Human Molecular Genetics, 2003, Vol. 12, No. 9 **1021–1028**
DOI: 10.1093/hmg/ddg111

Candidate DNA replication initiation regions at human trinucleotide repeat disease loci

Taurai Nenguke[1], Mirit I. Aladjem[2], James F. Gusella[3], Nancy S. Wexler and The Venezuela HD Project[4] and Norman Arnheim[1,*]

[1]Program in Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90089-1340, USA, [2]Laboratory of Molecular Pharmacology, National Cancer Institute, Bethesda, MD 20892, USA, [3]Department of Psychiatry, Molecular Neurogenetics Unit, Massachusetts General Hospital, Charlestown MA 02129, USA and [4]Department of Neurology, College of Physicians and Surgeons, Columbia University and the Hereditary Disease Foundation, New York, NY 10032, USA

Received December 19, 2002; Revised and Accepted February 24, 2003

The positions of DNA replication initiation regions (IRs) at three human trinucleotide repeat (TNR) disease loci were examined in order to characterize the role played by IRs in explaining the known locus-specific variation in TNR instability levels. Using three different normal cell lines, candidate IRs were identified at the *HD, SCA-7* and *SBMA* loci. At each locus the IR is less than 3.6 kb from the CAG/CTG repeat tract. Preliminary studies with a cell line homozygous for an *HD* disease mutation indicated no change in the position of the candidate IR in spite of the mutation. Comparison with experimental results from model systems suggests that a complex relationship may exist between instability and the proximity and/or orientation of the repeats with respect to an IR.

INTRODUCTION

At least 14 human diseases result from expansion of trinucleotide repeats (TNR) in or adjacent to protein-coding sequences (reviewed in 1,2). For most of the diseases, the repeat sequence consists of (CAG/CTG) triplets. Expanded tracts are likely to undergo further expansion during germline transmissions and often in particular somatic tissues depending upon the disease. The degree of germline instability of $(CAG/CTG)_n$ tracts varies significantly at different chromosomal positions, even when the disease alleles have identical numbers of repeats and the sex of the affected (transmitting) parent is the same (reviewed in 1–3). The reason for this inter-locus variation is not understood, but one possible explanation involves a DNA replication slippage model of expansion (4–6) that can be influenced by sequences adjacent to the repeat tract. In model systems, different instability levels have been attributed to both distance from a DNA replication initiation region (IR) as well as the orientation of the repeats with respect to the IR (1,7–11). Orientation is defined by whether the CAG or CTG tract is on the leading strand (or lagging strand) template during DNA replication.

In this study, we examined the location of IRs at three human TNR disease loci in normal cells; spinal cerebellar ataxia type 7 (*SCA-7*), Huntington disease (*HD*) and spinal and bulbar muscular atrophy (*SBMA*). Our results suggest that a complex relationship is likely to exist between instability and both proximity and repeat tract orientation.

RESULTS

We used well-established methods (12–14) for identifying replication initiation regions (IRs) using nascent DNA isolated from asynchronously growing cultured human cells and without any drugs to manipulate the cell's DNA metabolism. According to the principles of the nascent strand abundance assay, a population of short nascent DNA fragments will contain a higher concentration of sequences immediately adjacent to an IR, relative to sequences further away from the IR. The concentration of any target sequence in the nascent DNA preparation is quantified by comparing the rate of accumulation of PCR product with that of a series of standards with known amounts of total genomic DNA. A target sequence that is at a high concentration in nascent DNA will accumulate PCR products more rapidly than a target with a lower concentration. Thus, we identified candidate IRs by determining the concentration of nascent DNA sequences from different

*To whom correspondence should be addressed. Email: arnheim@usc.edu

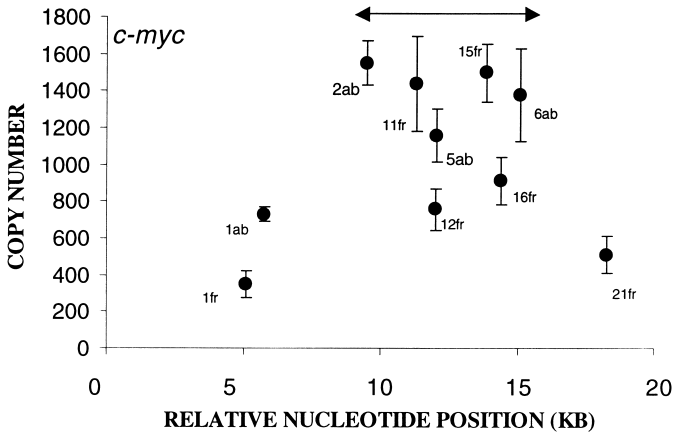

Figure 1. Abundance of nascent DNA templates for a series of *c-myc* markers plotted as a function of their relative position. The absolute position of each marker and the primer sequences can be found in the supplementary information. Each marker is represented by a solid circle and an error bar representing the 95% confidence interval based on duplicate PCR analyses (in some cases the variation is too small to observe on the graph). The name of each marker (without the locus prefix) is also given. Primer pairs with an 'fr' extension are identical to ones used previously (12). The double-headed (horizontal) arrow represents the region of origin activity characterized by others (12,16).

regions on both sides of a repeat tract using kinetic PCR (kt–PCR) (15) and primer pairs (markers) specific for each region. Those targets that are enriched in nascent DNA are defined as candidate IRs. Total genomic DNA was used to construct a standard curve for each primer pair. This standard curve was then used to estimate the copy number of the same target sequence in the nascent DNA preparation. For each gene studied, we used a BLASTN search to check that all primer pairs amplified a single copy genomic sequence. This is critical because multiple copies of a DNA segment, each with different chromosomal IR locations, will generate an enrichment and localization pattern that is the sum of all the individual members of the sequence family. Note, however, that the use of a total genomic DNA standard curve prevents a repeated sequence family lacking IR activity from appearing enriched.

Validation of the assay using *c-myc*

We determined the replication initiation sites in the well-characterized IR at the human *c-myc* locus (12,16) to validate the nascent strand purification and real-time PCR methodologies. DNA was purified from a lymphoblastoid cell line (HDG-7) and nascent strands were isolated using lambda exonuclease digestion. This enzyme preferentially digests 5′ phosphorylated DNA (single or double stranded) but cannot degrade Okazaki fragments that retain the 5′ RNA segment laid down during the initiation of lagging strand synthesis. In this study, an IR was detected in a ~6–7 kb segment. The region with the highest initiation activity (Fig. 1) was located between markers c-myc-2ab and c-myc-6ab as previously reported (12,16). The maximal enrichment in these regions was almost 4-fold when the highest copy number in this interval was compared with the average copy number of the two most distal flanking markers (c-myc-1 fr and c-myc-21fr). A different lymphoblastoid cell line (GM03469) was also studied using nascent DNA isolated by the same method (data not shown). As in the HDG-7 line, the markers between c-myc-2ab and c-myc-6ab were enriched (maximum, 5.5-fold) relative to the most distal flanking markers.

Analysis of the *SCA-7* locus

Having validated the assay, we studied the normal *SCA-7* locus in the HDG-7 cell line. Nascent strands were isolated by lambda exonuclease digestion of total genomic DNA. Compared to the two most distal flanking markers, an enrichment of ~4.8-fold was observed for marker SCA7-6ab located about 745 bp from the repeat tract (Fig. 2). Significant enrichment was also seen with marker 6gh, an additional 200 bp further away from the repeat tract than 6ab. The same highly reproducible enrichment pattern was seen in four additional experiments (data not shown). Two of these independent studies used the GM03469 cell line and nascent DNA prepared by lambda exonuclease digestion. DNA from the other two experiments used extruded, size fractionated DNA. With an average of 11 markers per test, SCA7-6ab or SCA7-6gh gave the highest enrichment compared to the average copy number of the two most distal flanking markers used in each study (data not shown).

The *SCA-7* CAG strand is the sense strand and codes for poly-glutamine. The candidate IR is located between 700 and 900 bp upstream of the repeats (relative to the direction of transcription). This defines the CAG strand as the template for Okazaki fragment synthesis during DNA replication in cells with an unexpanded repeat tract (Fig. 2).

Candidate initiation region at the *HD* locus

To examine the Huntington disease (*HD*) locus, nascent strands were obtained from K562 cells by lambda exonuclease digestion of total genomic DNA. One marker located about 3.6 kb away from the repeat tract (HD-26ab; spanning nucleotides 17502–17714) was highly enriched relative to the two most distal flanking markers (Fig. 3).

Four additional experiments, each starting with total genomic DNA from an independently grown cell culture (three using K562 and one using GM03469), were carried out using nascent DNA isolated by strand extrusion, sucrose gradient centrifugation, lambda exonuclease treatment and size selection by gel electrophoresis (average size, 1 kb). The HD-26ab marker was enriched 3–10-fold in each case with no other marker showing any significant enrichment. Surprisingly, neither primer pair that slightly overlapped the HD-26ab marker (HD-24cd, nt 17239–17551 or HD-6cd, nt 17693–17906) showed any enrichment (also see Fig. 4). Since we did not observe significant IR activity immediately surrounding HD-26ab, we were particularly concerned that identification of the *HD* origin might be an artifact in some way attributable to a characteristic of the region amplified by the HD-26ab primer pair. However, agarose gel electrophoresis of PCR product made using total genomic DNA gave a single band of approximately the same intensity as that of the other *HD* markers. In addition, electrophoresis and routine PE-5700 melting profile analysis (60–94°C) confirmed that the kt–PCR product was not the

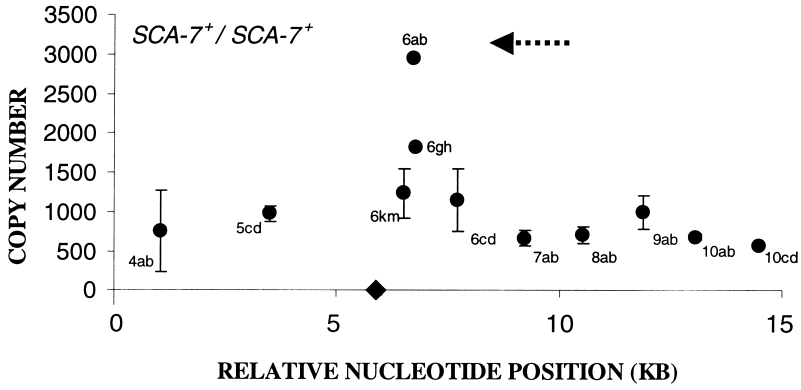

Figure 2. Abundance of nascent DNA templates isolated from an unaffected cell line using a series of *SCA-7* markers that are plotted as a function of their relative position. The absolute position of each marker and the primer sequences can be found in Supplementary information. Every marker is represented by a solid circle, each with an error bar representing the 95% confidence interval based on duplicate analyses (in some cases the variation is too small to observe on the graph). The name of each marker (without the locus prefix) is also given. The direction of transcription (arrow) and the TNR tract location (solid diamond) in the first *SCA-7* exon are indicated.

result of primer–dimer formation or some other amplification artifact. A genome-wide BLASTN search (17) of the HD-26ab primer set gave only one significant hit for each primer (in the same clone and the expected distance apart), indicating it is a single copy sequence.

In a preliminary experiment, we determined whether the candidate *HD* IR location was in the same place in a chromosome with an expanded *HD* allele. The HDG-7 cell line is derived from a rare individual homozygous for the *HD* mutation (46 and 42 repeats). Nascent strands were isolated by extrusion of total genomic DNA, sucrose gradient centrifugation, lambda exonuclease digestion and finally gel electrophoresis (size range 0.3–1.5 kb). The HD-26ab marker (Fig. 4) showed the highest enrichment (5.7-fold) suggesting that the candidate *HD* IR that we identified functions in chromosomes with both normal and expanded *HD* alleles.

The CAG repeat strand in the *HD* gene is the sense strand and codes for polyglutamine. The position of the candidate IR is upstream (relative to the direction of transcription) of the repeats. Based on this information the CAG strand is the template for Okazaki fragment synthesis during DNA replication.

Analysis of the SBMA locus

Finally, we examined the *SBMA* gene in HDG-7 cells using nascent strands isolated from total genomic DNA run on a sucrose gradient and then subjected to lambda exonuclease digestion (Fig. 5). Markers SBMA-15ab and SBMA-16cd show enrichments of ~5-fold compared with the most distal markers. The repeat tract sits between these two markers, ~550 bp from SBMA-15ab and 1 kb from SBMA-16cd. These data were verified using five independent nascent strand preparations from GM03469, HDG-7 or K562 cells and DNA fragments ranging in size from 0.3 to 1.5 kb (data not shown). The average enrichment for the initiation region was 2.8-fold (range 2–4), and in every case, SBMA-15ab and SBMA-16cd had greater enrichments than any other of the 11 markers in the same experiment (data not shown).

DISCUSSION

A review (3) of available pedigree, single sperm typing and small pool PCR data comparing the germline expandability at different TNR loci showed variation of up to 100-fold after correcting for the number of triplet repeats at each locus. One possible explanation for the variation could be the influence of functional elements adjacent to the repeats. These functional elements could affect the expansion process to different degrees at different chromosomal positions. In *E. coli* and yeast an IR is an example of such an element (1,7,9,10). Experiments examining the effect of TNR orientation in these model systems showed that an IR located so that the CAG strand is the lagging strand template during DNA replication has a greater expansion frequency than when CTG is the lagging strand template. Conversely, contractions occur more frequently when the lagging strand template contains CTG rather than CAG.

Relative to their respective candidate IRs, the *HD* and *SCA-7* repeats are in the orientation that makes the CAG strand the template for Okazaki fragment synthesis. In a preliminary study on the *HD* locus we showed that expanded alleles also have an IR at the same position. If true for expanded *SCA-7* alleles, the extensive germline expandability of the disease genes at these two loci would fit well with the repeat tract orientation model.

The germline instability at an expanded *SBMA* locus is 8–10-fold lower than at *HD* and 50–60-fold lower than *SCA-7* (3). Interestingly, our results suggest that the orientation of the repeats with respect to the candidate IR at the normal *SBMA* locus may not be constant from replication to replication. IR localization experiments of the kind described here reveal a probability distribution of initiation points for a population of dividing cells. For any individual cell entering S phase, replication of the leading strand is likely to be initiated at a single nucleotide somewhere within the IR region (13,18).

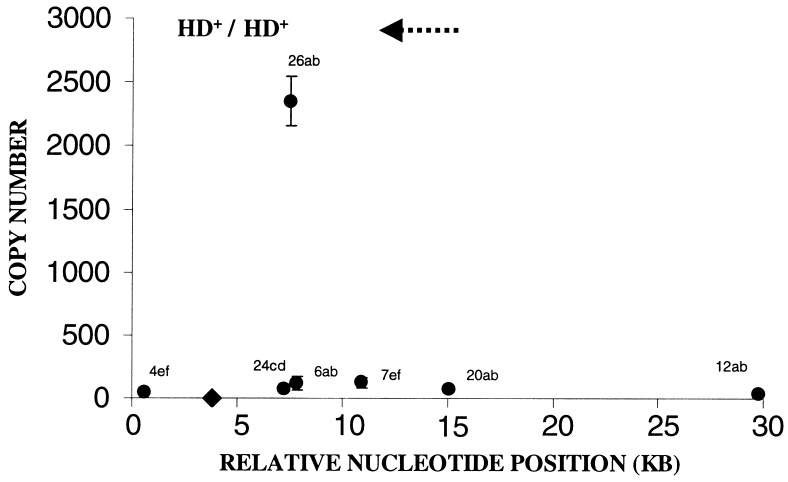

Figure 3. Abundance of nascent DNA templates isolated from an unaffected cell line (HD^{+}/HD^{+}) using a series of *HD* markers plotted as a function of their relative position. The absolute position of each marker and the primer sequences can be found in the Supplementary Material. Every marker is represented by a solid circle, each with an error bar representing the 95% confidence interval based on duplicate analyses (in some cases the variation is too small to observe on the graph). The name of each marker (without the locus prefix) is also given. The direction of transcription (arrow) and the TNR tract location (solid diamond) in the first *HD* exon are indicated. A high density of repeated DNA prevented analysis from 15 to 30 kb.

Since IR activity is found on both sides of the *SBMA* repeat tract, initiation events are expected to occur on both sides of the repeats so that sometimes the CAG strand would be a template for the lagging strand, and sometimes a template for the leading strand. If the IR location is the same in expanded *SBMA* alleles this might help to explain the lower instability at *SBMA* compared with the other two loci. Taken together, our data suggest that the replication of the SBMA locus may initiate from at least two replication initiation sites, each utilized at a particular cell cycle by some cells within the population (19–21). The emerging replication profile is similar to that found in other loci, such as the Chinese hamster *dhfr* locus, the rhodopsin locus and the murine β-globin locus (19–21).

Evidence from model systems also supports the idea that the proximity of the TNR to an IR can affect instability (8,9). Recent experiments (11) in monkey cells using a plasmid with a $(CAG/CTG)_{79}$ tract and an SV40 virus origin of replication were interpreted as showing that proximity of the viral origin to the repeats is more important than orientation. When the lagging strand template contained CAGs, only expansions were found when the SV40 origin was 103 bp away from the repeat tract. In the same orientation, most if not all mutations were contractions when the repeats were 234, 536 or 667 bp away from the origin.

Expanded *HD* and *SCA7* loci are characterized by high levels of expansion and very few contractions. The fact that this genetic behavior is compatible with the location of the TNR tract as far away as 700–900 bp or even 3.6 kb from the candidate IRs does not support the simple idea that the triplets must be less than 234 bp from the IR for expansions to always outnumber contractions.

It is too early to decide the relative role of orientation and proximity either in model systems or in humans. First, the greatest distance between repeat tract and viral origin tested in the SV40 system was only 667 bp. Whether there is any pattern to the relationship between IR location and expansion as the repeats move further away from the IR remains to be determined. Given our preliminary data on *SCA-7* and expanded *HD* alleles, distances up to at least 3.6 kb should be examined. Second, the SV40 vector is not integrated into a monkey cell chromosome and the possibility exists that some fundamental difference in replication initiation and/or Okazaki fragment processing may exist between chromosomal and SV40 IRs.

The first mechanism proposed to explain TNR expansions involved replication slippage and depended on the initiation of Okazaki fragment synthesis within the TNR region itself (4–6). Data on human germline TNR instability at the *HD* locus obtained by sperm typing was tested for its fit to a mutation mechanism that required Okazaki fragment initiation within the repeats (22). The computational analysis considered the absolute length (and variation in length) of Okazaki fragments and the length of the TNR region in calculating whether Okazaki fragment synthesis was initiated in the repeat tract (22). The actual size distribution of expanded *HD* alleles in sperm was found to fit this model very well. This computational study assumed that the repeat tract was far enough away from an IR so that the chance for an Okazaki fragment to initiate within the repeats was stochastic. Additional mathematical analysis is required to explore whether the chance of initiation within the repeats varies with the distance from the IR. In addition, Cleary *et al.* (11) have speculated that the position within the repeat tract where initiation takes place may also be important for instability. How this position varies with the distance from an IR could also be computed.

The impact of our experimental findings on the roles played by repeat tract proximity and/or orientation in human TNR instability must be viewed with some reservations. First, we used the term ‘candidate IR’ throughout to emphasize that definitive proof requires confirmation using independent IR identification technologies. This is particularly relevant to

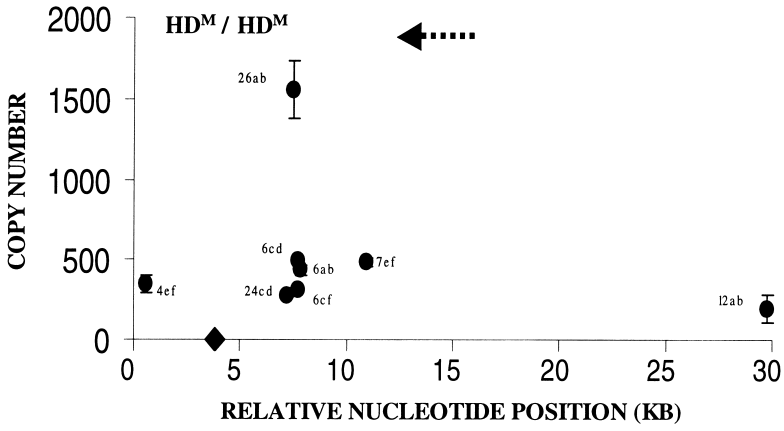

Figure 4. Abundance of nascent DNA templates isolated from a cell line carrying two mutant disease alleles at the *HD* locus (HD^M/HD^M). The *HD* markers are plotted as a function of their relative position. The absolute position of each marker and the primer sequences can be found in the Supplementary Material. Every marker is represented by a solid circle, each with an error bar representing the 95% confidence interval based on duplicate analyses (in some cases the variation is too small to observe on the graph). The name of each marker (without the locus prefix) is also given. The direction of transcription (arrow) and the TNR tract location (solid diamond) in the first *HD* exon are indicated.

the *HD* locus where we found a highly localized IR. Sharply delineated IRs are not unprecedented however and the *Lamin B2* and *HSP70* genes in human and the Ig heavy chain locus in mouse have been localized to ~600 bp or less (reviewed in 23). It has been suggested that the more care that is taken to isolate real replication intermediates (for example, eschewing the use of drugs affecting DNA metabolism, as in this paper), the more likely it is that origin activity will appear to be restricted to a narrower initiation region (24).

Second, locating an IR necessitates using cultured cells grown in the laboratory. Germline or somatic cell IRs cannot be studied in the adult mammal. However, with some notable exceptions (25), the use of a particular origin of replication does not usually vary among cell lines from different sources (26,27) (and as we have shown in this paper). Although it is likely that an origin detected in a cell line will also be used in adult tissues of the organism, proof must wait for technological advances that make it possible to test this idea directly.

Third, and in contrast to the studies on model organisms, there is no direct experimental data (including ours) demonstrating that either human germline or somatic instability (1) depends on DNA replication that accompanies cell division. Meiotic recombination and DNA damage repair-related mechanisms for instability have also been implicated based on studies in model organisms (1,2,28–40). However, if cell division-dependent replication does contribute to instability, our results could be highly relevant. In addition to inter-locus variation, the well-known paternal bias in $(CAG/CTG)_n$ tract instability (reviewed in 1,2) could be explained since male but not female germline cells continue division throughout adulthood.

Finally, another *cis*-factor proposed to explain inter-locus variation in instability at TNR loci (3) is the close proximity of a CpG-rich region. The TNR region of *SCA-7* and *HD* are included within a CpG-rich region but the repeats at the *SBMA* locus are ~1 kb away from such a region. In light of our data, it is interesting that a correlation has been drawn between the presence of CpG-rich regions and IRs (41–44). Isolation of sequences enriched for CpG islands (un-methylated CpG-rich regions usually associated with transcription start sites) are also enriched for IRs (45). At the present time, the methylation state of the CpG-rich regions in *HD, SBMA* and *SCA-7* are not known. Variation in instability among human TNR loci may be related to CpG-rich regions because of (i) a correlation between IRs and CpG islands, (ii) some other property of CpG-rich regions (e.g. transcriptional regulation) (46,47) or (iii) both (i) and (ii).

MATERIALS AND METHODS

Cells and cell culture

Cells were grown at densities between 10^5 and 10^6 cells/ml in RPMI 1640 medium (Sigma, St Louis, MO, USA) supplemented with L-glutamine, bicarbonate, penicillin/streptomycin (Sigma), and 15% heat-inactivated fetal bovine serum (Gibco Life Technologies, Grand Island, NY, USA). Three different cell lines were used: (i) erythroid leukemia cell line K562; (ii) HDG-7, a lymphoblastoid cell line from a homozygous *HD* patient; and (iii) GM03469, a normal lymphoblastoid line (Coriell Repository Camden NJ, USA).

Total genomic DNA purification

Cells were washed three times with ice cold PBS, lysed in SDS and incubated with proteinase K overnight at 37°C. Total genomic DNA was purified by a gentle phenol/chloroform extraction followed by ethanol precipitation and re-suspension in TE buffer.

Lambda exonuclease digestion

Total genomic DNA (~20 µg) was sheared by 10 passages through a $25^{3/8}$ gage needle, treated with T4 polynucleotide

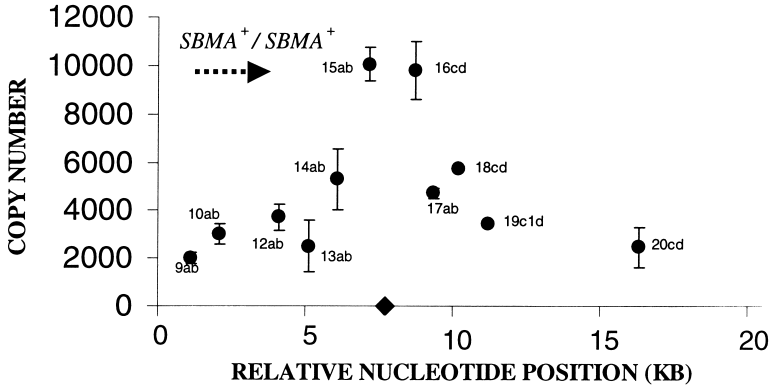

Figure 5. Abundance of nascent DNA templates isolated from an unaffected cell line using a series of *SBMA* markers plotted as a function of their relative position. The absolute position of each marker and the primer sequences can be found in Supplementary Material. Every marker is represented by a solid circle, each with an error bar representing the 95% confidence interval based on duplicate analyses (in some cases the variation is too small to observe on the graph). The name of each marker (without the locus prefix) is also given. The direction of transcription (arrow) and the TNR tract location (solid diamond) in the first *SBMA* exon are indicated.

kinase in the presence of 80 units/100 µl of RNAsin (Promega, Madison, WI, USA) and digested with lambda exonuclease (New England Biolabs, Beverly, MA, USA) overnight at 37°C (12,48). An internal control consisting of 250–500 ng of linearized pGem plasmid (Promega, Madison, WI, USA) was added to the DNA to evaluate the efficiency of the lambda exonuclease digestion. In a successful reaction, the plasmid band could not be detected in a small aliquot run on an agarose gel. The reaction mixture was phenol/chloroform extracted, ethanol precipitated and re-suspended in TE buffer before PCR.

Extrusion of nascent DNA

Total genomic DNA in TE buffer (pH 7.6) was extruded overnight at 50°C with gentle shaking (12) to allow the complementary nascent strands in replication bubbles to anneal.

Sucrose gradient centrifugation

Extruded DNA was loaded directly on 5–30% sucrose gradients and centrifuged at 15–20°C for 18 h at 24 k rpm in a SW27 rotor (14) using a L5-65 ultracentrifuge. Fractions were collected and a small sample of each fraction was analyzed by agarose gel electrophoresis to determine the size range. The appropriate fractions were pooled.

Gel purification of nascent DNA fragments

DNA fragments from the desired fractions were separated on a 1% agarose gel at 30–50 V/13 cm gel. The appropriate size range was excised and recovered using the Qiaex II purification kit (Qiagen, Valencia, CA, USA). The size range selected was between 0.3 and 1.5 kb although in a few cases the largest fragments were 3 kb.

Kinetic PCR

For any one experiment identical amounts of nascent strand DNA was used for amplification of each of the markers in a kinetic-PCR reaction using a PE 5700 thermocycler. Typical standards for quantifying the starting amount of each marker template included 20, 5 and 1 ng of total genomic DNA from the same cell line. In every experiment all samples (including standards) were run in duplicate. A typical PCR program included a hold step at 94°C for 3 min, then 29–31 cycles at 94°C for 30 s, 60°C for 30 s and 72°C for 1 min 10 s. The final extension step (72°C for 5 min) was followed by melting profile analysis (60–94°C) to confirm that a single product was produced with the T_m characteristic of the expected product. Primer pairs were used only if agarose gel electrophoresis showed the products were of the expected size rather than the result of primer–dimer formation or some other amplification artifact. The primer sequences can be obtained in Supplementary Material. For primer design, Repeat Masker (http://repeatmasker.genome.washington.edu/) was used to filter out repetitive sequences from the regions of interest. Primer pairs were designed so that most could be run under the same conditions. Fifty-microliter reactions were carried out in FailSafe PreMix with Buffer E (Epicentre technologies, Madison, WI, USA) containing 1×SYBR Green (PE Biosystems Foster city CA, USA) and 2 µM ROX (PE Biosystems Foster city CA, USA). Each primer was used at a concentration of 0.5 µM. For efficient amplification of some targets with a high GC content, additions of DMSO (up to a final concentration of 5%) were required. For every primer pair studied, the same reaction mixture and PCR conditions were used for both the standards and the nascent DNA to ensure equal amplification efficiency and thus accurate estimation of the template copy in the nascent DNA samples.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

The authors acknowledge helpful discussions on kt–PCR with Russ Higuchi and Bob Watson (Roche Molecular Systems) and

grants from the NIGMS (N.A.), NINDS and the Hereditary Disease Foundation (J.F.G. and N.S.W.).

REFERENCES

1. Wells, R.D., Warren, S.T. and Sarmiento, M. (1998) *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA.
2. Cummings, C.J. and Zoghbi, H.Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.*, **9**, 909–916.
3. Brock, G.J., Anderson, N.H. and Monckton, D.G. (1999) *Cis*-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum. Mol. Genet.*, **8**, 1061–1067.
4. Eichler, E.E., Holden, J.J., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A. and Nelson, D.L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.*, **8**, 88–94.
5. Richards, R.I. and Sutherland, G.R. (1994) Simple repeat DNA is not replicated simply. *Nat. Genet.*, **6**, 114–116.
6. Kunst, C.B. and Warren, S.T. (1994) Cryptic and polar variation of the fragile-X repeat could result in predisposing normal alleles. *Cell*, **77**, 853–861.
7. Freudenreich, C.H., Stavenhagen, J.B. and Zakian, V.A. (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell Biol.,* **17**, 2090–2098.
8. Kang, S., Jaworski, A., Ohshima, K. and Wells, R.D. (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli. Nat. Genet.*, **10**, 213–218.
9. Maurer, D.J., O'Callaghan, B.L. and Livingston, D.M. (1996) Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **16**, 6617–6622.
10. Miret, J.J., Pessoa-Brandao, L. and Lahue, R.S. (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **95**, 12438–12443.
11. Cleary, J.D., Nichol, K., Wang, Y.H. and Pearson, C.E. (2002) Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.*, **31**, 37–46.
12. Tao, L., Dong, Z., Leffak, M., Zannis-Hadjopoulos, M. and Price, G. (2000) Major DNA replication initiation sites in the C-Myc locus in human cells. *J. Cell. Biochem.*, **78**, 442–457.
13. Gerbi, S.A. and Bielinsky, A.-K. (1997) Replication initiation point mapping. *Meth. Compan. Meth. Enzymol.*, **13**, 271–280.
14. Giacca, M., Pelizon, C. and Falaschi, A. (1997) Mapping replication origins by quantifying relative abundance of nascent DNA strands using competitive polymerase chain reaction. *Meth. Compan. Meth. Enzymol.*, **13**, 301–312.
15. Higuchi, R., Fockler, C., Dollinger, G. and Watson, R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, **11**, 1026–1030.
16. Malot, M. and Leffak, M. (1999) Activity of the C-Myc replicator at an ectopic chromosomal location. *Mol. Cell. Biol.*, **19**, 5685–5695.
17. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.*, **25**, 3389–3402.
18. Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M. and Falaschi, A. (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science*, **287**, 2023–2026.
19. Aladjem, M.I., Rodewald, L.W., Lin, C.M., Bowman, S., Cimbora, D.M., Brody, L.L., Epner, E.M., Groudine, M. and Wahl, G.M. (2002) Replication initiation patterns in the beta-globin loci of totipotent and differentiated murine cells: evidence for multiple initiation regions. *Mol. Cell Biol.*, **22**, 442–452.
20. Dijkwel, P.A., Mesner, L.D., Levenson, V.V., d'Anna, J. and Hamlin, J.L. (2000) Dispersive initiation of replication in the Chinese hamster rhodopsin locus. *Exp. Cell Res.*, **256**, 150–157.
21. Bogan, J.A., Natale, D.A. and Depamphilis, M.L. (2000) Initiation of eukaryotic DNA replication: conservative or liberal? *J. Cell Physiol.*, **184**, 139–150.
22. Leeflang, E.P., Tavare, S., Marjoram, P., Neal, C.O., Srinidhi, J., MacFarlane, H., MacDonald, M.E., Gusella, J.F., de Young, M., Wexler, N.S. *et al.* (1999) Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum. Mol. Genet.*, **8**, 173–183.
23. Todorovic, V., Falaschi, A. and Giacca, M. (1999) Replication origins of mammalian chromosomes: the happy few. *Front. Biosci.*, **4**, D859–868.
24. Li, C.J., Bogan, J.A., Natale, D.A. and DePamphilis, M.L. (2000) Selective activation of pre-replication complexes in vitro at specific sites in mammalian nuclei. *J. Cell Sci.*, **113** (Pt 5), 887–898.
25. Ermakova, O.V., Nguyen, L.H., Little, R.D., Chevillard, C., Riblet, R., Ashouian, N., Birshtein, B.K. and Schildkraut, C.L. (1999) Evidence that a single replication fork proceeds from early to late replicating domains in the IgH locus in a non-B cell line. *Mol. Cell*, **3**, 321–330.
26. Kumar, S., Giacca, M., Norio, P., Biamonti, G., Riva, S. and Falaschi, A. (1996) Utilization of the same DNA replication origin by human cells of different derivation. *Nucl. Acids Res.*, **24**, 3289–3294.
27. Tao, L., Nielsen, T., Friedlander, P., Zannis-Hadjopoulos, M. and Price, G. (1997) Differential DNA replication origin activities in human normal skin fibroblast and HeLa cell lines. *J. Mol. Biol.*, **273**, 509–518.
28. Ashley, C.T. Jr and Warren, S.T. (1995) Trinucleotide repeat expansion and human disease. *A. Rev. Genet.*, **29**, 703–728.
29. Cohen, H., Sears, D.D., Zenvirth, D., Hieter, P. and Simchen, G. (1999) Increased instability of human CTG repeat tracts on yeast artificial chromosomes during gametogenesis. *Mol. Cell Biol.*, **19**, 4153–4158.
30. Fortune, T.M., Vassilopoulos, C., Coolbaugh, M.I., Siciliano, M.J. and Monckton, D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.*, **9**, 439–445.
31. Jakupciak, J.P. and Wells, R.D. (2000) Genetic instabilities of triplet repeat sequences by recombination. *IUBMB Life*, **50**, 355–359.
32. Jankowski, C., Nasar, F. and Nag, D.K. (2000) Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. *Proc. Natl Acad. Sci. USA*, **97**, 2134–2139.
33. Lia, A.S., Seznec, H., Hofmann-Radvanyi, H., Radvanyi, F., Duros, C., Saquet, C., Blanche, M., Junien, C. and Gourdon, G. (1998) Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum. Mol. Genet.*, **7**, 1285–1291.
34. McMurray, C.T. (1995) Mechanisms of DNA expansion. *Chromosoma*, **104**, 2–13.
35. Mitas, M. (1997) Trinucleotide repeats associated with human disease. *Nucl. Acids Res.*, **25**, 2245–2254.
36. Monckton, D.G., Cayuela, M.L., Gould, F.K., Brock, G.J., Silva, R. and Ashizawa, T. (1999) Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. *Hum. Mol. Genet.*, **8**, 2473–2478.
37. Paques, F. and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.*, **63**, 349–404.
38. Seznec, H., Lia-Baldini, A.S., Duros, C., Fouquet, C., Lacroix, C., Hofmann-Radvanyi, H., Junien, C. and Gourdon, G. (2000) Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum. Mol. Genet.*, **9**, 1185–1194.
39. Sinden, R.R. (1999) Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am. J. Hum. Genet.*, **64**, 346–353.
40. Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M. and Bates, G.P. (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.*, **15**, 197–200.
41. Delgado, S., Gomez, M., Bird, A. and Antequera, F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J.*, **17**, 2426–2435.
42. Araujo, F.D., Knox, J.D., Ramchandani, S., Pelletier, R., Bigey, P., Price, G., Szyf, M. and Zannis-Hadjopoulos, M. (1999) Identification of initiation sites for DNA replication in the human dnmt1 (DNA-methyltransferase) locus. *J. Biol. Chem.*, **274**, 9335–9341.
43. Rein, T., Zorbas, H. and DePamphilis, M.L. (1997) Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Mol. Cell Biol.*, **17**, 416–426.

44. Phi-van, L. and Stratling, W.H. (1999) An origin of bidirectional DNA replication is located within a CpG island at the 3′′ end of the chicken lysozyme gene. *Nucl. Acids Res.*, **27**, 3009–3017.
45. Antequera, F. and Bird, A. (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr. Biol.*, **9**, R661–667.
46. Robertson, K.D. and Jones, P.A. (1997) Dynamic interrelationships between DNA replication, methylation, and repair. *Am. J. Hum. Genet.*, **61**, 1220–1224.
47. Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression—belts, braces, and chromatin. *Cell*, **99**, 451–454.
48. Cimbora, D.M., Schubeler, D., Reik, A., Hamilton, J., Francastel, C., Epner, E.M. and Groudine, M. (2000) Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region. *Mol. Cell. Biol.*, **20**, 5581–5591.