story of human illness and trinucleotide repeats. Trends Genet 20:51–58

Cancel G, Gourfinkel-An I, Stevanin G, Didierjean O, Abbas N, Hirsch E, Agid Y, Brice A (1998) Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado–Joseph disease. Hum Mutat 11:23–27

Carvalho DR, La Rocque-Ferreira A, Rizzo IM, Imamura EU, Speck-Martins CE (2008) Homozygosity enhances severity in spinocerebellar ataxia type 3. Pediatr Neurol 38:296–299

Chan NL, Hou C, Zhang T, Yuan F, Machwe A, Huang J, Orren DK, Gu L, Li GM (2012) The Werner syndrome protein promotes CAG/CTG repeat stability by resolving large (CAG)(n)/(CTG)(n) hairpins. J Biol Chem 287:30151–30156

Chang CH, Chiu CF, Wang HC et al (2009) Significant association of ERCC6 single nucleotide polymorphisms with bladder cancer susceptibility in Taiwan. Anticancer Res 29:5121–5124

Chiu CF, Tsai MH, Tseng HC, Wang CL, Tsai FJ, Lin CC, Bau DT (2008) A novel single nucleotide polymorphism in ERCC6 gene is associated with oral cancer susceptibility in Taiwanese patients. Oral Oncol 44:582–586

Compe E, Egly JM (2012) TFIIH: when transcription met DNA repair. Nat Rev Mol Cell Biol 13:343–354

Coutinho P (1992) Doença de Machado–Joseph—Tentativa de definição. University of Porto, Portugal

Coutinho P, Andrade C (1978) Autosomal dominant system degeneration in Portuguese families of the Azores Islands. A new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions. Neurology 28:703–709

Cunningham JM, Vierkant RA, Sellers TA et al (2009) Cell cycle genes and ovarian cancer susceptibility: a tagSNP analysis. Br J Cancer 101:1461–1468

Curtin NJ (2012) DNA repair dysregulation from cancer driver to therapeutic target. Nat Rev Cancer 12:801–817

Dudbridge F (2008) Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered 66:87–98

Ezzatizadeh V, Pinto RM, Sandi C, Sandi M, Al-Mahdawi S, Te Riele H, Pook MA (2012) The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. Neurobiol Dis 46:165–171

Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev 25:2041–2056

Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol 32:361–369

George Priya Doss C, Nagasundaram N, Chakraborty C, Chen L, Zhu H (2013) Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: a molecular dynamics approach. Hum Genomics 7:10

Goula AV, Berquist BR, Wilson DM 3rd, Wheeler VC, Trottier Y, Merienne K (2009) Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. PLoS Genet 5:e1000749

Hubert L Jr, Lin Y, Dion V, Wilson JH (2011) Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. Hum Mol Genet 20:4822–4830

Jeon S, Choi JY, Lee KM, Park SK, Yoo KY, Noh DY, Ahn SH, Kang D (2010) Combined genetic effect of CDK7 and ESR1 polymorphisms on breast cancer. Breast Cancer Res Treat 121:737–742

Jiang H, Tang B, Xu B, Zhao GH, Shen L, Tang JG, Li QH, Xia K (2005) Frequency analysis of autosomal dominant spinocerebellar ataxias in Han population in the Chinese mainland and clinical and molecular characterization of spinocerebellar ataxia type 6. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 22:1–4

Jin T, Zhang J, Li G, Li S, Yang B, Chen C, Cai L (2013) TP53 and RPA3 gene variations were associated with risk of glioma in a Chinese Han population. Cancer Biother Radiopharm 28:248–253

Kawaguchi Y, Okamoto T, Taniwaki M et al (1994) CAG expansions in a novel gene for Machado–Joseph disease at chromosome 14q32.1. Nat Genet 8:221–228

Kovtun IV, McMurray CT (2001) Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet 27:407–411

Kovtun IV, McMurray CT (2008) Features of trinucleotide repeat instability in vivo. Cell Res 18:198–213

Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature 447:447–452

Kovtun IV, Johnson KO, McMurray CT (2011) Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. Aging (Albany NY) 3:509–514

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Mol Cell Biol 24:2286–2295

Larsen E, Gran C, Saether BE, Seeberg E, Klungland A (2003) Proliferation failure and gamma radiation sensitivity of Fen1 null mutant mice at the blastocyst stage. Mol Cell Biol 23:5346–5353

Laugel V, Dalloz C, Durand M et al (2010) Mutation update for the CSB/ERCC6 and CSA/ERCC8 genes involved in Cockayne syndrome. Hum Mutat 31:113–126

Li Y, Jin G, Wang H et al (2007) Polymorphisms of CAK genes and risk for lung cancer: a case-control study in Chinese population. Lung Cancer 58:171–183

Lin Y, Wilson JH (2007) Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Mol Cell Biol 27:6209–6217

Lin Z, Zhang X, Tuo J et al (2008) A variant of the Cockayne syndrome B gene ERCC6 confers risk of lung cancer. Hum Mutat 29:113–122

Lopes-Cendes I, Maciel P, Kish S et al (1996) Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado–Joseph disease. Ann Neurol 40:199–206

Lopez Castel A, Tomkinson AE, Pearson CE (2009) CTG/CAG repeat instability is modulated by the levels of human DNA ligase I and its interaction with proliferating cell nuclear antigen: a distinction between replication and slipped-DNA repair. J Biol Chem 284:26631–26645

Lopez Castel A, Cleary JD, Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol 11:165–170

Ma H, Hu Z, Wang H et al (2009) ERCC6/CSB gene polymorphisms and lung cancer risk. Cancer Lett 273:172–176

Ma H, Chen J, Pan S, Dai J, Jin G, Hu Z, Shen H, Shu Y (2011) Potentially functional polymorphisms in cell cycle genes and the survival of non-small cell lung cancer in a Chinese population. Lung Cancer 73:32–37

Maciel P, Gaspar C, DeStefano AL et al (1995) Correlation between CAG repeat length and clinical features in Machado–Joseph disease. Am J Hum Genet 57:54–61

Maciel P, Costa MC, Ferro A et al (2001) Improvement in the molecular diagnosis of Machado–Joseph disease. Arch Neurol 58:1821–1827

Manley K, Shirley TL, Flaherty L, Messer A (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. Nat Genet 23:471–473

Martins S, Coutinho P, Silveira I, Giunti P, Jardim LB, Calafell F, Sequeiros J, Amorim A (2008) Cis-acting factors promoting the CAG intergenerational instability in Machado–Joseph disease. Am J Med Genet B Neuropsychiatr Genet 147B:439–446

Maruyama H, Nakamura S, Matsuyama Z et al (1995) Molecular features of the CAG repeats and clinical manifestation of Machado–Joseph disease. Hum Mol Genet 4:807–812

Mason AG, Tome S, Simard JP, Libby RT, Bammler TK, Beyer RP, Morton AJ, Pearson CE, La Spada AR (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

Michiels S, Danoy P, Dessen P, Bera A, Boulet T, Bouchardy C, Lathrop M, Sarasin A, Benhamou S (2007) Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers. Carcinogenesis 28:1731–1739

Mittal U, Srivastava AK, Jain S, Mukerji M (2005) Founder haplotype for Machado–Joseph disease in the Indian population: novel insights from history and polymorphism studies. Arch Neurol 62:637–640

Moe SE, Sorbo JG, Holen T (2008) Huntingtin triplet-repeat locus is stable under long-term Fen1 knockdown in human cells. J Neurosci Methods 171:233–238

Mu D, Wakasugi M, Hsu DS, Sancar A (1997) Characterization of reaction intermediates of human excision repair nuclease. J Biol Chem 272:28971–28979

Nakamori M, Pearson CE, Thornton CA (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. Hum Mol Genet 20:580–588

Oakley GG, Patrick SM (2010) Replication protein A: directing traffic at the intersection of replication and repair. Front Biosci (Landmark Ed) 15:883–900

O'Hoy KL, Tsilfidis C, Mahadevan MS, Neville CE, Barcelo J, Hunter AG, Korneluk RG (1993) Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. Science 259:809–812

Pearson CE (2003) Slipping while sleeping? Trinucleotide repeat expansions in germ cells. Trends Mol Med 9:490–495

Pearson CE, Nichol Edamura K, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6:729–742

Pinto RM, Dragileva E, Kirby A et al (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet 9:e1003930

Pritchard C, Zhu N, Zuo J et al (1992) Recombination of 4p16 DNA markers in an unusual family with Huntington disease. Am J Hum Genet 50:1218–1230

Richards RI, Sutherland GR (1992) Dynamic mutations: a new class of mutations causing human disease. Cell 70:709–712

Rosenberg RN (1983) Dominant ataxias. Res Publ Assoc Res Nerv Ment Dis 60:195–213

Sequeiros J, Martins S, Silveira I (2011) Epidemiology and population genetics of degenerative ataxias. In: Subramony S, Durr A (eds) Ataxic disorders, Chap. 14, Handbook of Clinical Neurology, Vol. 103 (3rd series). Elsevier, Edinburgh, pp 225–248

Shah KA, Shishkin AA, Voineagu I, Pavlov YI, Shcherbakova PV, Mirkin SM (2012) Role of DNA polymerases in repeat-mediated genome instability. Cell Rep 2:1088–1095

Silva-Fernandes A, Costa Mdo C, Duarte-Silva S et al (2010) Motor uncoordination and neuropathology in a transgenic mouse model of Machado–Joseph disease lacking intranuclear inclusions and ataxin-3 cleavage products. Neurobiol Dis 40:163–176

Takano H, Onodera O, Takahashi H et al (1996) Somatic mosaicism of expanded CAG repeats in brains of patients with dentatorubral-pallidoluysian atrophy: cellular population-dependent dynamics of mitotic instability. Am J Hum Genet 58:1212–1222

Tanaka F, Sobue G, Doyu M et al (1996) Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeats in dentatorubral-pallidoluysian atrophy, Machado–Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. J Neurol Sci 135:43–50

Tanaka F, Ito Y, Sobue G (1999) Somatic mosaicism of expanded CAG trinucleotide repeat in the neural and nonneural tissues of Machado–Joseph disease (MJD). Nihon Rinsho 57:838–842

Tome S, Manley K, Simard JP et al (2013) MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet 9:e1003280

Tuo J, Ning B, Bojanowski CM et al (2006) Synergic effect of polymorphisms in ERCC6 5′ flanking region and complement factor H on age-related macular degeneration predisposition. Proc Natl Acad Sci USA 103:9256–9261

van de Warrenburg BP, Hendriks H, Durr A, van Zuijlen MC, Stevanin G, Camuzat A, Sinke RJ, Brice A, Kremer BP (2005) Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch–French cohort. Ann Neurol 57:505–512

van den Broek WJ, Nelen MR, van der Heijden GW, Wansink DG, Wieringa B (2006) Fen1 does not control somatic hypermutability of the (CTG)(n)*(CAG)(n) repeat in a knock-in mouse model for DM1. FEBS Lett 580:5208–5214

Wold MS (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem 66:61–92

Yoon SR, Dubeau L, de Young M, Wexler NS, Arnheim N (2003) Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc Natl Acad Sci USA 100:8834–8838