Meiotic recombination hot spots and human DNA diversity

Alec J. Jeffreys*, J. Kim Holloway, Liisa Kauppi, Celia A. May, Rita Neumann, M. Timothy Slingsby and Adam J. Webb

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

Meiotic recombination plays a key role in the maintenance of sequence diversity in the human genome. However, little is known about the fine-scale distribution and processes of recombination in human chromosomes, or how these impact on patterns of human diversity. We have therefore developed sperm typing systems that allow human recombination to be analysed at very high resolution. The emerging picture is that human crossovers are far from randomly distributed but instead are targeted into very narrow hot spots that can profoundly influence patterns of haplotype diversity in the human genome. These hot spots provide fundamental information on processes of human crossover and gene conversion, as well as evidence that they can violate basic rules of Mendelian inheritance.

Keywords: recombination; crossover; conversion; hot spot; haplotype; linkage disequilibrium

1. INTRODUCTION

This review summarizes our recent work on characterizing the molecular basis of meiotic recombination in the human genome. We are attempting to address three simple and inter-related questions. First, how are meiotic exchanges distributed along human chromosomes at the kilobase level of resolution? Second, what processes operate during human meiotic recombination and to what extent do model organisms such as yeast provide information relevant to understanding human recombination? And third, how do patterns and processes of recombination influence sequence and haplotype diversity in human chromosomes?

These are fundamental questions relevant to understanding the dynamics of human DNA evolution and the origins of human populations (Pääbo 2003). They also underpin genetic association analysis of complex disease (Jorde 2000) by providing a mechanistic basis for interpreting haplotype structures in the human genome. Despite their importance, our understanding of these basic issues has until recently been very limited. The reason is simple: meiotic crossovers are so spread out over the human genome that markers even a megabase apart will only recombine at a frequency of 1% or so. Linkage maps determined by classic family analyses are therefore limited in resolution to 0.1–1 cM, corresponding to 0.1–1 Mb DNA on average. This approach would require the analysis of 50 000 children of fully informative parents to identify just one crossover in a chosen kilobase interval: this is clearly impractical!

The alternative approach is LD analysis, which increases the potential resolution by using haplotype diversity to infer the recombination events that have accumulated, not in a single family, but instead over the many thousands of generations in the history of a contemporary population. This approach, first used to identify a putative recombination hot spot near the β -globin gene (Chakravarti *et al.* 1984), is, however, contentious (Hedrick 1987). Many factors in addition to recombination can influence levels of LD between markers; these include recurrent mutation and more importantly population processes such as selection, genetic drift, population bottlenecks, migration and admixture. As a result, neither recombination rates nor distributions can be safely deduced from LD data alone. Also, haplotype diversity information does not readily distinguish between the two products of meiotic recombination, namely crossover and gene conversion.

Despite these limitations, there is clear evidence that meiotic recombination events are, to some extent at least, non-randomly distributed along human chromosomes, both at the megabase level as revealed by comparing physical and linkage maps (Yu *et al.* 2001), and at the kilobase level as shown by family and single sperm typing near the β -globin gene (Smith *et al.* 1998; Schneider *et al.* 2002), the *PGM1* gene (Yip *et al.* 1999) and in the MHC locus (Cullen *et al.* 1995, 1997, 2002). However, the latter studies still lack the resolution needed to properly define recombination distributions and processes.

We have therefore developed an alternative two-tier strategy, applicable in principle to any region of the human genome, which is capable of analysing recombination events at a level of resolution of 0.0001 cM or less. The first stage is to use LD analysis of SNP genotypes to screen the target region for evidence of localized historical recombination (figure 1a). Various outcomes are possible. First, LD may decrease slowly and fairly uniformly with physical distance between markers, in which case

* Author for correspondence (ajj@le.ac.uk).

One contribution of 18 to a Discussion Meeting Issue 'Replicating and reshaping DNA: a celebration of the jubilee of the double helix'.

Figure 1. Strategy for high-resolution analysis of meiotic crossovers in human DNA. (a) LD analysis. A chosen target region is subjected to high-density SNP discovery, followed by genotyping all SNPs in a reference human population. The haplotypes so revealed give information on LD and thus indirectly on the recombination events that have accumulated in the target DNA during the history of the population. Several outcomes are possible: (i) alleles show no correlation between SNPs, implying that the entire region has been subjected to extensive recombinational turnover; (ii) blocks exist in which markers show strong LD (LD blocks); free association between blocks implies clustering of historical recombination; and (iii) all markers are strongly associated, suggesting suppressed recombination across the entire region. (b) Sperm crossover analysis. A man is identified with multiple heterozygous SNP sites across the target region. Two rounds of repulsion-phase allele-specific PCR are used to selectively amplify recombinant molecules from batches of sperm DNA. Reciprocal exchanges (black-white, white-black) can be recovered using appropriate PCR primers. (c) Examples of crossover detection in the *TAP2* hot-spot region (Jeffreys *et al.* 2000) (see figure 2). Each PCR was seeded with DNA equivalent to 24 000 haploid genomes and the secondary PCR products were detected by agarose gel electrophoresis and Southern blot hybridization. The intense signals in sperm DNA are derived from genuine recombinant molecules, while the much fainter blood signals arise from bleed-through of one haplotype, as well as from PCR artefacts that arise by strand jumping between haplotypes.

recombination events (at least those generating haplotypes that have survived in the population) are likely to have been randomly distributed across the target. Second, there may be little if any LD even between very closely linked markers, implying extreme recombination activity over the entire region. Third, all markers may be in intense LD,

suggesting but not proving that the whole region is recombinationally inert. Fourth, there may be regions of intense LD (LD blocks) separated by intervals of free association. This outcome would suggest clustering of historical recombination events, though does not prove the existence of a localized hot spot: chance clustering of a few lucky

Figure 2. LD and sperm crossover activity in an 8 kb region of the *TAP2* gene. (a) Patterns of LD between 37 different SNP markers, established by genotyping a panel of 50 UK north Europeans. Each marker, shown as a tick below and to the right of the plot, is compared with all other markers. Lewontin's (1984) D' measure of LD for each marker pair is colour coded as indicated and plotted in the lower right section. ($|D'|$ is a useful LD measure for recombination analysis, taking the value of 0 for pairs of markers in free association, and 1 for markers showing complete association with at most only three of the four possible haplotypes; three haplotypes can be explained by mutation without any need to invoke recombination or recurrent mutation.) The corresponding likelihood ratio versus free association is plotted in the upper left section. (i) LD profiles for all markers; (ii) results from SNPs with a MAF of at least 0.2. Two clear LD blocks were identified. (b) Sperm crossover activity in this region. Fifty-five crossover molecules were recovered from 1.3×10^6 haploid genome equivalents of sperm DNA from a man heterozygous at markers indicated above the plot, giving a recombination rate of 8×10^{-5} per sperm in this region. The numbers of crossovers in each interval between heterozygous markers (italics) were used to estimate the recombination activity in cM Mb⁻¹. The dashed line shows the mean rate of crossover in the human genome. (Data are taken from Jeffreys *et al.* (2000).)

historical crossovers is perfectly capable of generating a block-like LD structure (Wang *et al.* 2002).

The next stage is to recover crossover molecules directly from large batches of sperm DNA by using allele-specific PCR directed to SNP sites that are heterozygous in the sperm donor (figure 1*b*). Thus, SNPs are used both as markers and as physical selectors of recombinant molecules. If LD blocks are found, then the assay is targeted to the region extending from one LD block to the next to see whether the inter-block region contains a genuine recombination hot spot. However, the technique is not trivial. It demands extreme allele-specificity of the PCR primers used to differentially amplify recombinant molecules, and also needs to minimize jumping PCR artefacts arising by annealing between single-stranded DNA generated from each of the parental haplotypes. Nevertheless, once optimized, the technique can be applied to targets up to 10 kb long and can survey up to 50 000 sperm per PCR, allowing recombinant molecules to be quantitatively recovered and subsequently mapped from millions of sperm (Jeffreys *et al.* 2000). Blood DNA provides an important negative control, allowing true meiotic exchanges to be distinguished from PCR artefacts (figure 1*c*). It is important to stress that the method cannot be applied to any man, but only to those carrying SNP heterozygosities at the selector sites and at additional internal markers needed to map sites of exchange.

2. THE *TAP2* GENE: A TEST CASE

To test this two-tier strategy, we focused on the *TAP2* gene located in the class II region of the MHC (Jeffreys *et al.* 2000). Extensive family analyses had identified MHC-recombinant children who showed some evidence for clustering of exchange points within this gene (Cullen *et al.* 1995, 1997). We commenced by resequencing DNA around this putative hot spot to discover as many SNPs as possible. All SNPs were then genotyped in a panel of north European semen donors, allowing both LD analysis and the identification of informative donors for subsequent recombination analysis. LD analysis of all SNPs either on haplotype data or on unphased diploid genotype data (using maximum-likelihood approaches to extract haplotypes from which LD could be estimated) revealed an intense LD block over the first half of the test region, with most markers in strong and highly significant LD. By contrast, the picture from the second half was chaotic, with some markers in free association and others in complete LD (figure 2*a*). However, exclusion of low-frequency SNPs dramatically simplified this picture, revealing two clear LD blocks separated by a 2 kb region of free association in which only the closest markers showed significant LD. The reasons for this clarification are twofold. First, markers with low-frequency alleles have little statistical power to measure LD levels accurately. Second, and more seriously, rare alleles are likely to have arisen by mutation relatively recently on haplotypes that have not yet had the opportunity to be scrambled by recombination (Jeffreys *et al.* 2001). Thus rare alleles will tend to identify extended young haplotypes that can span regions of active recombination, whereas common (and thus most likely old) alleles give information on ancient haplotypes maximally disrupted by recombination.

To test whether this region of localized LD breakdown inside the *TAP2* gene corresponded to a crossover hot spot, we analysed millions of sperm for exchanges between the two LD blocks (figure 2*b*). Mapping exchange points revealed a highly localized hot spot *ca.* 1 kb wide inside the second intron of the gene and located within the interval of LD breakdown. Flanking DNA showed much lower levels of crossover activity, below the mean rate of 1 cM Mb⁻¹ seen across the human genome. Exchanges appeared to be smoothly and symmetrically distributed across the hot spot, and recombination was reciprocal in that indistinguishable crossover rates and distributions were seen when both orientations of crossovers were analysed (figure 1*b*). The only precisely mapped maternal crossover identified in families (Cullen *et al.* 1995) was located within the sperm hot spot, suggesting that the same hot spot functions in both male and female meiosis. However, the rates are not identical; the crossover frequency in sperm across the hot spot is 8×10^{-5} per gamete, while the female rate estimated from (extremely limited) family data (Cullen *et al.* 1997) and from population genetic approaches relating LD to recombination rate is roughly 30-fold higher (Jeffreys *et al.* 2000).

3. THE MHC CLASS II REGION: THE BIGGER PICTURE

To gain a better view of LD block structure and the general prevalence of crossover hot spots, we broadened this analysis to a 240 kb segment extending upstream of *TAP2* (Jeffreys *et al.* 2001). Previous analyses of MHC-recombinant children had shown that this entire region recombines at a rate fairly close to the genome average of 1 cM Mb⁻¹, with clear evidence for at least some degree of crossover clustering, particularly in the region between the *HLA-DOA* and *RING3* genes (Cullen *et al.* 1997).

Low-density SNP genotyping was used to broadly map LD block structure in this region, and then high-density SNP analysis was used to refine LD block boundaries (figure 3). The resulting LD profile revealed a remarkably crisp succession of three extended LD blocks 40–90 kb long, all freely associating with each other. The inter-block regions contained much smaller LD patches just a few kilobases in length, raising the possibility that there might be more than one hot spot per inter-block interval. Sperm typing these regions confirmed this prediction, revealing a cluster of three hot spots of various intensities around *HLA-DOA* and two hot spots in and near *HLA-DMB*. Each showed a morphology strikingly similar to the *TAP2* hot spot.

The recombinational behaviour of DNA within LD blocks remains much less certain. Historical recombination events have accumulated in these regions, as shown by the decreasing levels of association between markers with physical distance within LD blocks (figure 3). Population genetic analysis and the direct estimation of sperm exchange rates at the ends of LD blocks next to hot spots both point to a very low recombination rate of very roughly 0.05 cM Mb⁻¹ (Jeffreys *et al.* 2001). If correct, then *ca.* 95% of all sperm crossovers in this MHC region cluster into these six hot spots, with the most intense (hot-spot *DNA3*) accounting for some 72% of all exchanges in this region. Satisfyingly, the summed sperm crossover rate

over the entire region is indistinguishable from the paternal rate measured in families (2×10^{-3}). However, considerable caution is needed in interpreting the behaviour of LD blocks. Detection of sperm crossovers would be formidably difficult at this very low rate of 0.05 cM Mb^{-1} (corresponding to one exchange per kilobase per 2×10^6 sperm!) and has yet to be systematically investigated. Haplotype analysis in one LD block points to the likely involvement of gene conversion in LD block diversification (Kauppi *et al.* 2003), creating problems in estimating crossover rates from population data. Finally, it is possible that LD blocks could hide additional hot spots that have failed to leave their imprint on block structure; direct evidence for this possibility has come from recent studies of a non-MHC hot spot near a minisatellite that shows strong association between markers right across the hot spot, even for high-frequency SNP alleles (A. J. Jeffreys and R. Neumann, unpublished data).

4. RECOMBINATION IN THE PSEUDOAUTOSOMAL PAIRING REGION PAR1

The 2.6 Mb long PAR1 region at the end of chromosomes Xp/Yp undergoes an obligate crossover at male meiosis, essential for sex chromosome pairing and correct disjunction. As a result, the entire region appears to serve as a male-specific recombination hot domain with a mean rate of male crossover of 20 cM Mb^{-1} , some 20-fold higher than the mean autosomal rate of male recombination. Both linkage analysis (Rouyer *et al.* 1986) and single sperm typing (Lien *et al.* 2000) have suggested that most or all of PAR1 is fairly uniformly active in recombination. Higher-resolution analysis of PAR1 would therefore provide a system for the study of sperm crossover rates and distributions, and their relationship to LD, in a region where female recombination can be essentially ignored (May *et al.* 2002).

Haplotype analysis in and around the *SHOX* gene, located in PAR1 some 500 kb from the Xp/Yp telomere, revealed an LD pattern dramatically different from that seen in the MHC (figure 4a). Significant LD extended at most only a few kilobases, and evidence that associated markers were organized into LD blocks was at best equivocal. At first glance, this LD pattern is exactly what would be expected if the entire region were recombining uniformly at the high rate of 20 cM Mb^{-1} in males. However, sperm analysis of a region within the *SHOX* gene, chosen on the basis that it was flanked on each side by groups of markers showing at least some degree of association, revealed a highly non-random distribution of crossover exchange points (figure 4b). The hot spot so defined was intense, with a peak activity of 300 cM Mb^{-1} , higher than at any of the MHC hot spots. Thus hot spots are not confined to the MHC and must also play at least some role in PAR1 recombination. The key question now is how hot spots are distributed along PAR1. Once additional hot spots are identified, they might provide an experimentally tractable system for analysing the recombinational behaviour of DNA not only within but also between hot spots. However, the identification of additional PAR1 hot spots will not be trivial, given that LD patterns do not give clear clues as to their whereabouts, at least around the *SHOX* gene.

5. CROSSOVER HOT SPOTS AND MINISATELLITES

Minisatellites include some of the most unstable tandem repeat DNA loci in the human genome (see Bois & Jeffreys 1999). Repeat instability is largely restricted to the germline and, for the most unstable loci, frequently involves recombinational interactions between alleles. These interactions usually result in the gene conversion-like transfer of blocks of repeats from one allele to the other, without exchange of flanking markers (figure 5a). These events are almost always restricted to the repeat array and can sometimes be very complex, with the transferred segment being a patchwork of DNA from various regions of the donor allele, and with the insertion site in the recipient allele also showing rearrangements such as target site duplications. In rare cases, conversions can include SNP markers adjacent to the repeat array. Conversions at some but not all minisatellites are heavily polar, being largely restricted to one end of the repeat array and pointing to the potential role of flanking DNA in driving repeat instability.

The meiotic recombinational basis of minisatellite instability prompted us to analyse crossover activity near minisatellites, using approaches similar to those used to characterize the MHC and *SHOX* hot spots. All three loci tested to date show significant levels of exchange activity, not only within the repeat array—leading to unequal and equal exchanges between alleles—but also in the flanking DNA (figure 5b) (Jeffreys *et al.* 1998a,b; Buard *et al.* 2000). These crossovers, however, occur at a much lower rate than the conversion-driven mutations within the repeat array. In one case analysed in detail (minisatellite MS32), these crossover breakpoints define a narrow and intense recombination hot spot centred upstream of the polar end of the minisatellite (figure 5c) (Jeffreys *et al.* 1998a). We have provisional evidence for a very similar hot spot upstream of the polar end of a second minisatellite (MS31) (C. Hollies and A. J. Jeffreys, unpublished data). Current evidence strongly suggests that it is the hot spot that drives minisatellite instability, and neatly explains polar mutation because only the end of the minisatellite embedded within the hot spot will be subjected to recombinational turnover. It therefore appears that minisatellites are occasionally generated as parasitic by-products of recombination hot-spot activity, and thus may serve as surrogate markers for hot spots. However, while minisatellites may associate with hot spots, the converse is not always true: none of the MHC hot spots or the *SHOX* hot spot contains repeat DNA.

6. DISTRIBUTION OF HOT SPOTS

Are there any clear rules that govern the location of human crossover hot spots? From the limited repertoire of hot spots identified to date, the answer appears to be no. In yeast, recombination-initiating meiotic DSBs preferentially associate with promoters (Baudat & Nicolas 1997; Gerton *et al.* 2000). Only the weakest human hot spot (*DNA1*, figure 3) localizes in or near a promoter. The others are found in a wide variety of genomic locations, with hot-spot centres located in exons, introns, dispersed repeats such as *Alu* elements and retroviral LTRs, and single-copy intergenic DNA (table 1). The similar overall

Figure 3. LD and sperm crossover hot spots in a 240 kb segment of the class II region of the MHC. Levels of LD were established for 151 different SNP markers with a MAF of at least 0.25 and plotted as in figure 2a. The apparent high density of SNPs near sites of LD breakdown is the result, at least in part, of our greater effort to discover additional SNPs needed to define LD block boundaries and to recover and map sperm crossovers. The regions of DNA assayed for sperm recombination are shown below, with the locations of crossover hot spots indicated. Arrows are scaled to reflect relative crossover activity. (LD and crossover data are taken in part from Jeffreys *et al.* (2001).)

recombination rates in genomes of widely differing sizes but similar gene content has led to speculation that there may be a strict association between genes and recombination (Thuriaux 1977). The fact that MHC hot spots appear not to be present (as far as we know) in the longest regions of non-coding DNA (figure 3) would lend some support to this idea, though there is clearly not a one-to-one relationship between genes and hot spots, as also noted in maize (Yao *et al.* 2002). However, the minisatellite MS32 hot spot is not located in a gene-rich region of DNA: the flanking megabase contains only eight known genes verified by Ensembl, and the minisatellite is located centrally in an extended 77 kb long non-coding region and is thus situated remote from any genes. Clearly, many more data will be needed to resolve this issue. In this

context, it will be of considerable interest to see how emerging haplotype maps of human chromosomes (Dawson *et al.* 2002; Gabriel *et al.* 2002; Phillips *et al.* 2003) correlate with gene distribution.

7. PROPERTIES OF HOT SPOTS

Several unifying features of hot-spot activity have emerged (Jeffreys *et al.* 2001). First, almost all crossovers are simple, involving a single site of exchange between recombining haplotypes. Only 1% or so of crossovers are more complex, showing switching back and forth between haplotypes at the site of exchange. These rare, complex exchanges presumably result from patchy repair of heteroduplex DNA generated during recombination. Second,

Figure 4. LD and crossover activity in the pseudoautosomal *SHOX* gene. (a) LD levels in north Europeans were determined from 42 different SNPs with a MAF of at least 0.25 and plotted as in figure 2a. Exons in the *SHOX* gene are indicated in black. (b) A 5 kb region was analysed for sperm crossovers in an informative semen donor. Crossover activity was determined from 59 crossover molecules recovered from 27 000 haploid genome equivalents of sperm DNA, and plotted as in figure 2b. (Data are taken in part from May *et al.* (2002).)

most (but not all) hot spots show fully reciprocal exchange, with reciprocal crossover products (figure 1b) arising in sperm at the same rate and with the same distribution across the hot spot. The third shared feature is the symmetrical distribution of exchange points across these crossover hot spots, which enables hot-spot centres to be localized with considerable precision (within ± 30 bp). The fourth feature is their surprisingly uniform width of 1–2 kb

(table 1); this common width, also seen at two mouse MHC hot spots characterized by sperm analysis (Guillon & de Massy 2002; Yauk *et al.* 2003), clearly points to common processes operating within these hot spots. Despite this fairly constant width, the peak intensity varies by orders of magnitude over different hot spots. Additionally, at least some of these hot spots show significant variation in crossover rates between men (Jeffreys *et*

Figure 5. Recombination products at human minisatellite MS32. (a) Structures of MS32 length-mutant molecules recovered from sperm DNA using small-pool PCR methods (Jeffreys *et al.* 1994). Array conversions dominate germline instability and can sometimes be complex. Junction conversions can be identified among these mutants by typing the status of flanking SNP markers. Other types of convertant, in particular flanking conversions as well as array conversions that do not alter allele length, cannot be detected using these approaches. (b) Structures of crossover molecules recovered from sperm DNA using allele-specific PCR methods as in figure 1 (Jeffreys *et al.* 1998a,b). Equal crossovers within the minisatellite generate recombinant repeat arrays but do not alter allele length. (c) Sperm crossover activity near MS32, plotted as in figure 2b. Flanking crossovers, unequal crossovers within MS32 and equal crossovers are shaded in dark grey, white and light grey, respectively. (Data are taken from Jeffreys *et al.* (1998a).)

al. 1998a; Jeffreys & Neumann 2002). Finally, we can find no significant primary DNA sequence similarities between these hot spots; this is not to say that sequence predictors of recombination do not exist, but rather that very extensive collections of hot-spot sequences may be needed to identify potentially fuzzy determinants of recombination. However, at present, it is not possible to predict hot-spot location from human DNA sequences.

8. HOT SPOTS CONTAIN SITES OF RECOMBINATION INITIATION

Sperm crossover analysis identifies sites of crossover resolution, not initiation, and it remains possible that initiat-

ing lesions such as DSBs occur remote from these hot spots. However, the MS32 hot spot is also active in gene conversion, as detected by localized conversion events that alter repeat copy number (figure 5a). Because conversions and crossovers in fungi result from alternative processing of the same initiating lesion, then this points to the MS32 hot spot containing a site of recombination initiation. Similarly, there is evidence for gene conversion within a mouse MHC hot spot not associated with repeat DNA (Guillon & de Massy 2002). We have therefore developed screening and selection strategies that enable large numbers of sperm to be surveyed for conversion events that result in the exchange of one or more markers between alleles but without exchange of flanking markers

Table 1. Properties of human crossover hot spots.
(Data are taken from Jeffreys *et al.* (1998a, 2000, 2001) and May *et al.* (2002).)

hot spot	location	centre point	95% width ^a (kb)	peak activity (cM Mb ⁻¹)
<i>DNA1</i>	class II MHC region	<i>HLA-DOA</i> promoter	1.9	0.4
<i>DNA2</i>	class II MHC region	intergenic, between two <i>Alu</i> repeats	1.3	8
<i>DNA3</i>	class II MHC region	intergenic, in <i>Alu</i> repeat	1.2	100
<i>DMB1</i>	class II MHC region	<i>HLA-DMB</i> exon/intron boundary	1.8	1
<i>DMB2</i>	class II MHC region	intergenic, single-copy DNA	1.2	20
<i>TAP2</i>	class II MHC region	<i>TAP2</i> intron, single-copy DNA	1.0	8
<i>SHOX</i>	Xp/Yp PAR1	<i>SHOX</i> exon	2.0	300
minisatellite MS32	chromosome 1q42.3	intergenic, in retroviral LTR element	1.5	50

^a The width within which 95% of crossovers occur.

(A. J. Jeffreys and C. A. May, unpublished data). Preliminary data confirm that human hot spots are very active in gene conversion, and further that this activity is focused at the centre of the hot spot as defined by crossover distributions. Thus, hot spots contain localized sites or zones of initiation and presumably act as foci for initiating DSBs that can generate either conversion or crossover products.

9. MEIOTIC DRIVE OF MARKERS IN HOT SPOTS

While most hot spots show fully reciprocal crossover (at least in those men analysed), there are exceptions. The best-understood example comes from hot-spot *DNA2* in the MHC (Jeffreys & Neumann 2002) (figure 3). Some men show fully reciprocal exchanges, but others show crossover asymmetry, with reciprocal (A, B) crossovers occurring at the same rate but mapping to different locations within the hot spot (figure 6). However, combining these A and B distributions yields the same overall distribution as seen in men that do not show asymmetry.

By analysing men with various different haplotypes of markers in and around the hot spot, it has been possible to trace the probable cause of this curious phenomenon to a single G/A transitional SNP, located in the middle of the hot spot, which when heterozygous appears to be sufficient to trigger asymmetry. The simplest explanation is that the G variant suppresses recombination initiation at the centre of the hot spot, or equally that the A allele promotes initiation. Crossover asymmetry then naturally arises as a consequence of DSBs arising preferentially on the A haplotype which then undergo resection to produce single-strand ends that invade the G haplotype. Early mismatch repair of the invading strands leads to loss of information from the initiating A chromosome and its replacement, by repair synthesis, with information from the G chromosome. Isomerization and resolution of the recombination complex then leads naturally to asymmetry (see the model in Jeffreys & Neumann (2002) for more details). Thus, the DSB-repair model developed for recombination in yeast (Szostak *et al.* 1983) provides a simple explanation for human crossover asymmetry.

Asymmetry is thus accompanied by the systematic replacement of A-haplotype alleles with markers from the G haplotype within the hot spot (figure 6). It follows that these crossovers in humans must be accompanied by gene conversion of markers adjacent to the site of exchange.

For hot-spot *DNA2*, this conversion process is highly directional, with strongly non-Mendelian over-transmission of the G allele and nearby markers to crossover progeny from G/A heterozygotes. While this meiotic drive in favour of the recombination-suppressing G allele is intense in crossover products, it is weak at the population level given the low rate of crossover in this hot spot. Nevertheless, even low levels of drive can strongly promote the eventual fixation of driven alleles, though surprisingly have little effect on the time taken to fixation of the driven variant (Jeffreys & Neumann 2002).

Crossover asymmetry is not unique to hot-spot *DNA2*. A very similar phenomenon has been seen at the *E_β* hot spot in the mouse MHC, and is strongly associated with haplotypic variation in recombination rates (Yauk *et al.* 2003). Similarly, the minisatellite MS32 hot spot shows evidence for directional conversion, and again there is evidence that this effect can be traced to a single SNP within the hot spot that influences the rate of recombination (Monckton *et al.* 1994; Jeffreys *et al.* 1998a). Thus asymmetry and the meiotic drive of recombination-suppressing variants appears to be a common phenomenon at mammalian hot spots. This raises the fascinating paradox of how hot spots can survive in the face of a process that promotes the fixation of variants that specifically extinguish hot-spot activity (Boulton *et al.* 1997; Jeffreys & Neumann 2002), and raises the possibility that hot spots may be relatively ephemeral over evolutionary time. This is testable.

10. FINAL COMMENTS

It is clear that yeast models of recombination are, to some extent at least, directly applicable to humans. However, the products of meiotic recombination do appear to differ substantially between the two species. We have no genetic evidence for extensive migration of Holliday junctions in human DNA and the generation of long tracts of heteroduplex DNA, as seen in yeast (Borts & Haber 1989). Instead, hot spots seem to be regions within which the entire recombination process (initiation, resection, invasion, branch migration and their resulting crossovers and conversions) is completely trapped.

All evidence to date points to hot spots dominating human recombination and playing a major role not only in structuring human DNA diversity into haplotype

Figure 6. Reciprocal crossover asymmetry and meiotic drive in a recombination hot spot. (a) Detection of asymmetry. Asymmetry arises if reciprocal (A, B) exchanges between the black and white haplotypes occur at the same rate but show exchange points mapping to different locations. If A exchanges are displaced 3' relative to B exchanges, then crossover progeny will show non-Mendelian over-transmission of alleles from the black haplotype for markers closest to the centre of the hot spot, as indicated graphically. (b) Sperm crossover asymmetry at the MHC hot spot *DNA22* located downstream of *HLA-DOA* (formerly *HLA-DNA*) (see figure 3) and centred at two *Alu* repeats. A and B exchanges arise at similar rates, as indicated, but are displaced on average by 430 bp. The strength of asymmetry is such that markers closest to the centre of the hot spot, in particular the central G/A polymorphism arrowed, show strong transmission distortion, with the G allele being transmitted to 87% of crossover products. Heterozygosity at this SNP appears to be sufficient to trigger asymmetry, with meiotic drive specifically in favour of the G allele in all G/A heterozygotes tested. (Data are taken from Jeffreys & Neumann (2002).)

blocks, but also stabilizing these blocks in different human populations even in the face of rapid turnover of the underlying haplotypes (Kauppi *et al.* 2003). However, a note of caution is needed. The regions studied to date are all atypical (the intensely selected MHC, the specialized PAR1 region and an unusually unstable minisatellite), and it is unclear whether these rules of recombination apply more generally in the human genome. The rapidly developing evidence for haplotype block structures in the human genome certainly provides clues in favour of a highly non-random distribution of crossovers at the molecular level (Gabriel *et al.* 2002; Reich *et al.* 2002). However, it remains possible that at least some of this LD structuring comes from the vagaries of population history, rather than from hot spots (Wang *et al.* 2002; Phillips *et al.* 2003). This issue will only be resolved by much more extensive analyses of the relationship between haplotypes and crossover distribution. Sadly, sperm typing cannot be scaled to the whole-genome level. The challenge therefore is to develop alternative methods for studying meiotic recombination in humans, analogous to the initiating DSBs in yeast that can be readily detected biochemically and which provide a surrogate physical measure of genetic recombination activity applicable across the entire genome (Gerton *et al.* 2000). It remains to be seen whether such biochemical readouts of human recombination activity can be developed.

We are grateful to Jane Blower and volunteers for providing the very large number of semen samples needed for this research, and to numerous colleagues for their help and advice, in particular Rhona Borts and Ed Louis for helping to educate us about the mysteries of yeast recombination. We thank the MRC, BBSRC, Wellcome Trust and particularly The Royal Society for their continued support of our research.

REFERENCES

- Baudat, F. & Nicolas, A. 1997 Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc. Natl Acad. Sci. USA* **94**, 5213–5218.
- Bois, P. & Jeffreys, A. J. 1999 Minisatellite instability and germline mutation. *Cell Mol. Life Sci.* **55**, 1636–1648.
- Borts, R. H. & Haber, J. E. 1989 Length and distribution of meiotic gene conversion tracts and crossovers in *Saccharomyces cerevisiae*. *Genetics* **123**, 69–80.
- Boulton, A., Myers, R. S. & Redfield, R. J. 1997 The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl Acad. Sci. USA* **94**, 8058–8063.
- Buard, J., Shone, A. C. & Jeffreys, A. J. 2000 Meiotic recombination and flanking marker exchange at the highly unstable human minisatellite CEB1 (*D2S90*). *Am. J. Hum. Genet.* **67**, 333–344.
- Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Waber, P. G., Boehm, C. D. & Kazazian, H. H. 1984 Non-uniform recombination within the human beta-globin gene cluster. *Am. J. Hum. Genet.* **36**, 1239–1258.
- Cullen, M., Erlich, H., Klitz, W. & Carrington, M. 1995 Molecular mapping of a recombination hotspot located in the second intron of the human *TAP2* locus. *Am. J. Hum. Genet.* **56**, 1350–1358.
- Cullen, M., Noble, J., Erlich, H., Thorpe, K., Beck, S., Klitz, W., Trowsdale, J. & Carrington, M. 1997 Characterization of recombination in the HLA class II region. *Am. J. Hum. Genet.* **60**, 397–407.
- Cullen, M., Perfetto, S. P., Klitz, W., Nelson, G. & Carrington, M. 2002 High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am. J. Hum. Genet.* **71**, 759–776.
- Dawson, E. (and 28 others) 2002 A first-generation linkage disequilibrium map of human chromosome 22. *Nature* **418**, 544–548.
- Gabriel, S. B. (and 17 others) 2002 The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229.
- Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O. & Petes, T. D. 2000 Global mapping of meiotic recombination hot spots and cold spots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **97**, 11 383–11 390.
- Guillon, H. & de Massy, B. 2002 An initiation site for meiotic crossing-over and gene conversion in the mouse. *Nature Genet.* **32**, 296–299.
- Hedrick, P. W. 1987 Gametic disequilibrium measures: proceed with caution. *Genetics* **117**, 331–341.
- Jeffreys, A. J. & Neumann, R. 2002 Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature Genet.* **31**, 267–271.
- Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L. & Armour, J. A. L. 1994 Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* **6**, 136–145.
- Jeffreys, A. J., Murray, J. & Neumann, R. 1998a High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hot spot. *Mol. Cell* **2**, 267–273.
- Jeffreys, A. J., Neil, D. L. & Neumann, R. 1998b Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J.* **17**, 4147–4157.
- Jeffreys, A. J., Ritchie, A. & Neumann, R. 2000 High-resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hot spot. *Hum. Mol. Genet.* **9**, 725–733.
- Jeffreys, A. J., Kauppi, L. & Neumann, R. 2001 Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222.
- Jorde, L. B. 2000 Linkage disequilibrium and the search for complex disease genes. *Genome Res.* **10**, 1435–1444.
- Kauppi, L., Sajantila, A. & Jeffreys, A. J. 2003 Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum. Mol. Genet.* **12**, 33–40.
- Lewontin, R. C. 1984 The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* **49**, 49–67.
- Lien, S., Szyda, J., Schechinger, B., Rappold, G. & Arnheim, N. 2000 Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. *Am. J. Hum. Genet.* **66**, 557–566.
- May, C. A., Shone, A. C., Kalaydjieva, L., Sajantila, A. & Jeffreys, A. J. 2002 Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene *SHOX*. *Nature Genet.* **31**, 272–275.
- Monckton, D. G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A. & Jeffreys, A. J. 1994 Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genet.* **8**, 162–170.
- Pääbo, S. 2003 The mosaic that is our genome. *Nature* **421**, 409–412.
- Phillips, M. S. (and 34 others) 2003 Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nature Genet.* **33**, 382–387.

- Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. & Altshuler, D. 2002 Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genet.* **32**, 135–142.
- Rouyer, F., Simmler, M. C., Johnsson, C., Vergnaud, G., Cooke, H. & Weissenbach, J. 1986 A gradient of sex-linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**, 291–295.
- Schneider, J. A., Peto, T. E., Boone, R. A., Boyce, A. J. & Clegg, J. B. 2002 Direct measurement of the male recombination fraction in the human beta-globin hot spot. *Hum. Mol. Genet.* **11**, 207–215.
- Smith, R. A., Ho, P. J., Clegg, J. B., Kidd, J. R. & Thein, S. L. 1998 Recombination breakpoints in the human β -globin gene cluster. *Blood* **92**, 4415–4421.
- Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. 1983 The double-strand-break repair model for recombination. *Cell* **33**, 25–35.
- Thuriaux, P. 1977 Is recombination confined to structural genes on the eukaryotic genome? *Nature* **268**, 460–462.
- Wang, N., Akey, J. M., Zhang, K., Chakraborty, R. & Jin, L. 2002 Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.* **71**, 1227–1234.
- Yao, H., Zhou, Q., Li, J., Smith, H., Yandau, M., Nikolau, B. J. & Schnable, P. S. 2002 Molecular characterization of meiotic recombination across the 140 kb multigenic *a1-sh2* interval of maize. *Proc. Natl Acad. Sci. USA* **99**, 6157–6162.
- Yauk, C. L., Bois, P. R. & Jeffreys, A. J. 2003 High-resolution sperm typing of meiotic recombination in the mouse MHC E_β gene. *EMBO J.* **22**, 1389–1397.
- Yip, S. P., Lovegrove, J. U., Rana, N. A., Hopkinson, D. A. & Whitehouse, D. B. 1999 Mapping recombination hot spots in human phosphoglucomutase (*PGM1*). *Hum. Mol. Genet.* **9**, 1699–1706.
- Yu, A. (and 10 others) 2001 Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953.

GLOSSARY

- DSB: double-strand break
 LD: linkage disequilibrium
 LTR: long terminal repeat
 MAF: minor allele frequency
 MHC: major histocompatibility complex
 SNP: single-nucleotide polymorphism