ARTICLE

High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot

Alec J. Jeffreys+, Alistair Ritchie and Rita Neumann

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

Received 20 December 1999; Revised and Accepted 22 January 2000

DDBJ/EMBL/GenBank accession nos AJ251484–AJ251485

Little is known about the nature of recombination hotspots in the human genome and the relationship between crossover activity and patterns of linkage disequilibrium. We have therefore used both haplotype analysis and direct detection of crossovers in sperm to characterize a putative recombination hotspot in the TAP2 gene within the class II region of the MHC. Haplotype diversity provided evidence for a localized hotspot within intron 2 of this gene. Sperm DNA typing using allele-specific PCR primers to selectively amplify recombinant TAP2 molecules revealed a highly localized meiotic crossover hotspot ~1.2 kb long, unusually abundant in sequence polymorphisms and flanked by DNA much less active in recombination. Sperm crossover appeared to be fully reciprocal, and almost all crossover products were simple, involving a single exchange between adjacent heterozygous markers. This hotspot appears to be much more active in female than male meiosis. No primary sequence similarities could be found between any of the very few well defined crossover hotspots in the human genome, all of which show recombinationally active domains 1–2 kb long. Direct comparison of recombination frequency and haplotype diversity in TAP2 showed that linkage disequilibrium measures were a poor predictor of crossover frequency in this region, with non-recombining markers sometimes in free association and with examples of pairs of markers spanning the recombination hotspot showing substantial or even absolute linkage disequilibrium.

INTRODUCTION

There is abundant evidence that meiotic crossovers are non-randomly distributed in the human genome. Comparisons of linkage maps with physical maps or sequences of human chromosomes have revealed evidence for long domains showing elevated recombinational activity (1–6), though nothing is known about regional crossover activity within these hot domains. Higher resolution analysis of patterns of linkage disequilibrium (LD) and the physical mapping of occasional crossover breakpoints detected in families have led to the provisional identification of a number of putative crossover hotspots in the human genome, including examples in the β-globin gene cluster (7,8), the major histocompatibility complex (MHC) (9,10) and PGM1 (11). However, analysis of localized breakdown of LD is a potentially unreliable tool for defining crossover hotspots (12). Also, the full localization of a recombination hotspot requires the mapping of large numbers of crossovers over a small and defined interval of DNA, an approach which is not feasible in human pedigrees.

To date, only two human hotspots have been accurately localized and characterized. The first lies within duplicated 24 kb elements 1.5 Mb apart on chromosome 17, which undergo unequal exchange resulting in duplication or deletion of the 1.5 Mb region and causing Charcot–Marie–Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsy (HNPP), respectively (13,14). Analysis of numerous unequal crossovers in patients has revealed a localized hotspot ~1 kb long in the 24 kb elements, providing an important model for analysing processes of unequal but not equal crossover (15). The second well-defined meiotic recombination hotspot lies adjacent to the unstable human minisatellite MS32 (D1S8) (16). It is again short (1.5 kb long) and appears to be responsible for driving recombination not only in DNA flanking MS32 but also within the minisatellite repeat array, resulting in germline-specific minisatellite instability arising through inter-allelic conversion plus unequal and equal crossovers within the repeat array (17). The MS32 hotspot was discovered by the direct analysis of large numbers of recombinant molecules recovered directly from sperm DNA using allele-specific PCR directed to heterozygous single nucleotide polymorphism (SNP) sites near the repeat array (16). To test whether this approach could be used more generally to explore meiotic crossover activity in non-repeat regions of the human genome, we have analysed a putative recombination hotspot in the MHC.

*To whom correspondence should be addressed. Tel: +44 116 252 3435; Fax: +44 116 252 3378; Email: ajj@le.ac.uk
Haplotyping of the MHC class II region has revealed domains of LD in free association across a 15 kb interval between TAP1 and TAP2 (18,19). Cullen et al. (9) characterized 11 pedigree recombinants mapping to a 475 kb region between HLA DPB1 and DQB1 and spanning this interval. Two crossovers, both maternal, localized to the TAPI–TAP2 interval, indicating a significant clustering in this region (9). Furthermore, both co-localized to a 850 bp region within intron 2 of the TAP2 gene, providing provisional evidence for a localized hotspot. However, this clustering of two crossovers within the TAPI–TAP2 interval could have arisen by chance (P = 0.11), and it remains possible that there is no local hotspot but rather that the entire 15 kb interval shows modest recombinational enhancement. We have therefore used haplotype diversity analysis to refine the location of this putative hotspot, and have shown that high resolution characterization of hotspot activity is indeed possible by direct analysis of sperm DNA.

RESULTS

Polymorphisms in the LMP7–TAP2 gene region

The recombination breakpoint mapping analysis of Cullen et al. (9) suggested a meiotic crossover hotspot within the second intron of the TAP2 gene. To screen for additional polymorphisms to aid LD and crossover analysis, a 9.7 kb segment centred on the putative hotspot and extending from within the LMP7 gene to intron 6 of TAP2 was completely resequenced in eight Caucasians. Two insertion/deletion (indel) polymorphisms were identified, together with 45 SNPs including all four previously identified by Cullen et al. (9). The SNPs included 32 base transitional polymorphisms of which eight were at CpG doublets, plus 13 transversions (Fig. 1A). Only two SNPs were in coding sequence; T41G→A in TAP2 exon 5 resulted in the replacement V379I reported previously (20), whereas T42G→T in exon 6 was a silent substitution. There were two instances of pairs of SNPs separated by a single base; subsequent analysis showed absolute LD between these pairs and each pair was therefore treated as a single polymorphic site. The resulting 45 polymorphic sites (T1–T45) were spread throughout the LMP7–TAP2 region, with some clustering between markers T17 and T30 in a 1 kb region in TAP2 intron 2.

Haplotype diversity

TAP2 haplotypes were determined by using allele-specific PCR to separate haplotypes in 30 UK Caucasians followed by typing all 45 SNPs on each of the 60 separated chromosomes using allele-specific oligonucleotide (ASO) hybridization. This survey revealed 28 different haplotypes (Fig. 1B). Haplotype diversity was investigated by LD analysis of all pairwise combinations of SNPs. Two measures of LD were used. The first is the standardized LD coefficient Δ (21) which is a measure of absolute association such that Δ = 0 only when there are just two haplotypes per pair of markers (12). The second is Lewontin’s coefficient D* (12,22) which measures the degree of complete association such that D’ = 1 only if all four haplotypes exist for the marker pair; D* is therefore useful for signalling the existence of obligate recombinants. Both Δ and D* revealed three distinct domains in the TAP2 region (Fig. 2; data not shown). Domain 1, at least 4.2 kb long and extending to marker T15, showed strong and highly significant LD between almost all pairs of markers, defining just two major haplotypes (a–j and k–t) (Fig. 1B). Despite the strong LD, obligate recombinants do exist in this domain, as shown by the existence of pairs of markers showing all four haplotypes (e.g. markers T10/T13 and T1/T8). Domain 2 spanned the 1 kb long SNP cluster inside TAP2 intron 2 and was characterized by a substantial though erratic reduction in LD and the appearance of multiple haplotypes, even between closely linked markers (Fig. 2). This was seen most dramatically for markers T23 and T24 that are separated by only 16 bp yet are in substantially free association (Δ = 0.32, 95% CI 0.06–0.50; D* = 0.59) and show all four haplotypes, indicating active recombinational exchange. Domain 3 extending downstream from marker T32 showed strong LD, in particular when estimated by D’, and the complete absence of any obligate recombinants. Closer inspection revealed an underlying framework of just two major haplotypes defined by markers T33, T37 and T39 (Fig. 1B). All other variants showed minor alleles at low frequency (0.07–0.18), which in all cases were restricted to very similar or identical haplotypes in domain 3 [for example, variants T42T and T45T found in absolute association on the identical 3 haplotypes a, b, n and z (Fig. 1B)]. Furthermore, sequence analysis of the corresponding LMP7–TAP2 region in the chimpanzee and gorilla showed in every case that these low frequency alleles corresponded to the derived, not ancestral, state (Fig. 1B). Haplotype diversification in domain 3 can therefore be completely explained by relatively recent base substitutional changes in the two framework haplotypes creating SNPs with low allele frequencies that remain totally associated with their parent haplotype, with no need to invoke any recombinational exchange. Domain 1 and domain 3 haplotypes were in completely free association, with no pairs of domain 1 and domain 3 framework markers showing significant LD (Δ = 0–0.28, D* = 0.02–0.47 for all pairwise comparisons of framework markers). This is consistent with a recombination hotspot localized approximately to domain 2 and flanked by regions less active in meiotic recombination.

Detection of meiotic crossovers in sperm DNA

Haplotype analysis identified a putative recombination hotspot within TAP2 intron 2 but gave little information on crossover frequency and distribution in this region. Two male subjects were therefore chosen for detecting recombinant molecules directly in sperm DNA. Both were heterozygous at the domain 1 framework markers T4 and the closely linked markers T6 plus T7, as well as at the domain 3 markers T37 and T39. Multiple batches of sperm DNA each containing 10 000–24 000 amplifiable progenitor TAP2 molecules were screened for recombinants using two rounds of allele-specific PCR, first directed to sites T4 and T39 in repulsion phase, and second to the nested sites T6/T7 plus T37, again in repulsion, to selectively amplify molecules carrying crossovers in this region (Fig. 3i). This nested-PCR strategy provided sufficient allele specificity to allow the amplification and detection of putative recombinant sperm DNA molecules (Fig. 3i). Parallel analyses of crossover-free control blood DNA yielded much fainter signals derived almost certainly from inter-haplotype jumping PCR artefacts; these will arise during the later stages of the first round of PCR by single primer extension from non-recombinant molecules creating incomplete single-stranded
Figure 1. Single nucleotide polymorphisms and haplotype diversity in the TAP2 gene region. (A) Distribution of SNPs and indels identified by resequencing eight Caucasians. The locations of polymorphisms T1–T45 are indicated, where position 1 corresponds to base 112,741 in the MHC class II sequence of Beck et al. (42) (GenBank accession no. X87344). Exons are indicated by boxes. Sites T20 and T29 each contain two SNPs separated by one base pair and in absolute LD in Caucasians. Site T35 is a 3 bp indel with alleles TGTAAACAAAAATG (+) and TGTAAAAATG (−), and T39 is a 1 bp indel with alleles CTCAAAATC (−) and CTCACAATC (+); additional bases in the + alleles are indicated in bold. (B) Haplotype diversity determined from the analysis of 60 Caucasian chromosomes. Each polymorphism where the less common allele has a frequency of at least 0.2 is shown as a large black or white circle corresponding to the first and second allele indicated in (A) (e.g. T1G is black, T1T is white). Polymorphisms with the minor allele at a frequency of <0.2 are similarly shown as small grey or white circles. Allele shading was chosen to highlight regions of LD in domains 1 and 3, and is arbitrary in domain 2. The numbers observed of each of the different haplotypes a–zb are shown on the right. The ancestral haplotype (anc) was deduced from sequences of the chimpanzee and gorilla TAP2 region (data not shown; GenBank accession nos AJ251484 and AJ251485, respectively). The double site T20 has the sequence CAA in both great apes compared with alleles CAG and TAA in humans. Further details are available at: http://www.le.ac.uk/genetics/ajj.
products that anneal to corresponding products from the other haplotype. Putative crossovers were successfully detected in sperm for sets of primers in both of the recombinant linkage phases, allowing reciprocal crossovers to be studied (Fig. 3i and ii).

Analysis of $2.4 \times 10^6$ progenitor TAP2 molecules from the sperm DNA of these two men yielded 176 crossover molecules, giving a recombination frequency of 0.007 cM across the 5.2 kb interval between T7 and T37. In contrast, $1.6 \times 10^6$ blood molecules yielded no crossovers with a signal intensity as high as that seen in sperm, providing strong evidence that these sperm molecules represent authentic meiotic recombinants and not PCR artefacts. Very similar numbers of sperm crossovers were detected for each orientation of allele-specific primers (94 versus 82 for orientation A and B, respectively), consistent with reciprocal meiotic crossover.

**Mapping crossover breakpoints**

The two men analysed for sperm crossovers were additionally heterozygous at 18 or 19 different markers between sites T7 and T37, allowing crossover breakpoints to be localized by ASO typing of each sperm recombinant (Fig. 3iii). Almost all of the crossovers (175/176) were simple, mapping to a single interval between adjacent heterozygous markers. Only one crossover was more complex, showing recombination between markers T28 and T29 accompanied by exchange at the upstream site T23 (Fig. 3iv); it is possible that the switch at T23 is a PCR artefact arising by base misincorporation at the earliest stages of amplification, though such switches in the uncloned PCR products of single DNA molecules are rare (16). None of the recombinants matched any of the haplotypes detected in Caucasians (Fig. 1B), nor contained inappropriate alleles at any of the homozygous SNP sites in these men, providing strong additional evidence that these are authentic crossovers and not contaminant molecules from other individuals.

Sperm crossover breakpoints were highly non-randomly distributed in the TAP2 gene, with 81% localizing to the 1.4 kb region between markers T15 and T30 though not mapping to any single interval in this region (Fig. 4). Crossover frequencies were used to estimate the crossover efficiency per unit length of DNA in each interval and to compare this efficiency with the mean rate of crossover in the human genome at male meiosis of 0.89 cM/Mb (23). This revealed a crossover hotspot ∼1.2 kb long with a peak activity of 8–10 cM/Mb (Fig. 4). This hotspot coincided exactly with the recombinationally active domain 2 provisionally identified by haplotype analysis (Fig. 1B). Furthermore, the single female meiotic recombinant localized exactly by Cullen et al. (9) also mapped within this hotspot. Outside the hotspot, the recombinational proficiency fell substantially to ∼0.4 cM/Mb, below the genome average rate; this will be an over-estimate if

![Figure 2](image-url). Linkage disequilibrium in the TAP2 gene region estimated using the standardized LD coefficient $\Delta$ (21) or Lewontin’s coefficient $D'$ (12,22). Analysis of $\Delta$ was restricted to the 30 framework polymorphic markers showing minor allele frequencies of at least 0.2. LD coefficients were determined for all pairs of adjacent markers (squares), and for pairs separated by one (circles) or two (+) intervening sites, and are plotted at the midpoint of each interval tested. The locations of the SNPs tested are shown above the plot. LD values were determined from data in Figure 1 as follows. Linkage disequilibrium was estimated as $D = x_{ij} - p_i q_j$, where $x_{ij}$ is the frequency of haplotype $A_iB_j$, and $p_i$ and $q_j$ are the frequencies of alleles $A_i$ and $B_j$ at loci $A$ and $B$, respectively. $\Delta$ is given by $D(\bar{p}_1\bar{q}_1/\bar{p}_2\bar{q}_2)^{1/2}$, where $\bar{p}_1$ and $\bar{q}_1$ are the frequencies of the other alleles at $A$ and $B$, respectively. $D'$ is given by $DD'_{\text{max}}$, where $DD'_{\text{max}} = \min[p_1q_1, p_2q_2]$ when $D < 0$ or $DD'_{\text{max}} = \min[p_1q_2, p_2q_1]$ when $D > 0$. 

![Diagram](image-url). Mapping crossover breakpoints in the TAP2 gene. Domain 2 is the recombinationally active domain, and is highlighted in light blue. The recombination hotspot is indicated by the purple box. The recombination efficiency per unit length of DNA is shown in orange, with the peak activity of 8–10 cM/Mb in the hotspot. The crossover frequency is shown in green, with the highest frequency in the hotspot. The crossover efficiency is shown in blue, with the highest efficiency in the hotspot. The crossover frequency is shown in red, with the highest frequency in the hotspot.
some of these crossovers are in fact inter-haplotype PCR artefacts that appear to occur at random along the test interval, as shown by mapping exchange points in blood PCR artefacts (data not shown). Finally, the morphology of the hotspot was indistinguishable for the two classes of reciprocal crossover products, consistent with fully reciprocal meiotic recombination, and was very similar in extent and intensity in the two men tested.

**DISCUSSION**

**Properties of recombination hotspots**

Sperm crossover analysis has directly established the existence of a highly localized meiotic reciprocal crossover hotspot in intron 2 of the TAP2 gene within which one, and probably both, of the maternal crossovers localized by Cullen et al. (9) map. The TAP2 hotspot shows features similar to those of the minisatellite MS32 hotspot (16) and the CMT1A/HNPP (15) unequal crossover hotspot. All three are short (1–2 kb) but do not show crossovers restricted to a single interval, and for TAP2 and MS32 at least are flanked by regions showing a recombinational efficiency significantly reduced compared with the genome average rate of crossover. None are associated with transcriptional promoters (9,15,24), as seen for crossover hotspots in yeast (25). Crossovers in the MS32 and TAP2 hotspots are almost always simple, with very few more complex events showing a patchwork of DNA from both haplotypes at the site of crossover [1 in 176 crossovers for TAP2, 4 in 187 events at MS32 (16)]. Such events almost certainly result from patchwork heteroduplex repair at the site of recombination leading to crossover with conversion of adjacent markers. Curiously, these events are relatively much more common in CMT1A unequal exchanges (15), suggesting possible differences in the processing of unequal and equal crossover complexes in the human genome. It is also likely that gene conversions without crossover occur in the TAP2 hotspot and contribute to haplotype diversity. For example, haplotypes g and h (Fig. 1B) are identical over domains 1 and 3 but show a switch of three adjacent markers (T18–20) within the hotspot, consistent with conversion but potentially explicable by two successive crossover events. Further analysis of conversion activity within human crossover hotspots will require the development of methods for screening sperm for such relatively subtle events (26).

The availability of three well-defined recombination hotspots (TAP2, MS32, CMT1A/HNPP) has allowed us to search for shared sequence motifs previously suggested as being involved in crossover activity. These include the bacterial g sequence (27), the minisatellite core sequence [for CMT1A (15)], homologues of mouse MT element and Lmp2 crossover hotspot motifs [for CMT1A (14)], (TGGA) repeat, and translin binding motifs (28; J.A.L. Armour, personal communication). None of these motifs were significantly shared by any of the three hotspots compared with the 4 kb of recombinationally less active DNA flanking each of these hotspots (data not shown). Furthermore, none of these hotspots shared significant regions of similarity identifiable by dot matrix analysis or by searching for any significantly shared motifs as short as 9 bp. These findings are similar to those from comparisons of different human minisatellites that also failed to find any hints of primary sequence
determinants of recombinational activity (24). It therefore appears that hotspot activity in the human genome is not predictable from primary DNA sequence but may instead reflect open chromatin domains, as in yeast (29–31), that allow access of the meiotic recombinational machinery.

The TAP2 hotspot and nucleotide diversity

Definition of the TAP2 hotspot was greatly facilitated by the abundance of SNPs in this region which contrasts with the low levels of polymorphism in the TAP1 and TAP2 loci themselves (20). Domains 1 and 3 either side of the recombination hotspot show a nucleotide diversity (θ) estimated from the normalized number of variant sites (32) of 0.0015 and 0.0013, respectively, similar to estimates obtained from more global surveys of non-coding and silent substitution polymorphisms in the human genome [θ = 0.0009–0.0011 (33,34)]. In contrast, domain 2 spanning the hotspot shows a significantly higher diversity (θ = 0.0055, χ²[1 df] = 17, P < 0.001). A similar trend is seen with estimates of the observed heterozygosity per base pair (0.0018, 0.0054 and 0.0009 for domains 1, 2 and 3, respectively), suggesting some connection between SNP diversity and recombinational activity. One possibility is that recombination is either base mutagenic or exposes SNPs to biased mismatch repair in heteroduplex DNA formed in recombination complexes and that SNPs favoured in repair can move relatively easily to fixation by meiotic drive. This predicts that the crossover hotspot would show accelerated evolutionary divergence. However, comparison of human, chimpanzee and gorilla sequences shows no evidence for sequence hyperdivergence in domain 2, either across all species or for those substitutions that have accumulated specifically in the human lineage subsequent to the divergence of humans and chimps (data not shown). This suggests either that the hotspot is not a site of accelerated evolution or that it has only appeared very recently in human evolution. The alternative and more plausible explanation for the SNP cluster is therefore selection acting on the MHC. For example, selective sweeps operating on loci both upstream and downstream of the hotspot would reduce SNP diversity in domains 1 and 3, but less so in domain 2 where active recombination would allow variants to escape extinction by recombining onto selectively favoured haplotypes.

Recombination and linkage disequilibrium

Analysis of sperm crossover activity in the LMP7–TAP2 region has allowed us to investigate directly the relationship between LD and recombination frequency for all pairs of markers tested in the two men analysed, and to test for goodness-of-fit with the theoretical prediction for a finite population at crossover/drift equilibrium (35) (Fig. 5). For the markers that each show similar frequencies of both alleles and which define the underlying haplotype frameworks, there is a general trend towards lower LD at higher crossover frequencies, as expected, with D′ rather than D giving the better correlation with crossover frequency. However, the scatter even with D is substantial, and greater than can be accounted for by sampling errors in LD and crossover frequency estimates. Examples of major anomalies include markers T17 and T23 that bracket the peak of crossover activity in man 2 (Fig. 4) but which show strong disequilibrium (D = 0.82, 95% CI 0.60–0.93), and markers T27–29 which are in free association with marker T31 (D = 0.03, 95% CI 0–0.29) yet show a very low crossover frequency with this marker. For
pairs of markers where at least one marker shows a low minor allele frequency, there is no correlation at all between LD and crossover frequency (Fig. 5), indicating that new mutation followed by haplotype drift, and not recombination, is instrumental in determining levels of disequilibrium for low frequency alleles. The most notable example of such an anomaly is for markers T5A, T35* and T44C which span the hotspot but are in absolute LD; these three variants only exist on haplotype p seen five times in the survey (Fig. 1B) and have all presumably arisen by recent mutation on a haplotype which then attained a significant population frequency without being disrupted by recombination. We therefore conclude that disequilibrium measures are poor predictors of recombination frequency within this region of the MHC and that the use of LD to infer hotspots should be treated with caution, as noted previously (12). However, selection acting on the MHC could substantially influence haplotype frequencies and LD measures, and it will be important to extend this analysis to genomic regions less prone to selection to see whether there is an improved correlation between LD and recombination frequency.

Hotspot activity in male and female meiosis

Although the correlation between the standardized LD coefficient $\Delta$ and sperm crossover frequency is poor, the overall trend is consistent with an effective population size $N_e$ for humans of $\sim$100 000 (Fig. 5), an order of magnitude higher than usual estimates (34,36). This discrepancy could result from the sperm recombination rate in the two men tested (who share a common haplotype h) being unusually low, although analysis of a third man heterozygous for haplotypes d and r (Fig. 1B) showed a similar sperm crossover frequency in the TAP2 region (data not shown). Another explanation is that the TAP2 hotspot is more active in female than male meiosis (specifically, a 20-fold increase would yield a value of $N_e$ of 10 000). The 11 class II crossovers analysed by Cullen et al. (9) were located in a 0.74 cM interval (37) and would therefore suggest that the hotspot but are in absolute LD; these three variants only exist on haplotype p seen five times in the survey (Fig. 1B) and have all presumably arisen by recent mutation on a haplotype which then attained a significant population frequency without being disrupted by recombination. We therefore conclude that disequilibrium measures are poor predictors of recombination frequency within this region of the MHC and that the use of LD to infer hotspots should be treated with caution, as noted previously (12). However, selection acting on the MHC could substantially influence haplotype frequencies and LD measures, and it will be important to extend this analysis to genomic regions less prone to selection to see whether there is an improved correlation between LD and recombination frequency.

Figure 5. Relationship between crossover frequency and linkage disequilibrium in the TAP2 gene region. Sperm recombination frequencies between each pair of informative markers in each of the two men analysed (Fig. 4, crossovers in orientation A and B pooled) were compared with the LD coefficients $\Delta$ and $\Delta'$ determined from data in Figure 1. Recombination data for both men are shown, with separate plots for marker pairs each showing a minor allele frequency ($<0.2$, markers T30, T32, T35 and T36, bottom; all suc...
MATERIALS AND METHODS

SNP detection and haplotyping

Four overlapping 2.6 kb segments covering 10 kb of the TAP2 region were PCR-amplified from human genomic DNA and resequenced using BigDye Terminators (ABI, Warrington, UK) on an ABI 377 Automated Sequencer, with sequencing primers spaced 400–500 bp apart. SNPs were identified using ABI AutoAssembler software. New primers targeted to monomorphic regions were designed in cases where polymorphisms were found to affect PCR or resequencing primer sites, resulting in apparent loss of heterozygosity. The corresponding region in chimpanzee and gorilla was amplified as a series of overlapping ~1 kb amplicons, each of which was sequenced using the PCR primers. Human SNPs were typed by ASO hybridization to dot blots of PCR products using a tetrathymethylammonium chloride (TMAC) hybridization protocol (39) modified as follows. Dot blots on Hybond N° membranes (Amersham, Little Chalfont, UK) were prehybridized at 53°C for 10 min in 3 M TMAC, 0.6% SDS, 10 mM Na phosphate pH 6.8, 1 mM EDTA, 0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1% BSA (fraction V; Sigma, St Louis, MO) and 4 µg/ml yeast RNA, and hybridized in the same buffer for 1 h with 3 ng/ml 32P-labelled ASO (18-mer, with the SNP site located 8 nt from the 5′ end) in the presence of 60 ng/ml of unlabeled allelic ASO as competitor plus 10 µg/ml single-stranded herring sperm DNA. Filters were washed in 3 M TMAC, 0.6% SDS, 10 mM Na phosphate pH 6.8, 1 mM EDTA at 56°C for 20 min, rinsed in 3x SSC at room temperature, and autoradiographed. TAP2 haplotypes were established by allele-specific PCR directed to heterozygous SNP sites near the centre of the TAP2 region, in combination with universal (not allele-specific) primers at either end of the resequenced region, to amplify from genomic DNA two overlapping 4–7 kb regions per haplotype which were subsequently typed by ASO hybridization. Further details of SNPs and ASOs are provided at: http://www.le.ac.uk/genetics/ajj.

Sperm crossover detection and mapping

Allele-specific primers used were:

- T4T, 5′-TCC ACT TAC GTA CCG CAC CG-3′;
- T4A, 5′-TGC ACT TAC GTA CCG CAC CG-3′;
- T6/7TG, 5′-ACA AGA ACC AAA GCT CAG TAC G-3′;
- T6/7CC, 5′-ACA AGA ACC AAA GCC CAG TAC C-3′;
- T37C, 5′-CTC ACT TAC GTA CCG CAC CG-3′;
- T37A, 5′-TCC ACT TAC GTA CCG CAC CT-3′;
- T39, 5′-CTT CTA CCT TCC GCC ATG ATT T-3′;
- T39A, 5′-CTT CTA CCT TCC GCC ATG ATT G-3′.

Long PCR was carried out as described previously (17) with standard cycles of 96°C, 30 s; annealing temperature (T_a), 30 s; 70°C, 4.5 min. Optimized amplification conditions are given below as T_a number of cycles at that temperature.

Semen was provided from volunteers instructed to wash their penis prior to ejaculation, to minimize contamination with partner’s DNA. The preparation of sperm and blood DNA and all subsequent manipulations were carried out under conditions designed to minimize the risk of contamination (40,41). Genomic DNA was digested with HindIII, which cleaves outside the test interval. Multiple 8 µl PCRs each containing up to 144 ng DNA were amplified by touchdown long PCR using 0.2 µM allele-specific primers T4T plus T39+ (67°C, 3 cycles; 65°C, 10 cycles; 63°C, 14 cycles) or T4A plus T39 (66°C, 3 cycles; 64°C, 10 cycles; 63°C, 15 cycles). PCR products were digested with S1 nuclease to remove single-stranded DNA including panhandle artefacts (16) and reamplified with the nested allele-specific primers T6/7TG plus T37A (66°C, 4 cycles; 64°C, 10 cycles; 60°C, 15 cycles) or T6/7CC plus T37C (66°C, 4 cycles; 63°C, 10 cycles; 61°C, 14 cycles), respectively. Aliquots of these secondary PCR products were analysed by agarose gel electrophoresis and Southern blot hybridization with a 32P-labelled 2.4 kb probe extending from the beginning of TAP2 intron 2 to the beginning of intron 4. Secondary PCRs showing products from crossover molecules were re-amplified (64°C, 5 cycles; 63°C, 21 cycles) using the nested universal primers T2.0 (5′-CCT CAT AGG CTG AAG GTG CG-3′) and T7.5 (5′-ACA CTC ATG TCC ACA GCA GC-3′), and the status of SNPs determined by dot-blot ASO hybridization. The number of crossovers per interval was Poisson-corrected, as described elsewhere (16), for possible instances of more than one identical recombinant molecule being present in the same initial PCR. In practice, the effect of this correction was negligible except for crossovers in the T17/T23 interval in man 2, the frequency of which was increased by 18%. The number of input amplifiable molecules in the crossover analysis was estimated by Poisson analysis of 60 aliquots of HindIII-digested genomic DNA (2.7–5.4 pg DNA per aliquot) using long PCR amplification with universal primers spanning the test interval (41). The single molecule efficiency was consistently estimated at 50% (SE 12%), corresponding to one amplifiable TAP2 molecule per 6 pg of genomic DNA. This efficiency was further checked by analysis of blood DNA from an individual heterozygous at sites T4, T67, T37 and T39 but showing the opposite linkage phase between domain 1 and domain 3 markers (i.e. all TAP2 molecules in this man are ‘recombiant’ in the crossover assays); Poisson analysis using allele-specific primer combinations and cycling conditions exactly as in the crossover assay showed the same 50% efficiency in both orientations of crossover detection, indicating no significant loss of recombinant molecules under the conditions of crossover analysis.

ACKNOWLEDGEMENTS

We are grateful to Jane Blower and volunteers for supplying semen samples, to Kathryn Lilley for oligonucleotide synthesis and assistance with automated sequencing, to Yuri Dubrova for statistical advice and to colleagues for helpful discussions. This work was supported in part by an International Research Scholars Award to A.J.J. from the Howard Hughes Medical Institute and in part by grants from the Wellcome Trust, Medical Research Council, Biotechnology and Biological Sciences Research Council and the Royal Society.

REFERENCES
