Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex

Alec J. Jeffreys[1], Liisa Kauppi[1] & Rita Neumann[1]

There is considerable interest in understanding patterns of linkage disequilibrium (LD) in the human genome, to aid investigations of human evolution and facilitate association studies in complex disease[1–5]. The relative influences of meiotic crossover distribution and population history on LD remain unclear, however[5]. In particular, it is uncertain to what extent crossovers are clustered into 'hot spots,[6–8] that might influence LD patterns. As a first step to investigating the relationship between LD and recombination, we have analyzed a 216-kb segment of the class II region of the major histocompatibility complex (MHC) already characterized for familial crossovers[9]. High-resolution LD analysis shows the existence of extended domains of strong association interrupted by patchwork areas of LD breakdown. Sperm typing shows that these areas correspond precisely to meiotic crossover hot spots. All six hot spots defined share a remarkably similar symmetrical morphology but vary considerably in intensity, and are not obviously associated with any primary DNA sequence determinants of hot-spot activity. These hot spots occur in clusters and together account for almost all crossovers in this region of the MHC. These data show that, within the MHC at least, crossovers are far from randomly distributed at the molecular level and that recombination hot spots can profoundly affect LD patterns.

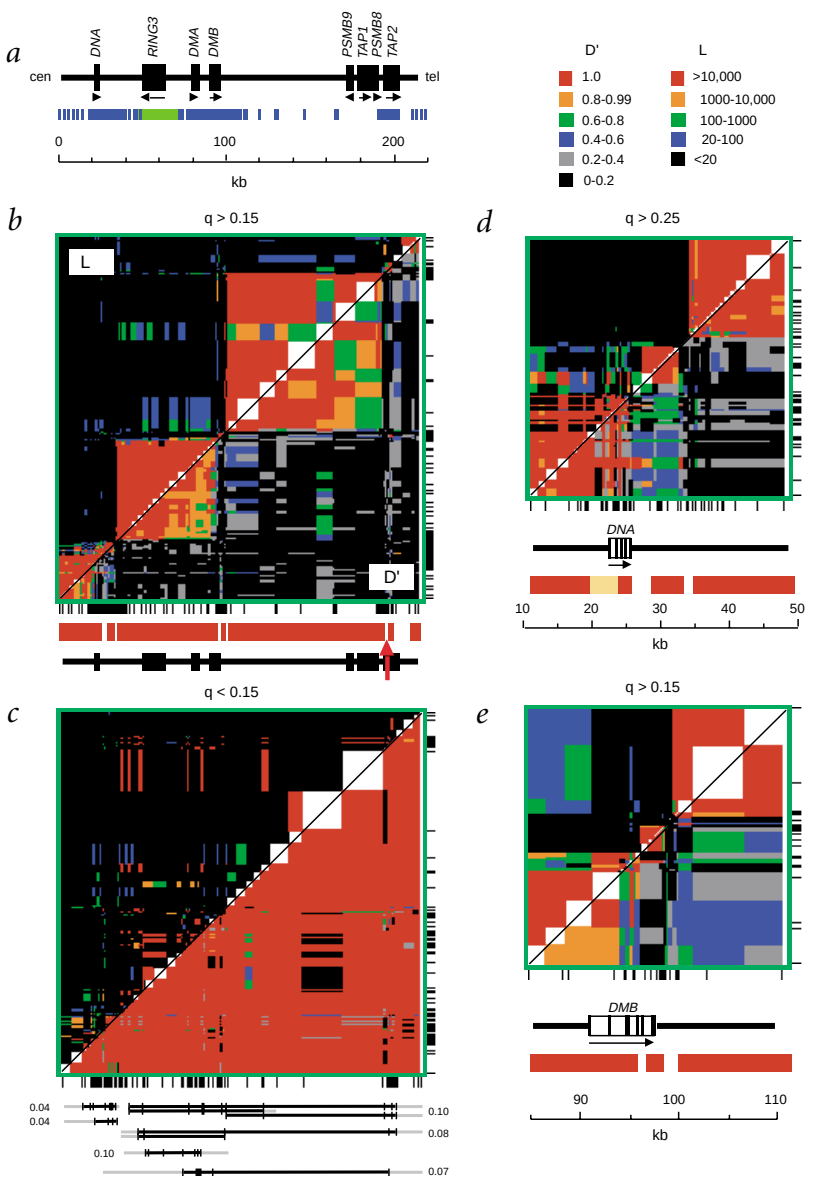

Fig. 1 Patterns of linkage disequilibrium (LD) in the class II region of the MHC. ***a***, The region extending from *HLA-DNA* to *TAP2*. Regions resequenced for SNP discovery are shown in blue; additional polymorphisms in the green region were identified in dbSNP. ***b***, LD in this region in north Europeans. Classic |D'| measures[27] of complete LD (lower right), where |D'|=1 for marker pairs showing only three haplotypes, are shown for all pairs of markers with minor allele frequencies of at least 0.15, together with the associated likelihood ratio (LR) versus free association (upper left), and color-coded as indicated at top right. We determined these measures from unphased diploid genotype data on 179 markers typed in a panel of 50 unrelated UK semen donors. Each point is plotted as a rectangle centered on each SNP (locations shown below and to right of plot) and extending half way to each adjacent marker. Domains showing strong LD are indicated below in red. The location of the previously characterized *TAP2* crossover hot spot[12] is indicated by the arrow. ***c***, Corresponding LD plot for 117 markers with minor allele frequencies less than 0.15. In most cases, the population sample size was too small to detect all four haplotypes, even for marker pairs in free association, and thus most SNP comparisons show D' values of 1 but without significant association. Extended haplotypes at least 30 kb long that are responsible for the significant (LR>10,000:1) long-range associations are shown below, together with haplotype frequencies and the locations of low frequency alleles restricted to each haplotype. The minimum and maximum extent of haplotypic identity for each haplotype lineage are shown in black and grey, respectively. In two cases, there were additional closely related but truncated haplotypes, as shown. ***d***, Expansion of plot *b* in the *HLA-DNA* region, after removal of all markers with minor allele frequencies less than 0.25. Domains of strong LD are shown in red, with a region of partial and erratic reduction of LD indicated in orange. ***e***, Expansion of plot *b* in the *HLA-DMB* region.

[1]*Department of Genetics, University of Leicester, Leicester, LE1 7RH, UK. Correspondence should be addressed to A.J.J. (e-mail: ajj@le.ac.uk).*

The class II MHC region selected for study extends from the *HLA-DNA* to *TAP2* genes (Fig. 1*a*)[10]. Cullen *et al.*[9,11] have shown that the limited numbers of meiotic crossovers detectable in families in this region occur at a slightly higher than genome-average rate of about 2 cM/Mb, with a roughly 4:1 bias towards maternal versus paternal crossovers. Recombinants were non-randomly distributed, with clear clustering in a 4-kb interval between *HLA-DNA* and *RING3,* and possible clustering in *TAP2*. Sperm crossover analysis of a 10-kb interval in the *TAP2* region confirmed a meiotic-crossover hot spot about 1 kb long and flanked by recombinationally inert regions in which single-nucleotide polymorphisms (SNPs) were in strong LD[12]. To see whether LD patterns could be used more generally to predict crossover distribution, we identified SNPs by resequencing at intervals of 5–20 kb across the target and tested associations by genotyping a panel of unrelated British semen donors of European descent. Additional SNP discovery in regions showing breakdown of LD refined the locations of putative recombination hot spots. In total, we discovered 264 SNPs and 22 1–11-bp insertion/deletion polymorphisms in 84 kb of resequenced DNA.

The LD patterns defined by these biallelic markers, plus ten additional SNPs identified near *RING3* in dbSNP[13], are shown in Fig. 1. Markers with high-frequency (>0.15) minor alleles, which are likely to be ancient and thus report on haplotypes maximally disrupted by recombination, reveal three extended domains of intense disequilibrium with abrupt LD breakdown near *HLA-DNA*, *HLA-DMB* and *TAP2* (Fig. 1*b*). Very similar LD profiles are obtained using higher minor-allele-frequency cut-offs (up to 0.35; data not shown), indicating that these are robust. Lower-frequency markers show a very different pattern, in part because of the lack of statistical power but also in that some markers show highly significant long-range associations extending across regions of LD breakdown (Fig. 1*c*). Haplotypes that carry multiple rare alleles creating these associations terminate near the presumptive hot spots in all but one case (Fig. 1*c*). This is consistent with these variants having arisen relatively recently in human evolution on haplotypes that nevertheless have been disrupted, albeit incompletely, by localized recombination.

Higher-resolution LD profiles of the *HLA-DNA* and *HLA-DMB* regions (Fig. 1*d*,*e*) reveal complex patterns of clusters of markers in strong LD within the regions where long-range LD breaks down. To investigate meiotic recombination activity, we subjected each of these regions to crossover analysis in sperm DNA using allele-specific PCR to selectively amplify DNA molecules that had crossed over between one domain of LD and the next (Fig. 2*a*)[12]. Single-molecule crossover products were

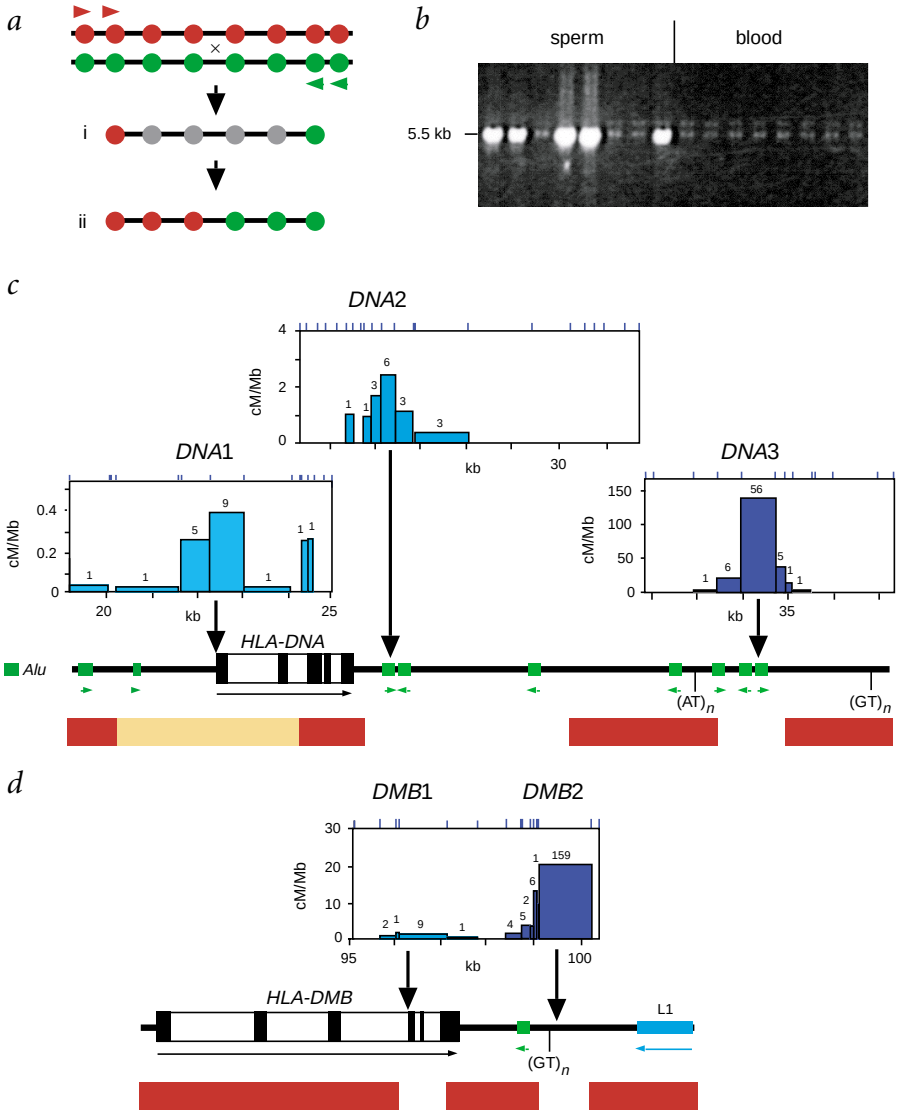

Fig. 2 Analysis of sperm crossover distribution near *HLA-DNA* and *HLA-DMB*. ***a***, Strategy for amplifying crossovers from sperm DNA. Batches of sperm DNA from a man showing multiple SNP heterozygosities (colored circles) across the test region are amplified using repulsion-phase allele-specific PCR directed to the outside SNP sites, followed by re-amplification using nested allele-specific primers, allowing recombinant molecules to be selectively amplified (i). SNP typing locates the crossover interval within each amplified recombinant (ii). ***b***, Examples of crossover detection in sperm and blood DNA from the same man over the 32–37-kb interval downstream of *HLA-DNA*. We seeded each PCR with 1,000 amplifiable molecules of each haplotype (12 ng genomic DNA) and detected re-amplified PCR products by agarose gel electrophoresis and staining with ethidium bromide. ***c***, Sperm crossover distribution in three test intervals around *HLA-DNA* that cover the three regions of complete or partial breakdown of LD; domains of strong LD (Fig. 1*d*) are indicated below in red. Region 1 yielded 19 crossover molecules from 3.7×10^6 amplifiable molecules of each haplotype in the man tested, region 2 gave 17 crossovers in 7.6×10^5 molecules, and region 3 yielded 70 crossovers from 5.7×10^4 molecules. The numbers of crossovers mapping to each interval between adjacent heterozygous SNPs (positions indicated above the plots), shown above the histograms, were used to estimate the recombinational proficiency of each interval in cM/Mb. The centers of hot spots *DNA*1–3 are indicated by arrows. Note the very different scaling of recombination activity in regions 1–3. The $(AT)_n$ and $(GT)_n$ microsatellites in region 3 define the 4-kb interval in which clustering of crossovers detected in families has been observed[9]. ***d***, Corresponding sperm crossover distribution in the *HLA-DMB* region, determined from 190 crossover molecules recovered from 7.0×10^5 progenitor molecules in the man tested. Hot spots *DMB*1 and 2 were not evident in family crossover data[9].

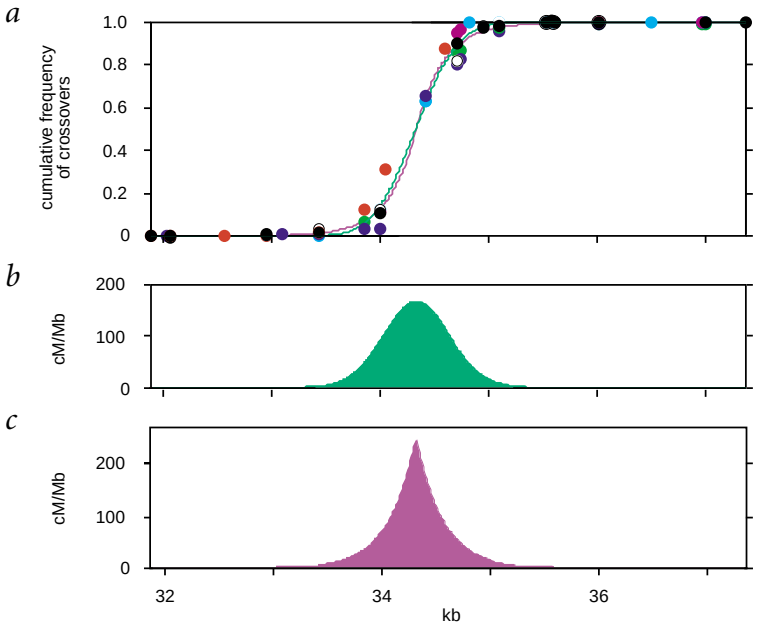

Fig. 3 Distribution of sperm crossovers within hot-spot *DNA3*. Sperm DNA samples from six men informative for SNP markers flanking this hot spot were screened for crossover molecules, yielding 70–126 crossovers per individual tested. ***a***, Cumulative frequency of crossovers across the test interval for each man (colored circles). One man (black circle) was also analyzed for reciprocal crossovers (open circle; green to red crossovers rather than red to green, Fig. 2*a*). We used combined distributions to determine the least-squares best-fit cumulative distribution assuming that crossover breakpoints either are normally distributed across the hot spot (green curve) or show a bidirectional exponential decay from a fixed point (purple curve). ***b***, Underlying morphology and intensity of the hot spot under the normal distribution model. We determined peak activity from the mean rate of crossover in the six men tested. The distribution has a standard deviation of 300 bp. ***c***, Hot-spot morphology under the exponential decay model. This model would, for example, be consistent with a recombination initiation complex binding to the center of the hot spot but then moving either 5′ or 3′ with a uniform probability of 0.004, per base pair traversed, of halting and effecting a crossover at that point.

recovered from sperm but not blood DNA (Fig. 2*b*), indicating that they are genuine products of meiotic recombination; additional criteria of authenticity are described elsewhere[12]. Location of the site of crossover in each recombinant molecule, using internal heterozygous SNPs, shows that almost all sperm crossovers (1,399 of 1,404) are simple and map to a single interval between adjacent heterozygous SNPs, as seen previously at the *TAP2* hot spot[12] and the minisatellite MS32-associated hot spot[14]. Only five crossovers are more complex, with switching between haplotypes near the site of crossover[12], suggesting that male crossover is rarely associated with patchy conversion of nearby SNPs[15].

Sperm crossovers are highly clustered in each test interval and define five narrow (1–2 kb) recombination hot spots, three around *HLA-DNA* (named *DNA*1–3) and two in the *HLA-DMB* region (*DMB*1,2; Fig. 2*c*,*d*). These hot spots all locate to intervals of LD breakdown, but vary considerably in peak intensity from only 0.4 cM/Mb for hot spot *DNA*1 to 140 cM/Mb for *DNA*3. Reciprocal recombinants (green-red rather than red-green crossovers; Fig. 2*a*) show the same rate and distribution for four of the hot spots. This is consistent with reciprocal crossover (or possibly with simple and continuous conversion tracts arising with equal likelihood in each orientation) rather than more complex conversion events (Fig. 3*a* for *DNA*3)[12]. Apparent deviations from fully reciprocal crossover at *DNA*2 will be discussed elsewhere. The cumulative frequencies of crossovers across each hot spot, including the *TAP2* and minisatellite MS32 hot spots (Fig. 3*a* for *DNA*3), show an apparently symmetric distribution within each hot spot, indistinguishable in the 2–6 men typed per hot spot, which in each case approximated to a normal distribution (Fig. 3*b*). Each hot spot can

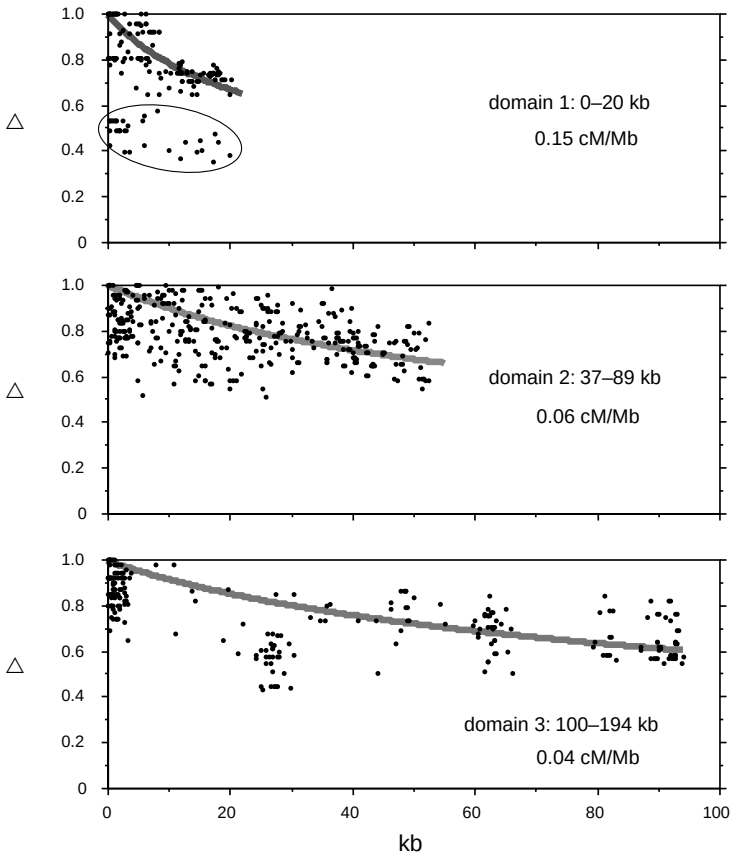

Fig. 4 Decay of association with distance within domains of linkage disequilibrium. We analyzed the three extended domains of LD shown in Fig. 1*b* using only SNPs with minor allele frequencies of at least 0.25 (23, 28 and 23 SNPs in domains 1, 2 and 3, respectively). The standardized LD coefficient Δ, a measure of absolute association such that |Δ|=1 only when there are just two haplotypes per pair of markers[28], was estimated for each pair of SNPs in each domain and plotted against physical distance between the markers. Lines show the relationship between |Δ| and physical distance expected if crossovers occur at random within each domain. We calculated these lines from $\Delta^2=1/(1+4N_e rd)$, where N_e is the effective population size, assumed to be 10,000 for humans[33], r is the recombination frequency per Mb and d is the inter-marker distance in Mb[34]. The corresponding recombination activity in cM/Mb is indicated for each domain; if this region of the MHC is unusually ancient, then N_e would be larger, resulting in a corresponding reduction in estimates of crossover activity. Points in domain 1 showing an unusually low level of LD (ellipse) resulted from just two SNPs in almost absolute association with each other and with relatively low minor-allele frequencies (0.25, 0.28).

therefore be defined by center location, normal distribution width and peak recombinational activity in cM/Mb. Other distributions are possible, however, such as bidirectional exponential decay from the center of the hot spot (Fig. 3*c*).

Sperm analysis around the low-activity recombination hot spot *DNA*1 (Fig. 2*c*) shows additional rare presumptive crossovers outside the hot spot that seem to occur at random along the test interval at a frequency of about 0.04 cM/Mb, suggesting that crossovers are not entirely restricted to hot spots. Similarly, each extended domain of LD (Fig. 1*b*) shows a gradual decrease of LD with distance between markers (Fig. 4) to an extent consistent, at recombination-drift equilibrium, with a uniform sex-averaged crossover activity of 0.04–0.15 cM/Mb. Although these estimates may be distorted by selection and the existence of ancient MHC haplotypes, they are nevertheless similar to the estimate from non–hot spot sperm crossovers.

We combined these crude estimates of inter–hot spot crossover frequency with hot-spot data to gain an overview of male meiotic recombination activity across the entire 216-kb MHC region (Fig. 5). The total sperm recombination frequency of 0.18 cM is similar to that seen in pedigrees (about 0.2 cM, based on 3–4 paternal crossovers[9]). This suggests that many sperm exchanges result from true meiotic crossovers rather than continuous conversion tracts extending out from a hot spot and encompassing SNP sites used to select recombinants. However, if such tracts occur, they would have to extend uninterrupted from one hot spot to the next to generate the observed strong correlation between hot spots and LD breakdown. Sperm crossovers occur across the region at 0.9 cM/Mb, very close to the genome average rate in male meiosis of 0.89 cM/Mb[16]. However, the great majority (about 94%) of crossovers lie within hot spots, with 72% in hot spot *DNA*3. All hot spots share a very similar width also seen at the MS32 hot spot that appears to drive minisatellite instability[14], although none are associated with tandem repeat DNA sequences, indicating that hot spots only occasionally create GC-rich minisatellites[14]. Similar clustering into 0.5–5-kb long hot spots has been observed in the mouse MHC, although in most cases limited numbers of pedigree crossovers have prevented a more precise definition of hot-spot morphology[7,17,18]. In yeast, meiotic crossovers also cluster into hot spots that tend to associate with GC-rich chromosomal domains[19] and locate to transcriptional promoters[19,20]; these associations are not evident in the human MHC, where only the weakest hot spot (*DNA*1) is located at a promoter and the most intense (*DNA*3) is centered in an intergenic *Alu* repeat.

hot spot	*DNA*1	*DNA*2	*DNA*3	*DMB*1	*DMB*2	*TAP2*	MS32
no. men tested	2	3	6	2	2	2	3
no. crossovers	69	237	661	36	358	141	250
center	3,238,030	3,234,020	3,226,050	3,164,030	3,160,780	3,065,340	–
95% width, kb	1.9	1.3	1.2	1.8	1.2	1.0	1.5
mcM, mean	0.5	3.7	130	3.1	28	5.8	43
mcM, range	0.5, 0.5	0.4–8.6	90–160	2.6, 3.6	24, 32	4.4, 7.1	19–89
location	promoter	intergenic, *Alu*	intergenic, *Alu*	intragenic, intron/exon?	intergenic	intron	RTLV-LTR

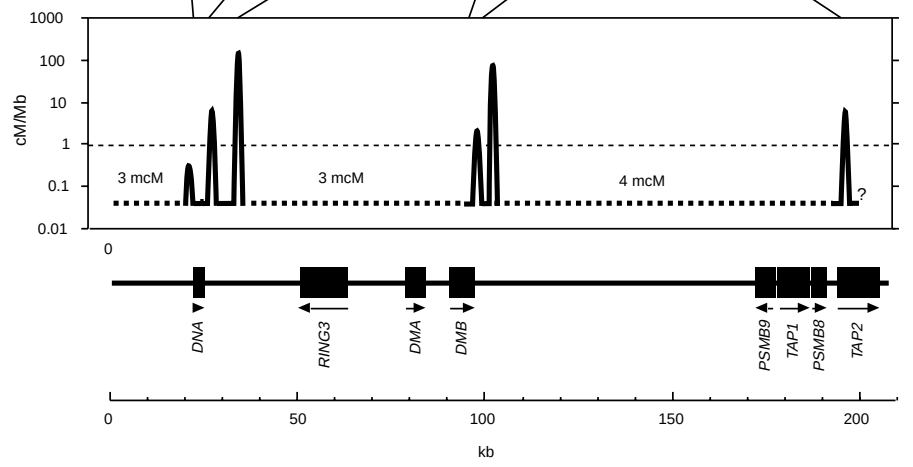

Fig. 5 Sperm crossover activity in the class II region of the MHC. The number of men tested and the total number of sperm crossovers mapped are given for each hot spot, together with approximate hot-spot center coordinate in the consensus sequence of the human MHC[10]. The width of each hot spot, within which 95% of crossovers occur, was determined by normal-distribution fitting (Fig. 3). The mean male linkage map distance contributed by each hot spot, plus range seen in the different men tested, was determined from the observed hot spot crossover frequency per sperm and is given in millicentimorgans (mcM, cM×10^{-3}); only the hot spot *DNA* 2 shows significant variation in activity between tested men. Inter–hot spot distances were estimated from data in Fig. 4. The background recombination rate of 0.04 cM/Mb is very approximate and should be treated with caution. The mean rate of male meiotic recombination in the human genome (0.89 cM/Mb)[16] is shown as a thin dashed line. *TAP2* and minisatellite MS32 estimates were from data published elsewhere[12,14].

The similar widths of the human hot spots suggest a common process operating at each hot spot. If these hot spots result from open domains in meiotic chromatin, as seen in yeast[21–23], this implies the existence of consistently narrow (1–2 kb) domains that allow access of the recombinational machinery. Alternatively, the fit of crossover data to an exponential-decay model (Fig. 3*c*) is compatible with loading of a recombination initiation complex at the center of a hot spot followed by stochastic drift away from the center. It is equally possible that crossover initiation at the hot-spot center could be followed by very limited branch migration of a Holliday junction prior to crossover resolution, leading to spreading of crossover sites to create the observed narrow distributions. However, as noted before for the *TAP2* and MS32 hot spots[12], there are no obvious primary recombination-promoting sequences shared by these hot-spot centers (data not shown).

The MHC hot spots are not randomly distributed but fall into clusters 60–90 kb apart, with 1–7 kb of DNA separating each hot spot within a cluster (Fig. 5). The abrupt decay of LD about 5 kb 3′ of the *TAP2* hot spot (Fig. 1*b*) suggests that this hot spot is also a member of a cluster, though a complex rearrangement polymorphism in this region (data not shown) has prevented further analysis of LD patterns and sperm crossovers. The reasons for clustering remain obscure. Crossovers may be initiated only in the most active hot spot in a cluster, with most being resolved in the same hot spot after very limited branch migration. Some Holliday junctions may, however, migrate out of the hot spot but then stall distally at preferred locations, creating secondary hot spots. But this does not account for the similar widths of all hot spots in a cluster, and it remains likely that each hot spot represents a separate site of crossover initiation.

This study indicates that, despite previous concerns[24], LD patterns can indeed locate crossover hot spots with considerable accuracy, at least within the MHC. The strong correlation between LD patterns and sperm hot-spot locations further suggests that the same hot spots function in female meiosis, though probably with different intensities[12]. The location of these MHC hot spots may have been influenced by selection and may not reflect crossover distribution elsewhere in the genome. However, the lengths of the extended domains of LD (60 and 90 kb; Fig. 1*b*) are consistent with estimates of north-European LD block length elsewhere in the human genome[5]; a strikingly similar pattern of 10–100-kb long LD blocks in a 500-kb interval on chromosome 5 is described in this issue[25]. It therefore seems that the LD blocks and associated hot spots seen in the MHC may occur more generally in the human genome. These LD block lengths are also similar to the average spacing of 50 kb between yeast hot spots[19], indicating the possibility of highly conserved mechanisms that control hot-spot spacing. The lengths of the MHC LD domains further imply that the human genome might be a mosaic of about 40,000 recombinationally suppressed segments of DNA. If correct, then whole genome association scanning for complex disease loci could eventually be approached more rationally by screening perhaps 2–5 SNPs per LD domain (80,000–200,000 in total), although more SNPs would be required to detect association with aetiological variants located within hot-spot clusters. The identification of SNPs that capture haplotype block information is discussed further elsewhere in this issue[26].

Methods

SNP discovery and genotyping. We used PCR primers designed from the current consensus MHC sequence[10] (28 Oct 1999 version; http://www.sanger.ac.uk/HGP/Chr6/MHC_991021.fasta) to amplify appropriate 1–5-kb DNA segments from 8 unrelated north-European UK semen donors. We resequenced these amplicons using BigDye Terminators (ABI) on an ABI 377 Automated Sequencer and identified SNPs using ABI AutoAssembler software. SNPs near *TAP2* have been described[12]. We identified additional SNPs around *RING3* in dbSNP (build ID 93; http://www.ncbi.nlm.nih.gov/SNP/). We genotyped SNPs in a panel of 50 unrelated north-European British semen donors by allele-specific oligonucleotide (ASO) hybridization to dot blots of appropriate long-PCR products[12]. Only 11 of the 14,580 genotypes were ambiguous. For details of SNPs, ASOs, genotypes and allele frequencies, see http://www.le.ac.uk/genetics/ajj/HLA.

Linkage disequilibrium analyses. We analyzed and plotted LD data using software written in True BASIC 4.1 by A.J.J. We estimated maximum-likelihood haplotype frequencies for each pair of SNP markers from the unphased diploid genotype data from the 50 men, assuming that all SNPs and haplotypes were in Hardy-Weinberg equilibrium (after Bonferroni correction, none of the SNPs showed significant departure from Hardy-Weinberg equilibrium; data not shown). We used these haplotype frequencies to estimate the level of LD between each pair of SNPs, using both the |D'| measure of complete association[27] and the |Δ| measure of absolute association[28]. We used the observed allele frequencies at each SNP to predict the haplotype frequencies expected if pairs of SNPs were in linkage equilibrium (LE). We then estimated the likelihood ratio (LR) in favor of significant LD as $\max(P[\text{data}])/P[\text{data}|\text{LE}]$, where $\max(P[\text{data}])$ is the probability of obtaining the observed unphased genotype data at the maximum-likelihood haplotype frequencies and $P[\text{data}|\text{LE}]$ is the probability of obtaining the data if haplotypes were at linkage-equilibrium frequencies.

We investigated haplotypes showing a highly significant (LR>10,000) sharing of low-frequency SNP alleles (Fig. 1*c*) by using additional SNPs present in the homozygous state to reconstruct partial but unambiguous haplotypes across the entire 216-kb region of the MHC in all 50 men typed. In all cases, these partial haplotypes show a region of haplotypic identity, common to all men sharing low-frequency alleles in strong association, that encompasses all of these alleles.

Sperm crossover analyses. We designed allele-specific PCR primers 16–21 nt long with at least 50% GC content for appropriate heterozygous SNP sites. Where necessary, the efficiency of primers with a lower GC content was improved either by adding a synthetic CCCC 5′ extension to the primer or by modifying the allele-specific primer with a synthetic 20-nt 5′ extension (55–60% GC) and using both the allele-specific primer and a primer identical to the 5′ extension to drive amplification[29]. The optimal annealing temperature of each primer was determined by amplifying genomic DNA at various annealing temperatures from individuals homozygous for either the appropriate allele or the alternative allele. We selected primers for crossover detection on the basis of high efficiency and excellent allele specificity; less than half of primers tested conformed to these criteria.

We detected recombinant molecules in sperm DNA using modifications of the procedures developed to analyze crossovers at the *TAP2* hot spot[12]. Sperm DNA prepared under conditions that minimized the risk of contamination[30] was first digested with a restriction endonuclease that cleaved outside the test interval, to render the DNA fully soluble and amenable to quantitative dilution. We used allele-specific PCR to establish linkage phase between the 5′ and 3′ SNP sites chosen to select crossovers. To detect recombinants, we amplified multiple aliquots of digested sperm DNA, each containing 0.2–2 amplifiable crossover molecules (3.6–360 ng genomic DNA, depending on crossover frequency), by long PCR[31] using two allele-specific primers in repulsion phase directed to SNP sites 6–8 kb apart. Typical touchdown PCR conditions were 1 cycle at 96 °C for 1 min, followed by 24 cycles of 96 °C for 20 s, 53–66 °C (depending on the primers) for 30 s and 61–66 °C for 6–8 min (depending on the length of the amplicon). The low extension temperature allowed efficient amplification, particularly across AT-rich DNA[32], and reduced the incidence of inter-haplotype-jumping PCR artifacts, which occasionally occur with extension at 70 °C (data not shown)[12]. We digested the primary PCR products with S1 nuclease to remove single-stranded DNA plus PCR artifacts[14] and re-amplified with nested repulsion-phase allele-specific primers as above. We analyzed aliquots of the secondary PCR products by agarose gel electrophoresis and detected DNA by staining with ethidium bromide. We re-amplified positive secondary PCR reactions using PCR primers located just within the secondary SNP sites and determined the status of internal SNP sites by ASO typing. We carried out Poisson estimation of the number of amplifiable input molecules and correction for PCRs containing more than one crossover molecule as described previously[12]; the single-molecule long-PCR efficiency was consistently estimated at 50% (1 amplifiable molecule of each haplotype per 12 pg genomic DNA), as seen previously[12]. For primer and PCR cycling details, see http://www.le.ac.uk/genetics/ajj/HLA.

Acknowledgments

We thank J. Blower and numerous volunteers for supplying semen and blood samples, K. Lilley for assistance with automated sequencing and oligonucleotide synthesis, R. Dalgleish for advice on PCR, J. Stead for web site construction and other colleagues for helpful discussions. This work was supported by grants to L.K. from the Instrumentarium Science Foundation and the Finnish Cultural Foundation and to A.J.J. from the Medical Research Council, Wellcome Trust and Royal Society.

Received 18 May; accepted 24 August 2001.

1. Collins, A., Lonjou, C. & Morton, N.E. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl Acad. Sci. USA* **96**, 15173–15177 (1999).
2. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).

3. Jorde, L.B. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* **10**, 1435–1444 (2000).
4. Ott, J. Predicting the range of linkage disequilibrium. *Proc. Natl Acad. Sci. USA* **97**, 2–3 (2000).
5. Reich, D.E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**,199–204 (2001).
6. Lichten, M. & Goldman, A.S.H. Meiotic recombination hotspots. *Annu. Rev. Genet.* **29**, 423–444 (1995).
7. Shiroishi, T., Koide, T., Yoshino, M., Sagai, T. & Morikawi, K. Hotspots of homologous recombination in mouse meiosis. *Adv. Biophys.* **31**, 119–132 (1995).
8. Petes, T.D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2**, 360–369 (2001).
9. Cullen, M. *et al.* Characterization of recombination in the HLA class II region. *Am. J. Hum. Genet.* **60**, 397–407 (1997).
10. Beck, S. *et al.* Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**, 921–923 (1999).
11. Cullen, M., Erlich, H., Klitz, W. & Carrington, M. Molecular mapping of a recombination hotspot located in the second intron of the human *TAP2* locus. *Am. J. Hum. Genet.* **56**, 1350–1358 (1995).
12. Jeffreys, A.J., Ritchie, A. & Neumann, R. High-resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hotspot. *Hum. Mol. Genet.* **9**, 725–733 (2000).
13. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
14. Jeffreys, A.J., Murray, J. & Neumann, R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2**, 267–273 (1998).
15. Borts, R.H. & Haber, J.E. Meiotic recombination in yeast—alteration by multiple heterozygosities. *Science* **237**, 1459–1465 (1987).
16. Gyapay, G. *et al.* The 1993-94 Généthon human genetic linkage map. *Nature Genet.* **7**, 246–339 (1994).
17. Steinmetz, M., Stephan, D. & Lindahl, K.F. Gene organization and recombinational hotspots in the murine major histocompatibility complex. *Cell* **44**, 895–904 (1986).
18. Snoek, M., Teuscher, C. & van Vugt, H. Molecular analysis of the major MHC recombinational hot spot located within the *G7c* gene of the murine class III region that is involved in disease susceptibility. *J. Immunol.* **160**, 266–272 (1998).
19. Gerton, J.L. *et al.* Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **97**, 11383–11390 (2000).
20. Baudat, F. & Nicolas, A. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc. Natl Acad. Sci. USA* **94**, 5213–5218 (1997).
21. Ohta, K., Shibata, T. & Nicolas, A. Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J.* **13**, 5754–5763 (1994).
22. Wu, T.C. & Lichten, M. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* **263**, 515–518 (1994).
23. Fan, Q.Q. & Petes, T.D. Relationship between nuclease-hypersensitive sites and meiotic recombination hot spot activity at the *HIS4* locus of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 2037–2043 (1996).
24. Hedrick, P.W. Inference of recombinational hotspots using gametic disequilibrium values. *Heredity* **60**, 435–438 (1988).
25. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. High-resolution haplotype structure in the human genome. *Nature Genet.* **29**, 229–232 (2001).
26. Johnson, G.C.L. *et al*. Haplotype tagging for the identification of common disease genes. *Nature Genet.* **29**,233–237 (2001).
27. Lewontin, R.C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1984).
28. Hill, W.G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231 (1968).
29. Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. & Monckton, D.G. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**, 204–209 (1991).
30. Jeffreys, A.J. *et al.* Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* **6**, 136–145 (1994).
31. Cheng, S., Fockler, C., Barnes, W.M. & Higuchi, R. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl Acad. Sci. USA* **91**, 5695–5699 (1994).
32. Su, X., Wu, Y., Sifri, C.D. & Wellems, T.E. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res.* **24**, 1574–1575 (1996).
33. Morton, N.E. *Outline of Genetic Epidemiology* (Karger, Basel, 1982).
34. Sved, J.A. Linkage disequilibrium and homozygosity of chromosomal segments in finite populations. *Theor. Popul. Biol.* **2**, 125–141 (1971).