Supplementary Information

Supplementary Information comprises a clinical description and three figures.

Clinical Description

Family 1

Patient DMGV165 was diagnosed clinically with myotonic dystrophy at age 28, after presenting to a hospital physician complaining primarily of excessive sleepiness and fatigue. He had a history of blepharitis, and cataracts at age 50 years. On last review at age 63, he had marked facial weakness with ptosis and dysarthria. Distal weakness and wasting of peripheral muscles consistent with DM1 were also readily evident, necessitating use of a walking stick when indoors, and wheelchair outdoors.

Patient DMGV14, the daughter of DMGV165, reported no symptoms suggestive of DM1, but had predictive testing for DM1 at 18 years due to the family history. Unusual TP-PCR traces were obtained, *i.e.* the expansion detected appeared longer from the 5'-flank than from the 3'-flank (data not shown). On most recent review by an experienced specialty doctor in clinical genetics (BB) at age 33, no muscle signs of DM1 were detectable, and the patient continues to undertake full-time employment in an office environment. She had a history of hypothyroidism.

The son of DMGV165, DMGV57, presented at age five after a schoolteacher raised concerns about his fine motor skills. DM1 testing was performed by Southern blot analysis of restriction digested genomic DNA, confirming an expansion of approximately 500 repeats. He required speech and language therapy in childhood, and a modified Malone procedure at age 17 for faecal incontinence. On review at age 36, examination by a consultant neurologist (RP) revealed typical features of DM1 including frontal balding and marked facial weakness with ptosis. Distal weakness

was severe, with finger flexion MRC grade 1 (insufficient to produce joint motion) bilaterally and finger extension grade 3 to 4. Biceps and triceps power were grade 3 bilaterally, and shoulder movements relatively preserved at grade 4. In the lower limbs, ankle dorsiflexion was virtually absent, and more proximal movements were grade 4 to 5. He was areflexic, and used ankle-foot orthoses for mobility.

Patient DMGV83, paternal uncle of 14 and 57, reported onset of muscle symptoms at age 38. He had a history of cataracts and implantation of a cardiac pacemaker-defibrillator. On most recent review at age 71, he had recently seen a speech and language therapist in relation to swallowing difficulty, and was referred for wheelchair provision due to worsening proximal weakness and falls.

Family 2

In family 2, the diagnosis of DM1 was first made in DMGV184. Having researched his own symptoms of muscle myotonia, recurrent pilomatrixoma, oligospermia, bowel symptoms with bacterial overgrowth, recurrent infections (due to low grade neutropenia with low IgG) and excessive daytime sleepiness, he requested investigation for myotonic dystrophy at age 27. Unidirectional TP-PCR from the 3' UTR of *DMPK* confirmed presence of an expanded CTG allele > 170 repeats.

His brother, DMGV182, requested testing at age 43 due to the family history. As for DMGV14, the TP-PCR trace extended much further from the 5'-flank than from the 3'-flank (data not shown).

DMGV206, the father of both brothers, underwent testing at age 64 with unidirectional TP-PCR confirming the presence of an expanded CTG allele. Subsequent ECG showed atrial flutter, for which DC cardioversion was arranged. At age 68 he kept generally good health with mild muscle symptoms, walking around 2

miles most days including uphill. DMGV242, the paternal aunt of the proband, was likewise investigated as part of cascade screening, and was confirmed to carry the familial CTG expansion by unidirectional TP-PCR. Her past medical history included recurrent hyperparathyroidism, osteopenia and irritable bowel syndrome. On last review at age 65 the patient had mild distal weakness, and had recently been referred for physiotherapy following a fall.

Family 3

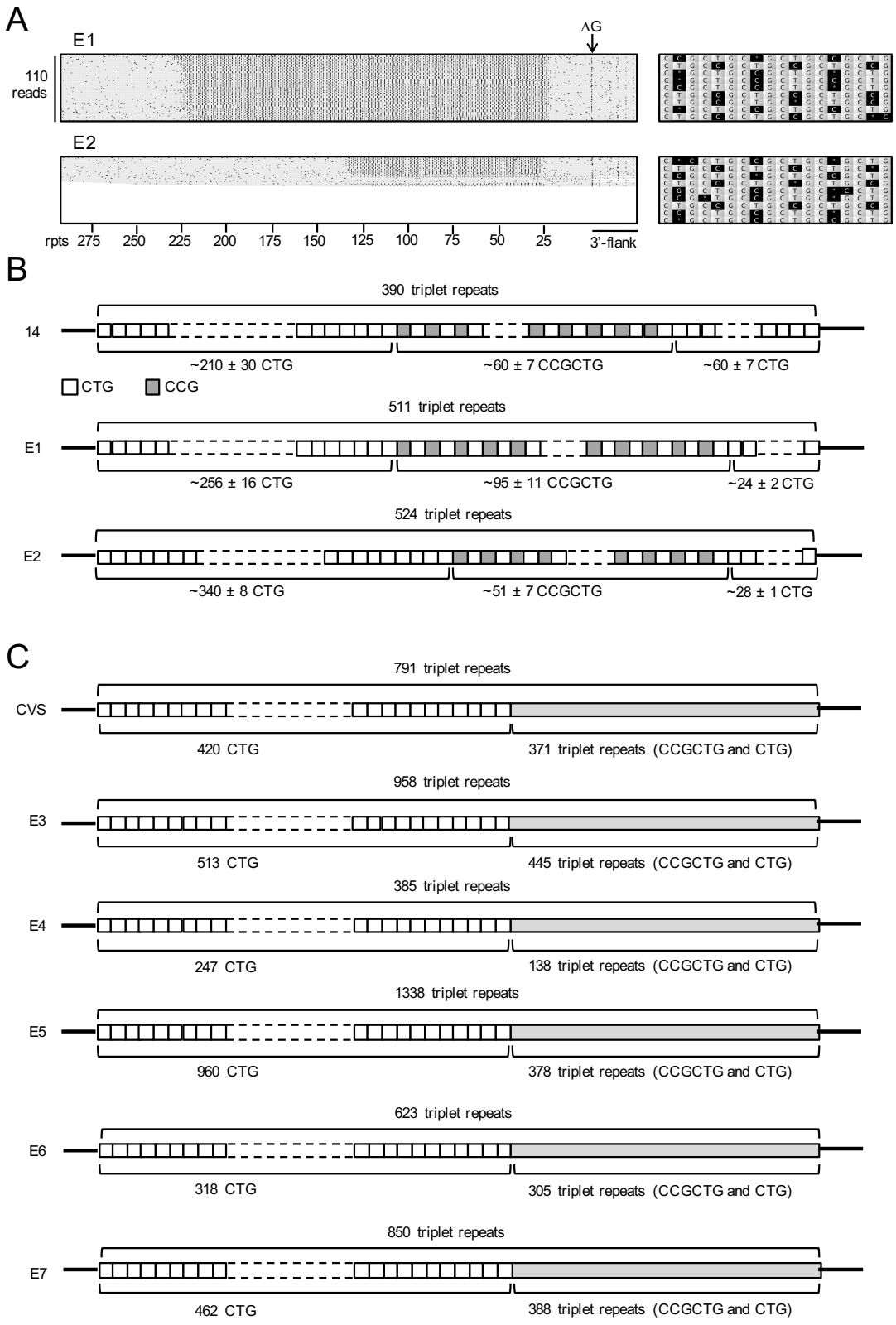
Patient DMGV234 was first to present in family 3, with weakness affecting his hands at age 32. A clinical diagnosis of DM1 was made, based on presence of myopathic facies, frontal balding, grip myotonia and percussion myotonia of the thenar eminence, and was supported by electromyography. On review at age 53, he had severe generalised muscle weakness requiring wheelchair use, marked grip and percussion myotonia, and bilateral ptosis. Investigation of excessive somnolence led to a diagnosis of chronic respiratory failure, which was managed with tracheostomy and domiciliary ventilation. He died suddenly at home overnight at age 54.

His daughter, DMGV15, had a previous history of resection subvalvular aortic stenosis in childhood and ligation of patent ductus arteriosus in infancy. She had predictive testing for DM1 at 22 due to the family history, and reported no symptoms at that time. The TP-PCR from the 3'-flank was at reduced intensity compared to that from the 5'-flank (data not shown). At age 46 the patient continued to report no significant symptoms of DM1, and was undertaking full time employment in the catering sector.

The brother of DMGV15 requested predictive testing for myotonic dystrophy at age 32, after developing weakness affecting his hands and neck. He was noted to have myopathic facies with temporal wasting. Unidirectional TP-PCR from the 3' flanking region of *DMPK* confirmed presence of an expanded CTG allele. On review by a consultant neurologist (RP) at age 43, a slightly unusual pattern of upper limb weakness with relative sparing of supraspinatus and deltoid was noted. Biceps and triceps were more severely impaired at MRC grade 4 with atrophy. Moderate distal weakness, typical of DM1 was also noted. He had ankle contractures, and mild long finger flexor contractures.

Bidirectional TP-PCR has since been carried out on blood DNA from DMGV15, demonstrating a characteristic ladder consistent with an expanded repeat in the 5' direction, and a ladder with reduced intensity in the 3' direction.

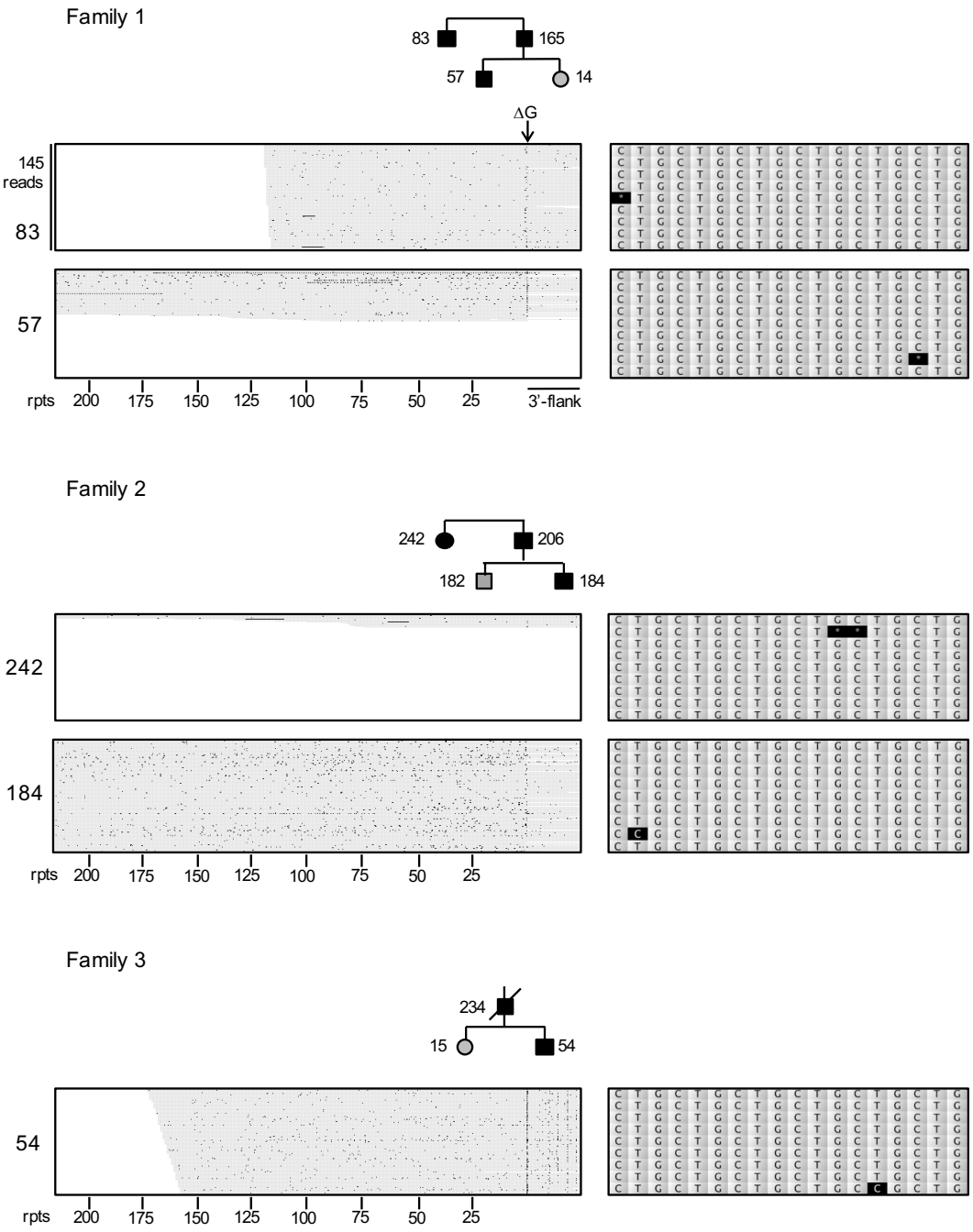
Supplementary Figure S1



Supplementary Figure S1 Predicted repeat composition of DM1 expanded alleles from DMGV14, seven of her WGA embryo DNA samples (E1 to E7) and a chorionic

villus sample (CVS). **(A)** Zoomed out (left hand panels) and zoomed in (right hand panels) view of PacBio sequence reads from expanded alleles derived from whole genome amplified DNA from two blastomeres, E1 and E2. The reads were aligned against a reference sequence comprising 600 CTG repeats and 72 bp of 3'-flanking sequence. Mismatches relative to the pure CTG repeats in the reference sequence appear as black boxes. The number of reads is shown to the left of the top left-hand panel. The boundary between the repeats and the 3'-flank, where a G nucleotide is frequently missing from the sequence reads, is marked ΔG. The distance in repeats (rpts) from the 3'-flank is shown below the lower panel. Variant repeats, comprising CCGCTG hexamer motifs, are present near the 3'-end of the expanded alleles in both E1 and E2. **(B)** The average triplet repeat composition of the expanded alleles from DMGV14, E1 and E2, determined by examining individual sequence reads. **(C)** The predicted triplet repeat composition of the expanded allele from a chorionic villus sample (CVS) and five further blastomere samples, E3-E7, each derived from a separate fertilisation. The allele structure was assumed to be broadly similar to those sequenced, *i.e.* pure CTG repeats near the 5'-end, CCGCTG hexamers close to the 3'-end, and a shorter region of pure CTG repeats near the 3'-end. The overall allele length and the number of CTG repeats near the 5'-end were determined from the lengths of undigested and digested expanded alleles in the SP-PCR and AciI digestion experiment (Fig 2). The restriction endonuclease AciI digests CCG and CGG variant repeats. The combined number of CCGCTG hexamers and pure CTGs near the 3'-end, shown as a grey rectangle, was estimated from the reduction in length of the expanded alleles following AciI digestion. All lengths are expressed as numbers of triplet repeats.

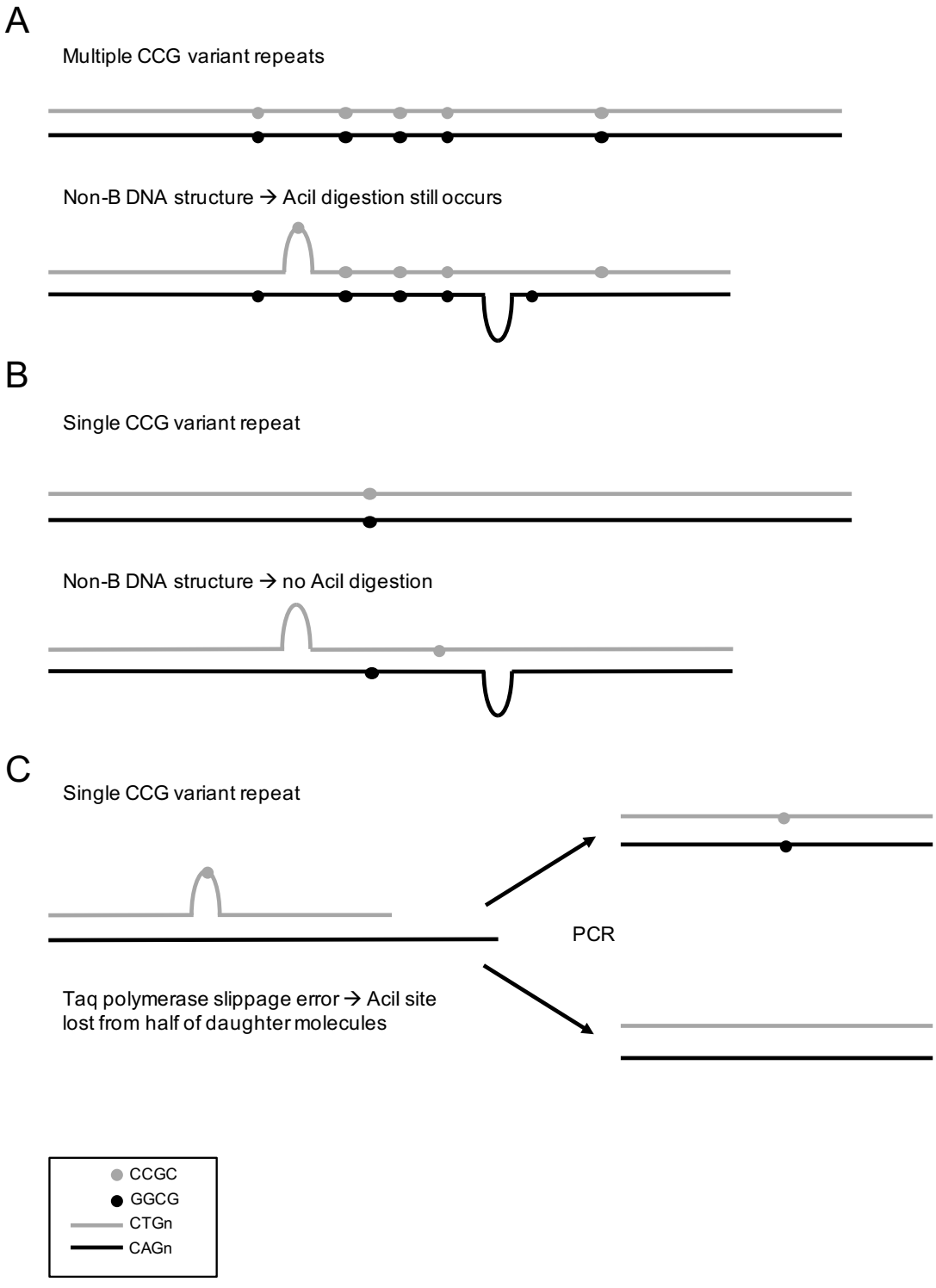
Supplementary Figure S2



Supplementary Figure S2 PacBio sequence reads from DM1 expanded alleles from other members of all three families contain no germ line variant repeat sequences. The left-hand panels show zoomed-out PacBio sequence reads, the right-hand panels show zoomed-in sequence reads. The reads were aligned against a reference sequencing comprising 600 CTG repeats and 72 bp of 3'-flanking sequence.

Mismatches compared to the reference sequence appear black. The number of reads in each zoomed-out panel is shown to the left of the top left panel for family 1. The boundary between the repeats and the 3'-flank, where a G nucleotide is frequently missing from the sequence reads, is marked ΔG. The distance in repeats (rpts) from the 3'-flank is shown below the lower left panel for family 1. The alleles all consist of pure CTG repeats, with a small number of mismatches. These have no clear pattern and may derive from a combination of somatic mutations and PCR and sequencing errors.

Supplementary Figure S3



Supplementary Figure S3 Slipped-strand products and PCR errors may explain why around 17% of PacBio sequence reads from DMGV182's expanded allele contained

no CCG variant repeats (Fig 3), but no individual band in a single molecule SP-PCR experiment remained completely undigested by AciI (Fig 4). **(A)** In this hypothetical expanded allele containing five CCG variant repeats, when slipped strand products with complementary loopouts form as shown, three from five of the CCGC sequences are still correctly base paired with GGCG, and the AciI restriction enzyme is able to digest the DNA at these three sites. The two remaining recognition sequences are lost and do not digest with AciI. **(B)** When there is a single CCG variant repeat, as in this hypothetical expanded allele, a slipped strand product with complementary loopout might disrupt base pairing between CCGC and GGCG, resulting in failure to digest with AciI. **(C)** If a Taq polymerase slippage error resulted in the omission of a CCG variant repeat from the newly synthesized strand, subsequent rounds of PCR would result in a mixture of products that contained CCG, and were digested with AciI, and products that had lost the variant repeat and remained undigested.

Around 70% of bands in the single molecule SP-PCR and AciI digestion experiment using DMGV182's DM1 alleles (Fig 4) were only partially digested by AciI. If the germline mutation that introduced variant repeats into DMGV182's expanded allele had introduced only a single CCG, both possible mechanisms (B and C) could contribute towards a proportion of individual molecules within a single band in the SP-PCR and digestion experiment (Fig 4) failing to digest with AciI. However, if multiple CCG variant repeats were present, the majority of bands would still digest with AciI. The subset of PacBio sequence reads that contained no CCG variant repeats might have been generated by the slippage error mechanism suggested in (C), resulting in the loss of the CCG variant repeat from a subset of the amplicons used to generate the sequencing library, and generating a substantial minority of pure CTG sequence reads. Alternatively, during bioinformatic processing to generate the ccs sequences from the raw reads, processing errors might have resulted in the loss of a single CCG repeat from a proportion of the reads.