Mutation mechanisms that underlie turnover of a human telomere-adjacent segmental duplication containing an unstable minisatellite.

Mark Hills[a], Jennie N. Jeyapalan, Jennifer L. Foxon and Nicola J. Royle

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK.

[a] M. Hills current address: Terry Fox Laboratory, 675 West 10th Avenue, Vancouver, B.C., Canada V5Z 1L3

Corresponding author – NJ Royle
email: njr@le.ac.uk
Tel: +44 (0)116-2522270
Fax: +44 (0)116-2523378

Abstract

Subterminal regions, juxtaposed to telomeres on human chromosomes, contain a high density of segmental duplications but relatively little is known about the evolutionary processes that underlie sequence turnover in these regions. We have characterised a segmental duplication adjacent to the Xp/Yp telomere, each copy containing a hypervariable array of the DXYS14 minisatellite. Both DXYS14 repeat arrays mutate at a high rate (0.3% and 0.2% per gamete) but linkage disequilibrium analysis across 27 SNPs and a direct crossover assay show that recombination during meiosis is suppressed. Therefore instability at DXYS14a and b is dominated by intra-allelic processes or possibly conversion limited to the repeat arrays. Furthermore some chromosomes (14%) carry only one copy of the duplicon, including one DXYS14 repeat array that is also highly mutable (1.2% per gamete). To explain these and other observations, we propose there is another low rate mutation process that causes copy number change of part or all of the duplicon.

155 words

Keywords: telomere, subterminal, minisatellite, VNTR, segmental duplication, recombination

Introduction

The distribution of homologous crossover events across human autosomes is not random during meiosis but clustered into hotspots 1-2kb in length that are separated by regions that show strong linkage disequilibrium (LD) [1] [2] [3]. The pseudoautosomal region (PAR1) on the short arm of the sex chromosomes (Xp/Yp) spans 2.7Mb of DNA from a Y specific Alu element to the Xp/Yp telomere. This region is atypical in that it supports an obligatory crossover in male meiosis and therefore the overall male recombination rate in PAR1 is 20 times greater than the genome average [4] [5] [6]. However, LD and recombination analysis around the SHOX gene (~500kb from the Xp/Yp telomere) has shown that most crossovers in this region of PAR1 are clustered in a recombination hotspot [7]. The 1kb sequence immediately adjacent to the Xp/Yp telomere contains a high density of single nucleotide polymorphisms (SNPs; 1 per 65bp among individuals from northern and western Europe) and insertion / deletion polymorphisms that show almost complete LD such that only three common haplotypes (A, B and C) are found in this population [8]. The strong LD suggests that this region of the genome, immediately adjacent to the Xp/Yp telomere, is inert for meiotic recombination. Similar properties of high SNP density and strong LD have been found adjacent to the 12q telomere [9] suggesting this may be a general feature of telomere-adjacent regions.

The Xp/Yp and other autosomal telomeres contain many sequence-variant telomere repeat units at the beginning of the repeat array. The distribution of the variant repeats is hypervariable between alleles [8] [10] [9] indicating that telomeres have a high germline mutation rate, measured at 0.6% per gamete [11]. Telomeres with related variant-repeat distributions define lineages that are in strong linkage disequilibrium with SNPs in the telomere adjacent sequence and this indicates that variability within the telomere repeat array arises from intra-allelic processes such as sister-chromatid exchange and replication slippage. The more internal subtelomeric domains of most human chromosomes contain complex patchworks of DNA sequences that are shared between non-homologous chromosomes (inter-chromosomal segmental duplications) [12] [13]. The duplicated segments vary considerably in length and in their composition of dispersed repeats, unique sequences and genes. This subtelomeric sequence organisation must have arisen recently through inter-chromosomal

exchanges (ectopic recombination) including translocations that have likely relocated subterminal sequences and their associated telomeres [10] [14] [15] [16].

The dynamics of subterminal regions up to and including the telomere are complex but it is possible that there is a barrier to homologous recombination close to the start of human telomeres. To explore this we have extended our analysis of the Xp/Yp telomere-adjacent sequence to incorporate the DXYS14 locus [17] [18] . DXYS14 comprises a hypervariable minisatellite (HSVNTR, detected by the probe 29c1, [19]) based on a 31bp GC-rich repeat unit and with an estimated mutation rate of 1.3% per gamete. Sequences homologous to the DXYS14 minisatellite are present in the chimpanzee, gorilla and orang-utan genomes though their locations with respect to the telomere are unknown [20].

Here we present evidence that the sequence adjacent to the Xp/Yp telomere includes a small segmental duplication, each copy containing a hypervariable DXYS14 repeat array. There is some evidence that one copy of this duplicon is part of a larger segmental duplication whose sequence organisation has not been fully determined (DC1321 in TCAG genomic duplications database). Therefore our analysis is confined to the sequence in the terminal 9kbp at XpYp where the DXYS14 containing duplicons are separated by 0.8kb of sequence rich in dispersed repeats. Both copies of the DXYS14 repeat arrays mutate in the male germline and yet they are embedded in a region of strong LD that extends from the telomere across both copies of the duplication. However, 14% of chromosomes carry only a single copy of the entire duplication unit, containing one minisatellte repeat array (DXYS14c) that also has a high mutation rate. We present data that indicates that a few chromosomes show additional copy number variation and carry three DXYS14 repeat arrays. The implications of these findings for mutation mechanisms within this Xp/Yp subterminal region are discussed.

Results

To extend the analysis of the region adjacent to the Xp/Yp telomere sequence information was taken from a clone containing the Xp/Yp telomere-adjacent DNA

[21] [8]; from a half-YAC clone (M57751, [22]); from a cosmid clone (29c1) containing part of the DXYS14 locus (X17009, [19]) and from sequence available in the GenBank and EMBL databases as part of the human genome project (accession number BX640545).

The Xp/Yp telomere-adjacent sequence is annotated such that the first base internal to the telomere repeat array is -1 (chrXY: 39 in Human Genome Sequence (HGS) March 2006 assembly). Dotplot analysis of ~10kb of sequence immediately adjacent to the Xp/Yp telomere confirmed preliminary PCR results (using primers 29c1B and Tsk8W, Figure 1a) indicating the presence of a sequence duplication (data not shown). The copy of the duplication closest to the telomere (A) extends from the –2238 position (chrXY: 2277 HGS March2006) and comprises 1863bp. The second copy of the duplication (B) begins at position -4954 (chrXY: 5667 HGS March2006) from the telomere and spans 2871bp. Both copies of the duplication contain sequences probably not found elsewhere in the genome, sequences that show partial homology to a family of human endogenous retroviruses (HERVs), Alu sequences and a small array of telomere-like repeats. The duplicated units differ by three insertion/deletions that result in the internal copy (B) containing an additional 1008bp of mainly unique DNA sequence. In addition both copies of the duplication contain a variable length array of 31bp GC-rich repeats that comprise the DXYS14 minisatellite (from here on referred to as DXYS14a and DXYS14b in copy A and B of the duplication respectively). The two copies of the duplication are separated by 839bp of sequence that includes two partial Alu elements and part of an LTR (Figure 1a). In addition to the insertion/deletion differences between the two copies of the duplication they show a 4% sequence divergence indicating the duplication is ancient. Furthermore, PCR analysis of DNAs from 80 individuals (parents in the CEPH family DNA panel) with the 29c1B and Tsk8W primers, which generate amplicons of different lengths from the two copies of the duplication, showed that some chromosomes carry only a single copy of the duplication unit (see below).

SNP detection adjacent to the Xp/Yp telomere.

A high density of DNA sequence polymorphisms in the terminal 1.2kb adjacent to the Xp/Yp telomere has been reported previously [8] [9]. Sequence analysis was

undertaken across the extended telomere-adjacent region to determine the density and distribution of DNA sequence polymorphisms. Seventeen candidate SNPs were identified between the -1890 SNP and the DXYS14b repeat array some of which are present in dbSNP and one novel SNP was identified internal to DXYS14b at the -7152 position (Figure 1a).

Linkage disequilibrium across the region

The polymorphic sites in the Xp/Yp telomere-adjacent sequence, from the -1890 SNP (previously reported as -1888) to the -13 SNP next to the telomere, are in almost complete LD suggesting there has been very little recombination in this region during recent evolution [8]. To determine whether strong LD extends into the region encompassing the newly identified SNPs that span the DXYS14a and DXYS14b repeat arrays, allele specific oligonucleotide (ASO) hybridization assays were developed for the new SNPs. Some of the DNA samples generated more than one product with the 29c1M and 29c1X primers or a product of unexpected, larger size. The products of expected size contained the sequences described here but sequence analysis of the larger products did not resolve the origin of these fragments. To avoid any incorrect genotyping, the four candidate SNPs in this region (-3640, -3644, -4781 and -4838) were excluded from subsequent analyses. Genotype analysis was undertaken using ASO hybridization [23] for a total of 26 SNPs including 11 new SNPs (between -7152 to -2124 (rs28406733) positions), the -1890 SNP and 14 of the 22 SNPs immediately adjacent to the telomere on a panel of 50 UK individuals of North European descent (Figure 1a and supplementary table 1).

The polymorphic sites all have high minor allele frequencies, the lowest being 0.33 at the -7152 SNP and all were in Hardy-Weinberg equilibrium. The genotyping data were used to determine the maximum likelihood frequency of the four possible haplotypes between pairs of SNPs and from this data, D′ (a measure of LD) and an odds ratio that measures the significance of an association between two SNPs was determined. Figure 1b shows that LD remains high throughout the region that encompasses both of the minisatellite repeat arrays, DXYS14a and b. Inclusion of the additional SNPs in haplotype analysis increased the number of haplotypes present in

the Caucasian population (haplotypes were deduced from genotypes, data not shown) nevertheless the majority of chromosomes (58%) can be assigned to just two haplotypes that differ at all the polymorphic sites analysed.

Assay for germline crossovers in the XpYp telomere adjacent region

The presence of strong LD in the Xp/Yp telomere adjacent region suggests that crossing-over in the germline is reduced in this region. However, recent data have shown that hidden crossover-hotspots can occur in regions that show comparatively strong LD [24]. This apparent discrepancy between LD and crossover analysis has been interpreted to mean that the recombination hotspot is relatively young and so LD between SNPs has not been fully eroded. To explore the crossover activity of the Xp/Yp telomere adjacent region experimentally, a PCR based crossover assay was undertaken [1].

The crossover analysis was conducted using two rounds of PCR with nested repulsion-phase primers on sperm donor 8 using allele specific primers at the -3119 and -2780 (rs28578105) SNPs with return primers at the -13 and -30 SNPs (Figure 2). The DXYS14a repeat array was not included because of difficulty in designing primers that generated full-length allele specific products across the array. Fifty crossover PCR reactions were conducted, each containing 25,000 amplifiable molecules of sperm DNA from donor 8. Crossover-assays using the equivalent amount of blood DNA from the donor were conducted as negative controls, as genuine crossover molecules are not present in blood DNA. In addition positive control experiments were established to ensure that, if present, a single recombinant molecule could be detected in the crossover PCRs (Figure 2 and materials and methods).

One candidate recombinant molecule was detected in the sperm crossover reactions and none in reactions containing the blood DNA (Figure 2). The breakpoint in the candidate recombinant molecule was mapped by ASO-hybridisation to the interval between the nested priming sites at -2780 (rs28578105) and at -2576 (rs28491545) (Figure 1a and Figure 2). While the candidate recombinant molecule may have arisen

from a genuine germline crossover event it cannot be excluded that it arose as a misprimed product generated by the repulsion phase primers, although no such misprimed products were detected among the 1.25 million molecules analysed from blood DNA. If the candidate recombinant molecule is genuine, the recombination rate in the 2760bp interval is 0.029cM Mb^{-1} (95% confidence interval 0.002 – 0.17), 34 times lower than the genome average rate. Therefore the experimentally determined recombination rate is consistent with the LD analysis and indicates that germline crossovers are rare in the sequence immediately adjacent to the Xp/Yp telomere.

Analysis of the DXYS14 repeat arrays

Previous analysis has shown that the DXYS14 minisatellite repeat array is highly variable and mutable in the germline but the duplicated nature of the repeat array was not fully understood [19]. To investigate the contribution that each DXYS14 repeat array makes to length variation in this region and to explore the germline mutation rate we developed PCR assays to amplify the DXYS14a and b repeat arrays (Figure 3a). Amplification across both arrays using primers 29c1A and 29c1T generated one or two large products, as expected, in individuals heterozygous for different length alleles at the DXYS14a and b arrays (data not shown). This showed that the DXYS14 a and b arrays analysed lie close to one another at the terminus of Xp/Yp. However, in a subset of individuals amplification with the 29c1A and T primers generated short products that confirmed a subset of chromosomes have only one copy of the repeat array, termed DXYS14c (Figure 3b). The frequency of the chromosomes carrying only one copy of the array (DXYS14c) was 14% (36/260) among individuals with northern or western European ancestry.

Analysis of parental DNA from the CEPH family panel showed that both the DXYS14a and DXYS14b arrays are hypervariable with heterozygosities of 96% and 94% respectively (50/52 at DXYS14a and 51/54 at DXYS14b, supplementary table 2). The estimated mutation rate determined from the heterozygosities for the DXYS14a and b repeat arrays is 0.2% per gamete (using the symmetric K allele model, [25]). Segregation analysis within the 40 CEPH families did not reveal any recombination between the DXYS14a and b arrays but four paternally derived mutations were identified. In CEPH family 45 the paternally inherited allele at

DXYS14a was mutated in child 03, showing a deletion of 400bp (~13 repeats). In CEPH family 104 the mutant allele previously identified in child 09 was confirmed [19] and we showed that the deletion of 250bp (~8 repeats) occurred at the paternal DXYS14a array. In CEPH family 884 the father is heterozygous for a chromosome carrying DXYS14a and b repeat arrays and a chromosome carrying a single DXYS14c array. Two paternally derived mutations were identified in this family, one a deletion of 175bp (~ 6 repeats) at the DXYS14b array in child 06 and the second an insertion of 220bp (~7 repeats) in the DXYS14c array in child 11 (Figure 3c). Based on the small number of mutants detected in these families, the mutation rates for the DXYS14 a and b repeat arrays are 0.3% (2/639) and 0.2% (1/644) per gamete respectively, which are in good agreement with the estimated mutation rates. The observed mutation rate for DXYS14c, 1.2% per gamete (1/83), is not significantly different from that at DXYS14 a or b (Fisher Exact Test p=0.3 and 0.2 respetively).

Copy Number changes of the DXYS14 repeat array.

As shown above, 14% of Caucasian chromosomes carry a single DXYS14 repeat array (DXYS14c). PCR analysis showed that these chromosomes lack all the DNA sequence between DXYS14a and b (data not shown) and therefore they have only one copy of the duplication unit. Chromosomes carrying DXYS14c retain the sequences located internally at least as far as the 29c1Q priming site (at -7335) and including the -7152 SNP site (Figure 1 and Supplementary table 1). The distally located sequences including all SNPs from -3119 to the telomere are also retained. DXYS14c repeat arrays have been found associated with chromosomes carrying either of the highly diverged haplotypes (A and B) across the telomere–adjacent SNPs (Supplementary table 1). Furthermore when the presence/ absence of the DXYS14c repeat array is treated as a bi-allelic marker it is not in linkage disequilibrium with surrounding SNPs (data not shown).

A further two parents from the CEPH family panel (102.02 and 1418.01) each contained a chromosome that has a copy of the DXYS14b repeat array but appear null for the DXYS14a repeat array as no amplicons were generated with the 29c1W and 29c1T primers (supplementary table 2). The absence of DXYS14a amplicons in these individuals could occur because (1) the DXYS14a alleles are very short and not

detected, (2) one of the PCR priming sites contains a novel SNP or (3) these chromosomes lack the DXYS14a repeat array.

Two CEPH families segregate chromosomes that appear to have three DXYS14 repeat arrays. PCR analysis using the primers that amplify the DXYS14a, b or c arrays showed that the father in family 1421 (1421.01) inherited a chromosome from his mother (1421.10) that contains a short DXYS14a array (575bp) and apparently two DXYS14b arrays (1525bp and 500bp). This chromosome was inherited by several of his children (1421.05, 06 and 08, data not shown). Similar PCR analysis in CEPH family 1423 revealed that the mother (1423.02) inherited a chromosome from her mother that appears to contain two DXYS14b (1325bp and 700bp) and one DXYS14a (1575bp) array. She passed this chromosome onto four of her children (1423.04, 07, 10, and 15; Figure 4a). Amplification with the primers 29c1B and 29c1L in family 1423 and subsequent sequence analysis, showed that individuals with the additional DXYS14b array also contain a partial duplication of the AluJo element in this region (Figure 4b). This was not seen in family 1421.

Two possible explanations can be offered for the presence of two DXYS14b arrays in the 1421 and 1423 families. Either one of the DXYS14 priming sites (29c1A or 29c1V) has been duplicated within or close to the DXYS14b repeat array and so two different sized products are generated from the DXYS14b repeat array or the entire DXYS14b repeat array with both priming sites has been duplicated and the chromosomes have two DXYS14b repeat arrays (not necessarily close together) and one DXYS14a repeat arrays. Moreover it seems likely that the partial or complete duplication events occurred independently as the chromosome in family 1421 is associated with haplotype B across the telomere-adjacent SNPs whereas in family 1423 it is associated with haplotype C and it also has a partial duplication of an Alu element.

Discussion

We have defined a segmental duplication in the sequence adjacent to the Xp/Yp telomere, each copy containing a hypervariable minisatellite repeat array (DXYS14a

or DXYS14b). The 4% sequence divergence between the two copies of the duplication (excluding the DXYS14 repeat arrays) indicates the duplication is ancient. SNPs throughout the region are in strong LD suggesting there is reduced recombination in the germline and this was confirmed by direct crossover analysis across part of the region (-13 to -3119). The DXYS14 repeat arrays are embedded within this region of strong LD but they have high germline mutation rates. Thus our data indicate the main mutation mechanism underlying hypervariability at the DXYS14 a and b repeat arrays is via intra-molecular processes or unequal conversion events that are confined to the DXYS14 repeat arrays and do not result in the exchange of flanking SNPs. Comparison of genetic and physical maps of human chromosomes indicates that the ends of chromosomes support a high level of homologous recombination during meiosis. This is particularly evident in the Xp/Yp pseudosautosoaml regions (PAR1) during male meiosis [17] [4] though our data indicate the region of elevated recombination does not extend to the telomere itself. This is consistent with observed distribution of synapsis initiation sites that occur in subtelomeric regions but not telomeres [26]. Moreover the strong LD across the ~7kb analysed in the Xp/Yp telomere-adjacent region is different from the patterns of LD seen around the SHOX gene [7] located internally in PAR1 and it may indicate that meiotic recombination is suppressed adjacent to the telomere. It is known that human telomeres can suppress gene expression (known as telomere position effect, TPE) and that TPE may be dependent on telomere length and chromatin organization [27]. It has also been noted that the Xp/Yp telomere adjacent sequence in particular, may be modified in some way rendering it resistant to some restriction endonucleases [28]. To determine the extent of LD and possible recombination suppression from the Xp/Yp telomere we have used LD analysis of SNPs in the region with SNPs around the PGPL gene that lies ~153kb internally (T. Slingsby and A.J. Jeffreys unpublished data). There is free association between the two locations indicating that any telomere-mediated crossover suppression does not extend this far internally.

Sex chromosomes that have only a single copy of the duplication unit adjacent to the Xp/Yp telomere and therefore only one DXYS14 repeat array (DXYS14c) are quite common among northern and western Europeans (14%). These chromosomes might represent the ancestral state, prior to formation of the duplication unit. However, it is notable that the chromosomes carrying a single copy of the duplication are associated

with either of the highly diverged haplotypes (A or B) across the SNPs immediately adjacent to the XpYp telomere [8]. This raises the possibility that mutation events leading to the loss of one duplicon have occurred more recently and more than once. We also present evidence that some chromosomes carry the DXYS14b repeat array but may lack the DXYS14a repeat array. In addition another two chromosomes (2/144) may carry three DXYS14 repeat arrays but it appears that the sequence organization on these two chromosomes arose independently as they are associated with different telomere-adjacent haplotypes and only one is contains a partial duplication of an Alu element.

In summary, the data presented here indicate that two different processes operate on the sequences adjacent to the Xp/Yp telomere repeat array. Firstly mutations that only affect the DXYS14 repeat arrays occur at a high rate by intra-molecular processes or conversions limited to the repeat arrays. Secondly mutation events that alter the copy number of the duplication unit or the DXYS14 repeat arrays also occur but rarely. The processes underlying these rare events are not known but intra-molecular recombination between the DXYS14a and b arrays would give rise to a chromosome carrying a single copy of the duplication unit and release a small circular copy. Similarly, unequal exchange of sister-chromatids or homologous chromosomes could result in chromosomes carrying a single or multiple copies of the duplicon.

Materials and Methods

DNA samples.

Human DNAs derived from lymphoblastoid cell lines comprising the Centre d'Etude du Polymorphism Humain (CEPH, Paris, France) panel of DNAs from families with Northern or Western European ancestry were used for analysis of the DXYS14 repeat arrays. The panel of DNAs from 50 anonymised semen and blood donors were a gift from A.J. Jeffreys (University of Leicester, UK) and had been collected with informed consent and approval from the Leicestershire Health Authority Research Ethics Committee [3]. These samples were used for SNP typing, LD analysis, analysis of the DXYS14 repeat arrays and for the recombination assay.

PCR

DNA was amplified using the polymerase chain reaction (PCR). Unless otherwise stated, reactions contained 50ng human genomic DNA, 1 μM of each primer, 0.05 U μl^{-1} *Taq* polymerase (ABgene) and PCR buffer (45mM Tris-HCl pH 8.8, 11mM $(NH_4)_2SO_4$, 4.5mM $MgCl_2$, 6.7mM β-mercaptoethanol, 4.4mM EDTA, 1mM of each dNTP and 113μg/ml BSA [29]) and were cycled at 96°C 30 seconds, appropriate annealing temperature 40 seconds and 70°C 3 minutes. Products longer than 3kb were amplified as above, with the addition of Tris-HCl (pH 8.8) to final concentration of 33mM; 0.1U μl^{-1} of *Taq* polymerase and 0.05U μl^{-1} of *Pfu* polymerase (Stratagene) [30]. The denaturation temperature was 94°C for 15 seconds and extension times were increased. All PCR reactions were carried out using a PTC200 DNA Engine™ (MJ Research, USA).

PCR Primers and optimal annealing temperatures

BX64A cagcaggatcggagagggac 64°C; BX64E aacgcggtctgcggtccag 64°C; Tsk8M gaccaggttttccagtgtgt 64°C; Tsk8R ctttccccagctatggcttc 65°C; Tsk8W aacctgccgcatctcacctg 65°C; TS-12rev accctctgaaagtggacc 65°C; 29c1A gtctggtgtggtctggaaaatg 64°C; 29c1B tctctttttggtccggttcccatc 66°C; 29c1E tgtgtgacccataggaga 66°C; 29c1L aaggaaggagaccaccactcc 64°C; 29c1M gaggagtggtggtctcctt 63°C; 29c1Y aagcacaagccatcgcgcccag 64°C; 29c1Q gcatctcaccatttcgctttcc 64°C; 29c1R aagagccgagcggagctgaaggg 63°C; 29c1T acctcctgtcaatatcccttccc 64°C 29c1X agactgccgctcctcctg 65°C; 29c1Arev cattttccagaccacaccagac 64°C; 29c1V ttgggaccaggctgccactcc 64°C; 29c1W aagcctcagggggtcccgac 64°C;

Sequencing and SNP detection

Gel purified PCR products were sequenced using the Big Dye Terminators (ABI) on an ABI 377 automated sequencer. Electropherograms were analysed and contiguous sequences created with the AutoAssembler software (ABI).

Four PCR amplicons defined by primers 29c1B+29c1L; 29c1M+29c1X; 29c1B+Tsk8R and Tsk8M+TS-12rev that cover the region between the DXYS14b

repeat array and the telomere (excluding the DXYS14a repeat array) were sequenced from 3 to10 selected individuals with northern or western European ancestry and candidate SNPs identified. To verify the candidate SNPs that reside in the interval between the DXYS14a and b repeat arrays, genomic DNA from five CEPH parents was amplified with 29c1A and 29c1T to generate long products encompassing both repeat arrays. These products were reamplified with 29c1B and 29c1L to generate a ~1.7kb amplicon that was sequenced with 29c1I and 29c1S. To identify SNPs centromeric to the DXYS14b repeat array an amplicon, defined by the primers BX64A (specific to duplication unit B) and 29c1Q, was sequenced from six individuals and an additional SNP was identified at the -7152 position.

SNP typing by allele specific oligonucleotide (ASO) hybridization

PCR products (29c1B+29c1L; 29c1M+29c1X; 29c1B+Tsk8R; Tsk8M+TS-12rev) were generated from the Xp/Yp telomere-adjacent DNA of the blood / semen donor panel and 3 to 100ng per 1 kb of each amplicon was spotted onto Hybond™-N^{fp} membranes (Amersham) to produce 96 well format dot-blots. ^{32}P end-labelled ASOs were hybridized sequentially to the dotblots in TMAC (tetramethylammonium chloride, Sigma Biochemical Company, UK) hybridization solution at 53°C [23]. Following post hybridization washes the dotblots were exposed to Fuji RX100 X-ray.

Allele-specific oligonucleotides for genotyping

13XYASO/T tgctgattggtccacttt, 13XYASO/A tgctgataggtccacttt; 30XYASO/T tgctccctcccacatcct, 30XYASO/A tgctcccacccacatcct; 146XYASO/G gtgagcagtagcaagatt, 146XYASO/A gtgagcaatagcaagatt; 176XYASO/T cagctcataaaggcagtg, 176XYASO/G cagctcagaaaggcagtg; 298XYASO/A aagtggcatgtccggagt, 298XYASO/G aagtggcgcgtccggagt; 415XYASO/C ggtgctcctctaggggtt, 415XYASO/T ggtgctcttctaggggtt (rs28494123); 427XYASO/A ggttatcaaccaggtgct, 427XYASO/G ggttatcgaccaggtgct (rs28718810); 540XYASO/C gctsatgcaacgggctgt, 540XYASO/T gctsatgtaacgggctgt (rs35793405); 544XYASO/C aggggctcatgyaacggg, 544XYASO/G aggggctgatgyaacggg (rs35832313); 554XYASO/A gttagagaccaggggcts, 554XYASO/T gttagagtccaggggcts (rs28675701); 652XYASO/A aggcccaagtccctaaat, 652XYASO/G aggcccaggtccctaaat (rs28419004); 842XYASO/C tcccgagctgagaaagac,

842XYASO/A tcccgagatgagaaagac (rs28705946); 1078XYASO/C tggtaaaccaaccacaga, 1078XYASO/A tggtaaaacaaccacaga (rs28516717); 1890XYASO/A cagtgaaagagccaagat, 1890XYASO/G cagtgaaggagccaagat; 2124XYASO/C agtaaatcaggcaaaatg, 2124XYASO/t agtaaattaggcaaaatg (rs28406733); 2180XYASO/C ttttatacttttgtttag, 2180XYASO/T ttttatatttttgtttag; 2283XYASO/A ggaagggaagggcagcag, 2283XYASO/G ggaaggggagggcagcag; 2372XYASO/G tcgtctagctatttatct, 2372XYASO/C tcgtctacctatttatct; 2576XYASO/T ggttcacttgaaagggtc, 2576XYASO/G ggttcacgtgaaagggtc (rs28491545); 2780XYASO/A caactgtaggggttaggg, 2780XYASO/C caactgtcggggttaggg (rs28578105); 3119XYASO/G tgggtgcgatggcacact, 3119XYASO/A tgggtgcaatggcacact; 5095XYASO/G tgccaaggtcatctagct, 5095XYASO/A tgccaagatcatctagct; 5263XYASO/G aggaagcgtggggaacgt, 5263XYASO/A aggaagcatggggaacgt; 5328XYASO/G aattcctggccctttaa, 5328XYASO/A aattcctagccctttaa; 5712XYASO/G gttttgcggttgggggtt, 5712XYASO/A gttttgcagttgggggtt; 7152XYASO/C agtgacggagggtgacgg, 7152XYASO/T agtgacgaagggtgacgg. The ASO assays for the -5886 and -5880 SNPs sometimes gave ambiguous results (possibly because the two SNPs are close together) and therefore they were not included in the analysis.

Linkage disequilibrium analysis

Genotype data from the panel of 50 semen donors were used to calculate LD between pairs of SNPs using the D′ measure and software developed by A.J Jeffreys [1]. This software also estimates the likelihood ratio (LR) of association versus linkage equilibrium. All the loci included in the analysis had a minor allele frequency greater than 0.2 and they were in Hardy-Weinberg equilibrium.

Southern blot analysis of PCR amplified DXYS14 repeat arrays.

The ^{32}P labelled 29c1 repetitive probe was hybridized to Southern blots of amplicons from the DXYS14 repeat arrays in phosphate-SDS solution (7% SDS, 0.5M Na_2HPO_4 pH7.2, 1mM EDTA) [31] at 65 °C for 5 hours to overnight. Following hybridization the blots were washed in 0.5 x SSC, 0.1% SDS and exposed to Fuji RX100 X-ray film.

Crossover assay

The crossover assay was conducted from the -13A/T to -3119A/G SNP positions in the Xp/Yp telomere-adjacent sequence, and then at the nested internal -30 A/T and -2780A/C (rs28578105) SNPs sites using a method reported previously [23] [1]. Primers used TS-13Arev accctctgaaagtggacca; TS-13Trev accctctgaaagtggacct; at the -3119SNP 29c1GaspA ttaagaaattcgctgggtgca or 29c1GaspG ttaagaaattcgctgggtgcg; TS-30Arev aatcagcaggatgtggga; TS-30Trev tatcagcaggatgtgggt; and at the -2780 SNP (rs28578105) 29c1NaspA ggggggacctcaactgta or 29c1NaspC ggggggacctcaactgtc.

Semen donor 8, selected for the crossover analysis, was heterozygous at every SNP site in the 3.1 kb target region (confirmed by allele specific PCR and sequence analysis). Donor 31 was selected for control experiments. Genomic DNA was extracted from blood and sperm DNA from each donor, digested with *Kpn*I and concentration determined. Experiments were conducted to determine the maximum amount of genomic DNA that could be included into each crossover PCR reaction without compromising amplification of a candidate recombinant molecule. For these experiments genomic DNA equivalent to 0.8 target molecules (determined by Poisson analysis) from donor 31 (Figure 2) was added to increasing amounts of genomic DNA (equivalent to 10,000 to 50,000 target molecules) from another donor that could not be amplified with the 29c1GaspA and TS-13Trev primers. This showed that single molecules from donor 31 could be amplified in a background of 25,000 molecules per reaction.

The crossover assay on donor 8 was conducted with the repulsion phase allele specific primers 29c1GaspA and TS-13Trev in 50 PCRs each containing sperm genomic DNA equivalent to 25,000 target molecules (number of molecules verified by Poisson analysis). Following a second round of repulsion phase PCR with nested primers 29c1NaspA and TS-30Trev, the PCR products were detected by Southern blot hybridization to the radioactively labelled Tsk8M-Tsk8G amplicon. Control crossover assays were conducted on the equivalent amount of blood DNA from donor 8 and no products were detected. The single candidate recombinant molecule was reamplified with 29c1NaspA and TS-51XY (aagcagactgcctgagccagc 64°C) primers and then genotyped at the intervening SNPs using the ASO hybridization to dotblots as described above.

Acknowledgements

We would like to thank Zankhna Mistry in particular for her contribution to the work presented here. We thank Rita Neumann for advice on the crossover assay; Pricilla Lo for technical assistance; Tim Slingsby and Alec Jeffreys for sharing their SNP data around the PGPL gene. We are also most grateful to Alec Jeffreys, Celia May and Mark Jobling for their advice and comments on the manuscript. This work was funded by grants to NJR from the UK MRC and M. Hills was supported by an MRC Ph.D. studentship.

References

[1] A.J. Jeffreys, L. Kauppi, R. Neumann, Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex, Nat Genet 29 (2001) 217-222.

[2] A.J. Jeffreys, R. Neumann, Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot, Nat Genet 31 (2002) 267-271.

[3] A.J. Jeffreys, C.A. May, Intense and highly localized gene conversion activity in human meiotic crossover hot spots, Nat Genet 36 (2004) 151-156.

[4] F. Rouyer *et al.*, A gradient of sex linkage in the pseudoautosomal region of the human sex-chromosomes, Nature 319 (1986) 291-295.

[5] G.A. Rappold, The pseudoautosomal regions of the human sex-chromosomes, Hum. Genet. 92 (1993) 315-324.

[6] R.J. Blaschke, G. Rappold, The pseudoautosomal regions, SHOX and disease, Curr Opin Genet Dev 16 (2006) 233-239.

[7] C.A. May, A.C. Shone, L. Kalaydjieva, A. Sajantila, A.J. Jeffreys, Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX, Nat Genet 31 (2002) 272-275.

[8] D.M. Baird, A.J. Jeffreys, N.J. Royle, Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere, EMBO Journal 14 (1995) 5433-5443.

[9] D.M. Baird, J. Coleman, Z.H. Rosser, N.J. Royle, High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution, Am J Hum Genet 66 (2000) 235-250.

[10] J. Coleman, D.M. Baird, N.J. Royle, The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q, Hum Mol Genet 8 (1999) 1637-1646.

[11] H.A. Pickett *et al.*, Telomere instability detected in sporadic colon cancers, some showing mutations in a mismatch repair gene, Oncogene 23 (2004) 3434-3443.

[12] H. Riethman *et al.*, Mapping and initial analysis of human subtelomeric sequence assemblies, Genome Res 14 (2004) 18-28.

[13] E.V. Linardopoulou *et al.*, Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication, Nature 437 (2005) 94-100.

[14] J. Flint *et al.*, Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains, Hum Mol Genet 6 (1997) 1305-1313.

[15] M. van Geel *et al.*, Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin, Genomics 79 (2002) 210-217.

[16] H. Riethman, A. Ambrosini, S. Paul, Human subtelomere structure and variation, Chromosome Res 13 (2005) 505-515.

[17] H.J. Cooke, W.R.A. Brown, G.A. Rappold, Hypervariable telomeric sequences from the human sex-chromosomes are pseudoautosomal, Nature 317 (1985) 687-692.

[18] J. Levilliers, B. Quack, J. Weissenbach, C. Petit, Exchange of terminal portions of X- and Y-chromosomal short arms in human XY females, Proc Natl Acad Sci U S A 86 (1989) 2296-2300.

[19] C.F. Inglehearn, H.J. Cooke, A VNTR immediately adjacent to the human pseudoautosomal telomere, Nucleic Acids Res. 18 (1990) 471-476.

[20] D.M. Baird, N.J. Royle, Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence, Human Molecular Genetics 6 (1997) 2291-2299.

[21] N.J. Royle, M.C. Hill, A.J. Jeffreys, Isolation of telomere junction fragments by anchored polymerase chain-reaction, Proc. R. Soc. Lond. B 247 (1992) 57-61.

[22] W.R.A. Brown *et al.*, Structure and polymorphism of human telomere-associated DNA, Cell 63 (1990) 119-132.

[23] A.J. Jeffreys, A. Ritchie, R. Neumann, High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot, Hum Mol Genet 9 (2000) 725-733.

[24] A.J. Jeffreys, R. Neumann, M. Panayi, S. Myers, P. Donnelly, Human recombination hot spots hidden in regions of strong marker association, Nat Genet 37 (2005) 601-606.

[25] A.J. Jeffreys, N.J. Royle, V. Wilson, Z. Wong, Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA, Nature 332 (1988) 278-281.

[26] P.W. Brown *et al.*, Meiotic synapsis proceeds from a limited number of subtelomeric sites in the human male, Am J Hum Genet 77 (2005) 556-566.

[27] J.A. Baur, Y. Zou, J.W. Shay, W.E. Wright, Telomere position effect in human cells, Science 292 (2001) 2075-2077.

[28] S. Steinert, J.W. Shay, W.E. Wright, Modification of subtelomeric DNA, Mol Cell Biol 24 (2004) 4571-4580.

[29] A.J. Jeffreys, V. Wilson, R. Neumann, J. Keyte, Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells, Nucleic Acids Res 16 (1988) 10953-10971.

[30] W.M. Barnes, PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates, Proc Natl Acad Sci U S A 91 (1994) 2216-2220.

[31] G.M. Church, W. Gilbert, Genomic Sequencing, Proc. Natl. Acad. Sci. USA 81 (1984) 1991-1995.

Figure Legends.

Figure 1.

a. Sequence organisation of the Xp/Yp- telomere adjacent region. The terminal 7811bp of the Xp/Yp telomere adjacent region (from chrXY:39 - 9828) is shown with the telomere on the right (orange/red arrowheads). The two copies of the duplicated region (A and B) are aligned and the two DXYS14 repeat arrays (a and b) are shown (blue boxes). Other tandem repeat arrays: monomorphic minisatellite (blue box), STRs (orange boxes) are shown. Dispersed repeats are shown, Alu (yellow boxes), other elements (dark blue boxes). SNPs are numbered from the telomere and those included in the LD analysis are shown as black arrow heads, others pink arrowheads. The DXYS14 repeat arrays are variable in length and so they have not been included in the numbering. Nucleotide positions from the HGS March2006 are shown in grey (e.g. chrXY: 39). Primers shown are labelled blue arrows (e.g. Tsk8R).

b. The 26 SNPs in the 7kb sequence immediately adjacent to the Xp/Yp telomere show strong linkage disequilibrium (LD). LD was measured in a North European population of 50 semen donors, using the D′ measure. Complete LD (D′=1) is given by pairs of markers that show only three of the four possible haplotypes. D′ values were plotted (lower right triangle) for all pair-wise combinations of SNPs with minor allele frequency >0.2 and colour coded as indicated. Regions of strong LD are shown in red and the position of each SNP is shown below. The upper left triangle shows the likelihood ratio of each pair of SNPs being in linkage equilibrium (free association).

Figure 2

Assay for crossovers in the Xp/Yp telomere adjacent region. Donor 8 is heterozygous at all SNPs in the region and the haplotypes of his chromosomes are shown. Crossover analysis on sperm DNA from donor 8 was conducted using two pairs of nested repulsion-phase primers at the -3119 and -13 then -2780 (rs28578105) and -30 SNPs. Results of the crossover assay are shown for 25 aliquots of sperm or blood DNA that each contained 25,000 molecules of the Xp/Yp region. Only one aliquot of the sperm DNA (08) generated a putative recombinant molecule and dotblot analysis of SNPs in the region, showed the breakpoint of the putative crossover molecule lies between the -2780 and -2576 (rs28491545) SNPs. Donor 31 contains one chromosome that can be amplified by the primer pairs used in the crossover assay on donor 8. As shown, genomic DNA from donor 31 was used in separate positive control experiments for the crossover assay (see materials and methods).

Figure 3.

Analysis of the DXYS14 repeat arrays.

a. Specific PCR amplification of the DXYS14 a and b repeat arrays was achieved by using sequence differences between the two duplication units such that DXYS14b is

amplified with primers 29c1A + 29c1V and DXYS14a with primers 29c1W + 29c1T. Amplification of DXYS14c is achieved with primers 29c1A + 29c1T.

b. Amplification of the DXYS14 repeat arrays in two donors. Donor 4 shows length heterozygosity at the DXYS14a and b arrays and does not contain a chromosome carrying DXYS14c. Donor 11 is heterozygous for a chromosome with one copy of DXYS14a and b and a chromosome carrying a single copy of DXYS14c.

c. Mutation analysis of DXYS14 repeat arrays in a family. Amplification of the DXYS14 repeat arrays in CEPH family 884 followed by Southern blot analysis showed normal Mendelian segregation of the DXYS14 repeat arrays in the children. However, child 884.06 inherited a length mutant DXYS14b array from the father (884.01), arising from a deletion event. Similarly child 884.11 inherited a mutant DXYS14c array following an insertion event in the paternal germline.

Figure 4.

Segregation of a chromosome that apparently carries three copies of the DXYS14 repeat array.
a. PCR amplification and Southern blot analysis of DXYS14 repeat arrays in family 1423, revealed that the mother (1423.02) contains an unusual chromosome that appears to carry two DXYS14b and one DXYS14a repeat array. The co-segregating DXYS14b bands in the mother are joined by a vertical line. She inherited this chromosome from her mother (1423.14) and passed it onto four of her children (1423.04, 07, 10, 15). The mother's other chromosome (inherited from her father 1423.13) carries one DXYS14b and one DXYS14a repeat array and it was inherited by the remaining four children (1423.03, 06, 08, 09).

b. PCR amplification with the 29c1B and 29c1L primers generated amplicons of the expected length in all members of family 1423 but in addition a larger amplicon was detected in all the individuals that contain the chromosome with three DXYS14 repeat arrays (1423.02, 04, 07, 10, 15, 14). Sequence analysis revealed that the larger fragment contained a partial duplication of the Alu element in this region.