SHORT REPORT

Attempts to detect retrotransposition and *de novo* deletion of Alus and other dispersed repeats at specific loci in the human genome

Caroline R Hollies[1], Darren G Monckton[2] and Alec J Jeffreys*,[3]

[1]Department of Pathology, University of Wales College of Medicine, Wales, UK; [2]Division of Molecular Genetics, University of Glasgow, Glasgow, Scotland, UK; [3]Department of Genetics, University of Leicester, Leicester, UK

Dispersed repeat elements contribute to genome instability by *de novo* insertion and unequal recombination between repeats. To study the dynamics of these processes, we have developed single DNA molecule approaches to detect *de novo* insertions at a single locus and Alu-mediated deletions at two different loci in human genomic DNA. Validation experiments showed these approaches could detect insertions and deletions at frequencies below 10^{-6} per cell. However, bulk analysis of germline (sperm) and somatic DNA showed no evidence for genuine mutant molecules, placing an upper limit of insertion and deletion rates of 2×10^{-7} and 3×10^{-7}, respectively, in the individuals tested. Such re-arrangements at these loci therefore occur at a rate lower than that detectable by the most sensitive methods currently available. *European Journal of Human Genetics* (2001) 9, 143–146.

Introduction

Human retroposons, in particular Alu and L1 elements, play a major role in genome remodelling, both by *de novo* insertion and by promoting deletions and re-duplications as a result of unequal recombination between dispersed repeat elements.[1] Mechanisms of re-arrangement, particularly in retroposition, are fairly well established;[2] in contrast, little is known about the dynamics of these events. *De novo* insertions are usually of complete Alu and partial L1 elements,[1,2] while unequal exchanges occur mainly between Alu elements.[3–5] Both types of event are thought to occur mainly in the germline,[2] though examples of both have been observed in somatic tissues.[3,6] There is evidence that at least some unequal exchange events may be due to meiotic recombination.[4] To date, such insertions and re-arrangements have only been characterised in patients, resulting in considerable uncertainties about the nature of these processes. We have

therefore developed highly sensitive methods for mutation detection, analogous to methods devised for studying *de novo* mutation at highly unstable minisatellites,[7] and have used these in an attempt to detect dispersed repeat-mediated instability in somatic and germline DNA at three different loci.

Methods

PCR amplification

Primers used were: MS1-1, 5′-GCT-GCT-GTG-CCT-TCC-CCG-GTT-C-3′; MS1-2, 5′-ACC-CAC-TGA-CTC-ACA-GCA-AGG-G-3′; MS1-3 5′-ggc-gga-tcc-cgg-gaa-TTC-AGT-CTC-CTA-GTG-AAG-GTG-C-3′; MS1-4, 5′-gcg-aat-tcg-gat-ccc-GGT-GAT-TTC-TCA-ATT-TGC-TGG-3′ (lowercase – extensions incorporating restriction endonuclease sites for subsequent cloning): 32-5F, 5′-CAG-TGC-TTG-GCA-CAT-AAT-GAG-CAC-3′; 32-5NF, 5′-CTC-TTT-CTA-GAA-GCC-GTT-AGA-GGA-G-3′; 32-5R, 5′-GAA-TCC-TAC-ATG-TAG-GCG-AGC-AGT-3′; 32-1.6F, 5′-GCC-AAC-AGT-GTA-CTT-TGA-AGA-GCA-3′; 32-1.3NR, 5′-GCA-GGT-AGA-TAG-TGG-CCC-AGA-G-3′; 32-1.3R, 5′-CCA-GGT-TCT-GGG-GTG-ACT-AG-3′: C1+2.4F, 5′-CCC-AAC-AGA-TTC-TCC-TAC-CC-3′; C1+2.4NF 5′-GCC-CAC-TAC-

TGG-GTC-CTT-CTG-CCC-AG-3'; C1+2.8R, 5'ATA-CAT-CCC-TCT-ACC-CCA-CC-3'; C1+6.5F, 5'-TGG-GTC-AAA-GGA-GTC-TTG-G-3'; C1+6.8NR, 5'-ATG-GAC-CCT-GGC-TGA-GGC-TGG-GTG-3'; C1+6.9R, 5'GCA-GCA-GAA-CTG-CTT-TCA-AAC-3'. MS32 and *C1NH* outside and nested PCR reactions employed a long PCR protocol.[7] Locus-specific probes (c. 300 bp long) were synthesised at MS32 by amplification of 1 ng MS32 cosmid DNA[8] using primer pairs 32-5NF and 32-5R (probe 1), and 32-1.6F and 32-1.3NR (probe 2) for 20 cycles, denaturing at 94°C for 15 s, annealing at 65°C for 30 s and 1 min extension at 70°C; and at *C1NH* by amplification of 200 ng genomic DNA using primers C1+2.4NF and C1+2.8R (probe 1), and C1+6.5F and C1+6.8NR (probe 2) for 30 cycles, denaturing at 96°C for 1 min, annealing at 62°C for 30 s and 1 min extension at 70°C. At MS32 semi-nested PCR was carried out using primers 32-5F and 32-1.3NR for 26 cycles, denaturing at 96°C for 15 s, annealing at 68°C for 30 s and 6 min extension at 70°C. Nested PCR at *C1NH* was carried out using primers C1+2.4NF and C1+6.8NR for 28 cycles, denaturing at 94°C for 15 s, and annealing and extending at 70°C for 6 min.

Size fractionation of genomic DNA to detect insertion events

A total of 25 μl sperm DNA was amplified in five 100-μl reactions for 12 cycles, denaturing at 96°C for 1 min, annealing at 63°C for 30 s and 4 min extension at 70°C, using the primers MS1-1 and MS1-2 to give a progenitor PCR product of 1.2 kb. Following amplification, 500 ng of pAT153-*Dde*I/*Bam*HI marker DNA was added to each sample as an internal control and separated by agarose gel electrophoresis. DNA between 1.4 and 9.5 kb was size-fractionated[7] to enrich for alleles containing insertions of between 200 and 7000 bp (consensus Alu and L1 elements have lengths of c. 300 and 6000 bp respectively). Fifty per cent of the purified DNA was re-amplified for between 12–16 cycles such that the residual progenitor allele was amplified to levels no greater than 1 ng, and again size fractionated. These PCR and size fractionation steps were repeated twice more using the outside primers, and four times with the nested primers. A sample of each reaction was analysed by agarose gel electrophoresis and ethidium bromide staining, followed by Southern hybridisation with ^{32}P-labelled total human DNA. Reactions were scaled down for subsequent analysis of 10-μg samples.

Size fractionation and screening of genomic DNA to detect deletion mutants

Genomic DNA was digested with restriction enzymes which released the target region (*Nco*I+*Kpn*I for MS32, *Dra*I for *C1NH*; Figure 2) and size-fractionated by agarose gel electrophoresis[7] to eliminate most or all progenitor molecules (c. 4.4 kb) and collect all possible Alu-Alu deletion mutant molecules (1.4–4.3 kb long). Multiple aliquots of each fraction, each containing DNA derived from 8–30 × 10[4]

haploid genomes, were screened for deletion mutants in 7-μl reactions by long PCR. MS32 amplifications were carried out using primers 32-5F and 32-1.3R for 26 cycles, denaturing at 94°C for 30 s, annealing at 68°C for 30 s and 6 min extension at 70°C; *C1NH* amplifications were carried out using primers C1+2.4F and C1+6.9R for 32 cycles, denaturing at 94°C for 15 s, and annealing and extending at 68°C for 5 min. Half of each reaction was analysed by agarose gel electrophoresis and Southern hybridisation with both ^{32}P-labelled end probes.

Results

Analysis of *de novo* insertions

The 1.2 kb target flanking the human minisatellite MS1 (*D1S7*) contains a number of short AT-rich potential integration sites[9] (Figure 1). To detect transposition events into this locus, size enrichment for target molecules longer than the progenitor (Methods) was carried out on two sperm DNA samples (1 × 25 μg plus 1 × 10 μg) and 10-μg samples of two blood DNAs, two colon tumour DNAs and two placental DNAs. No evidence for abnormal length mutants was detected. This technique was validated by amplifying two 100-μl reactions each containing 10 μg sperm DNA plus 1 μl of a 2 molecules/μl dilution of the pseudoinsertion allele (Figure 1) with the nested primers MS1-3 and MS1-4, then performing two cycles of size fractionation and re-amplification. Analysis of the final PCR products by hybridisation with a pseudoinsertion specific probe showed that a single molecule of the synthetic mutant could be recovered from a background of c. 3.3×10^6 normal molecules.
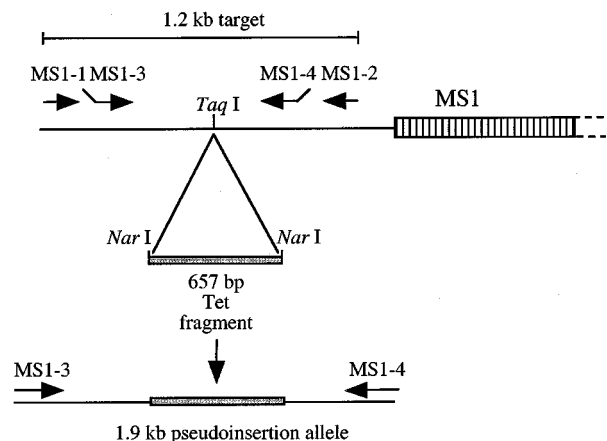

Figure 1 The MS1 target region for insertion analysis. PCR primers are shown as arrows. The 657 bp Tet fragment (shaded box) was isolated from plasmid pBR322 vector by *Nar*I digestion. This fragment was ligated to the target PCR product, amplified from human DNA with primers MS1-3 plus MS1-4 and digested with *Taq*I (compatible with *Nar*I sticky ends). The 1.9 kb pseudoinsertion allele was recovered by amplifying the ligation products with primers MS1-3 plus MS1-4, using the same PCR conditions as for primers MS-1 and MS-2 (Methods).
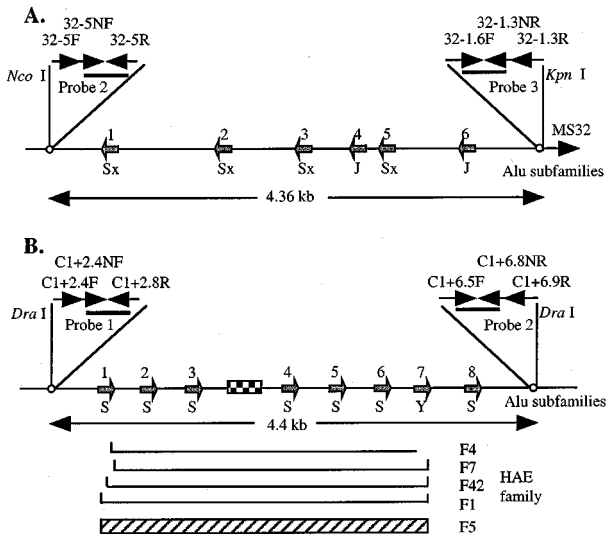
Figure 2 Alu-rich regions of the human genome analysed for instability. The locations and orientation of Alu elements are shown, together with their assigned number and Alu subfamily classification. The positions of PCR primers, end probes and restriction sites used in size enrichment are indicated above. (**A**) The MS32 target region upstream of minisatellite MS32[15]. (**B**) The C1NH gene target spanning intron 3, exon 4 and intron 4[16]. The location and extent of different deletions (brackets) and a duplication (hatched bar) in HAE patients[5] are shown below.

Analysis of Alu-mediated deletions

The first target lies 1.3–5 kb upstream of the end of human minisatellite MS32 (*D1S8*),[8] known to be active in meiotic recombination[10] (Figure 2A). The second target was a 4.4 kb region around exon 4 of the C1-inhibitor gene (*C1NH*, Figure 2B). Twelve out of 18 deletion breakpoints identified in type 1 hereditary angioedema (HAE) patients have been localised to Alu elements within this region.[5]

Evolutionary and population genetic estimates of instability

To investigate instability over evolutionary time, both loci were PCR amplified from human, chimpanzee, gorilla and orang-utan DNA using semi-nested (MS32) or nested (*C1NH*) primers and analysed by restriction enzyme digestion (Figure 3). The MS32 target showed no length changes in any of the species, and all changes in DNA fragment profiles could be attributed to restriction site gains and losses by base substitution. Similar analysis of the *C1NH* target showed no evidence for length variation except for a 200-bp size difference between orang-utan and the other three primates. Sequencing showed that this was the result of a 200-bp deletion entirely within Alu 2. Neither target has therefore altered its Alu content over approximately 34 million years of primate evolution (7 million years since the divergence of human, chimpanzee and gorilla, and 10 million years since the divergence of orang-utan from this lineage).[11] Assuming
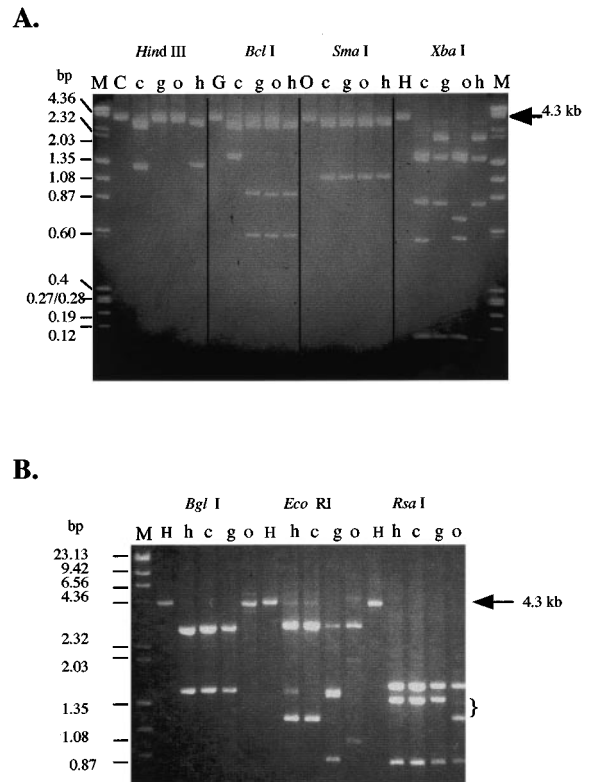

Figure 3 Comparison of the Alu-rich MS32 and C1NH targets in humans and apes. Target regions were amplified from human (H), chimpanzee (C), gorilla (G) and orang-utan (O) DNA using primers 32-5F & 32-1.3NR and C1+2.4NF & C1+6.8NR respectively and analysed by agarose gel electrophoresis either before (uppercase) or after (lowercase) digestion with the restriction enzymes indicated. M, markers = λDNA $\times$ *Hind*III + ϕX174 $\times$ *Hae*III. (**A**) Comparative digestion of the MS32 region. (**B**) Comparative digestion of the C1NH region. Bracket highlights the *Rsa*I DNA fragments showing the 0.2 kb deletion in the orang-utan C1NH target region.

a generation time of 10 years, this implies that selectively-neutral Alu-mediated re-arrangements in this region arise at a rate of $<0.9 \times 10^{-6}$ per generation (upper 95% C.I.).[12]

To investigate instability in human populations, both targets were PCR amplified from 20 unrelated Caucasians for MS32 and 39 unrelated Caucasians for *C1NH*. Both loci were monomorphic in size in all individuals tested (data not shown). Assuming an effective population size of 10 000 for humans, this suggests a rate in Caucasians for selectively neutral Alu-mediated re-arrangements of $<4.4 \times 10^{-6}$ per generation at MS32 and $<2.1 \times 10^{-6}$ at *C1NH* (upper 95% C.I.'s).

Analysis of deletion mutants in sperm and blood DNA

Size enrichment methods were then used in an attempt to detect *de novo* mutant molecules in sperm and blood DNA

that had arisen by Alu-mediated deletion (Methods). A survey of DNA from 10^7 sperm and 4.3×10^6 sperm at MS32 and *C1NH* respectively revealed no deletion mutants at either locus. Similar analysis of somatic (blood) DNA showed no deletion mutants at either locus in DNA from 2.5×10^6 diploid cells. The enrichment technique was validated for *C1NH* by mixing 4 μg sperm DNA with 6 pg (equivalent to one diploid genome) genomic DNA from HAE patient F1 (Figure 2B).[5] Size fractionation and screening revealed a single molecule of the deletion mutant of the correct size (1.2 kb), establishing that rare deletions can be detected in a background of 1.3×10^6 normal molecules (data not shown).

Discussion

Systems have been developed for detecting very rare insertion and deletion events using physical size-selection of DNA and PCR recovery of single molecules. Despite the sensitivity of these methods, no *de novo* insertion or deletion mutants were detected. The absence of insertion events in 50 μg of human sperm DNA indicates that the rate of *de novo* transposition into the 1.2-kb MS1 target is very low ($<2 \times 10^{-7}$ per sperm (upper 95% C.I.) and $<1 \times 10^{-6}$ in the other cell types investigated). Similarly, the lack of deletions in either of the Alu-rich loci investigated implies deletion rates in sperm of $<3 \times 10^{-7}$ at MS32 and $<7 \times 10^{-7}$ per sperm at *C1NH*, with maximum frequencies of somatic mutation at both of these loci of $<10^{-6}$ per cell.

The failure to detect any large insertions into the MS1 target region may be because the target is nonoptimal for integration of retroposons, or alternatively reflects an extremely low overall rate of transposition in the human genome. The latter possibility is supported by indirect estimates[13] of Alu and L1 insertion, calculated to be approximately 3.3×10^{-11}/kb target/generation, far lower than the upper limits obtained by sperm DNA analysis.

The frequency of gross genomic re-arrangement within the *C1NH* gene, estimated from the rate of occurrence of HAE, is between 7.6×10^{-5} and 1.5×10^{-6} per gamete.[14] These rates contrast with the population estimate of selectively neutral deletion rate of $<1 \times 10^{-6}$ per gamete and the direct estimate of overall deletion rate in sperm of $<7 \times 10^{-7}$. This inconsistency suggests, first, that most or all Alu-mediated re-arrangements in the *C1NH* gene are selectively disadvantageous, and second, that these re-arrangements may occur preferentially in the female germline. The stability of the Alu-rich region associated with MS32 can be explained by recent evidence demonstrating that the target locus investigated here lies within a region of intense linkage disequilibrium, within which meiotic crossover is strongly suppressed (M Panayi and AJ Jeffreys, unpublished data).

This work has demonstrated that the most sensitive germline mutation detection systems currently available are unsuitable for the isolation of re-arrangements occurring at a

rate below 10^{-7}. These difficulties highlight the need for more powerful and flexible systems to detect extremely low frequency re-arrangements at the single molecule level. Until such methods are developed, ascertainment of mutation in inherited disease will provide the only means of accessing such events in the human germline.

References

1 Deininger PL, Batzer MA: Alu repeats and human disease. *Mol Genet Metab* 1999; **67**: 183 – 193.
2 Kazazian HH, Moran JV: The impact of L1 retrotransposons on the human genome. *Nature Genet* 1999; **19**: 19 – 24.
3 Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM: The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. *Hum Mol Genet* 1998; **7**: 767 – 776.
4 Rouyer F, Simmler MC, Page DC, Weissenbach J: A sex-chromosome rearrangement in a human XX male caused by Alu-Alu recombination. *Cell* 1987; **51**: 417 – 425.
5 Stoppa-Lyonnet D, Duponchel C, Meo T *et al*: Recombinational biases in the rearranged C1-inhibitor genes of hereditary angioedema patients. *Am J Hum Genet* 1991; **49**: 1055 – 1062.
6 Miki Y, Nishisho I, Horii A *et al*: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* 1992; **52**: 643 – 645.
7 Jeffreys AJ, Neil DL, Neumann R: Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J* 1999; **17**: 4147 – 4157.
8 Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ: Characterisation of a panel of highly variable minisatellites cloned from human DNA. *Ann Hum Genet* 1987; **51**: 269 – 288.
9 Jurka J: Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci USA* 1997; **94**: 1872 – 1877.
10 Jeffreys AJ, Murray J, Neumann R: High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol Cell* 1998; **2**: 267 – 273.
11 Koop BF, Goodman M, Xu P, Chan K, Slightom JL: Primate η-globin DNA sequences and man's place among the great apes. *Nature* 1986; **319**: 234 – 238.
12 Kimura M: *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, 1983.
13 Kazazian HH: An estimated frequency of endogenous insertional mutations in human. *Nature Genetics* 1999; **2**: 130.
14 Hollies CR: The investigation of variation and *de novo* mutation in the human genome: PhD thesis. University of Leicester, 1999.
15 Murray J, Buard J, Neil DL *et al*: Comparative sequence analysis of human minisatellites showing meiotic repeat instability. *Genome Res* 1999; **9**: 130 – 136.
16 Carter PE, Duponchel C, Tosi M *et al*: Complete nucleotide sequence of the gene for human C1-inhibitor with an unusually high density of Alu elements. *Eur J Biochem* 1991; **197**: 301 – 308.