Classification: Biological Sciences (Genetics)

Title: Sperm crossover activity in regions of the human genome showing extreme breakdown of marker association

Authors: Adam J. Webb, Ingrid L. Berg and Alec J. Jeffreys

Author affiliation: Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

Corresponding author: Alec J. Jeffreys
Department of Genetics, University of Leicester, Leicester LE1 7RH, UK
tel. +44 116 2523435
fax. +44 116 2523378
email ajj@le.ac.uk

Manuscript information

No. of text pages: 25

No. of figures: 5

No. of tables: 0

No. of words in the abstract: 221

Total no. of characters in the manuscript: 36,779

Abbreviations footnote: allele-specific oligonucleotide (ASO), linkage disequilibrium (LD), linkage disequilibrium unit (LDU), recombination frequency (RF).
Abstract

Population diversity data have recently provided profound, albeit inferential, insights into meiotic recombination across the human genome, revealing a landscape dominated by thousands of crossover hotspots. However, very few of these putative hotspots have been directly analysed for crossover activity. We now describe a search for very active hotspots, using extreme breakdown of marker association as a guide for high-resolution sperm crossover analysis. This strategy has led to the isolation of the most active crossover hotspots yet described. Their morphology, sequence attributes and crossover processes are very similar to those seen at less active hotspots, but their activity in sperm is poorly predicted from population diversity information. Several of these new hotspots showed evidence for biased gene conversion accompanying crossover, in some cases associated with variation between men in crossover activity and with two hotspots showing complete presence/absence polymorphism in different men. Hotspot polymorphism is very common at less active hotspots but curiously was not seen at any of the most active hotspots. This contrasts with the prediction that extreme hotspots should be the most vulnerable to attenuation by meiotic drive in favour of mutations that suppress recombination, and should therefore show rapid rate evolution and thus variation in activity between men. Finally, these very intense hotspots provide a valuable new resource for dissecting meiotic recombination processes and pathways in humans.
Introduction

The recombinational exchange of DNA between homologous chromosomes at meiosis is vital to ensure correct chromosome segregation and also plays a major role in increasing haplotype diversity within populations. In humans, the combination of low average crossover frequency (~1% recombination frequency (RF) per Mb DNA) and small numbers of informative meioses in pedigree studies has limited the resolution of current linkage maps to the megabase level (1, 2). Much higher resolution profiles of recombination can instead be obtained indirectly through examining patterns of marker association (linkage disequilibrium, LD) – established through population dynamic processes and eroded by recombination – or directly through labour-intensive screening of millions of sperm for recombinant DNA molecules within short DNA intervals (typically < 10 kb).

Population LD (3, 4) and sperm DNA (5-14) analyses have firmly established that most crossover events in humans cluster into narrow hotspots spaced, on average, 50 kb apart. Recently, the International HapMap Project (15, 16) has mapped the LD landscape genome-wide at the kilobase level. These data allowed inference of the global recombination landscape at high resolution (4) using coalescent analyses whereby observed haplotypes are explained through in silico reconstruction using variable historical recombination rates. These analyses have identified ~33,000 putative crossover hotspots (LD hotspots) throughout the genome (4, 16, 17), and have provided insights into hotspot distribution and historical crossover activity, as well as identifying DNA sequence motifs associated with hotspots (16, 18).
In contrast, few human recombination hotspots have been directly characterised in sperm, and it is still unclear whether LD landscapes can accurately predict and locate genuine hotspots or correctly estimate their historical activity. To date, sperm surveys have only covered a total of 0.6 Mb of human DNA, identifying seven hotspots in a 216 kb region of the major histocompatibility complex (6, 10, 11), eight in a 206 kb region on chromosome 1 (7-9, 13), one in the β-globin gene cluster (5) and one in the SHOX gene located in the Xp/Yp pseudoautosomal region PAR1 (12). A near-contiguous 103 kb segment of chromosome 21 has also been screened for crossovers in sperm (14). These surveys have shown a good, if not perfect, concordance between the location of LD hotspots and sperm hotspots. They have revealed additional phenomena that could not have been detected from population data, including variation in hotspot activity between men (7-9, 14) as well as complete on/off polymorphism despite no changes in local DNA sequence (13). Meiotic drive in favour of a crossover-suppressing variant within the hotspot has been detected at two loci (8, 19), suggesting a mechanism for hotspot extinction (8, 20). Conversely, active sperm hotspots have been observed within regions of strong LD, consistent with these hotspots being young (9). These findings suggest that crossover hotspots might be transient features of the genome, turning over rapidly in evolutionary time; this is consistent with the markedly divergent LD landscapes of humans and chimpanzees (21-23).

The autosomal crossover hotspots analysed to date show sperm RFs ranging from 0.0005% (6) to 0.14% (5). These hotspots were identified in regions that were not unusually active in meiotic recombination as judged from linkage maps. It is therefore likely that the most active hotspots have yet to be characterised. Indeed, an intense
hotspot with 1.1% RF in sperm has been found in mice despite extremely limited surveys of the mouse genome (24). We now expand the current repertoire of human crossover hotspots by targeting sperm crossover assays to short intervals showing the most extreme LD breakdown in HapMap genotypes. This new panel of ‘super-hotspots’ will provide a valuable resource for population geneticists to explore the relationship between recombination and DNA diversity. It will also aid further studies into recombination through the analysis of frequencies and distributions of crossovers and gene conversions, helping to elucidate factors contributing to the regulation and evolutionary turnover of human crossover hotspots.

Results and Discussion

Selecting strong LD hotspots from genotype data. Strong recombination hotspots should create intervals of very substantial or complete breakdown of marker association. Such intervals can be identified by LD mapping (25), which provides a profile of the rate of LD breakdown along a chromosome in linkage disequilibrium units (LDUs). LD maps show a good correlation at the megabase level with linkage maps (26), and also with historical crossover rates estimated by coalescent analysis of genotype data (17). We therefore constructed LD maps of all the human autosomes using Phase II HapMap genotype data (16) on 2.4 million single nucleotide polymorphisms (SNPs) from 60 individuals from a Utah population with northern and western European ancestry (CEU). We scanned these maps for intervals of 5 kb containing a strong (> 3) LDU step and good SNP density (> 5 genotyped SNPs), and compared these intervals with coalescent analysis data available from HapMap
Fifteen strong LD hotspots (A-Q) each characterised by a major LDU step and substantial historical activity were selected for crossover analysis in sperm. We also analysed the most active LD hotspot reported by HapMap (16) (historical RF estimated at 1.2% by coalescent analysis). This LD hotspot is however unusually broad (25 kb) (Fig. 1B), and the LD map appeared to resolve it into a moderately strong LD hotspot 2.5 kb wide (R, chosen for further analysis) followed apparently by a succession of other weaker LD hotspots.

Each of the 16 targets was re-genotyped in a panel of 94 north European semen donors using HapMap SNPs plus additional markers in dbSNP. The LDU step at each target was confirmed, although 12 showed a drop in LDU step size, in some cases substantial (Fig. 1C). This drop is significant (paired t-test, \( P = 0.005 \)) and is due in part to increased marker density closing “holes” (intervals of free association between adjacent markers) in the HapMap LD map (26) and reducing step size. Ascertainment bias probably also contributes to this reduction – extreme LDU steps will tend to arise at those hotspots where haplotypes sampled in CEU HapMap individuals happen by chance to over-inflate LDU step size; genotyping additional individuals will remove this bias and thus reduce step size. In contrast to LDU, historical crossover rate estimates from coalescent analysis were not significantly different between CEU individuals and north European semen donors (\( P > 0.05 \)) (Fig. 1C) but showed little variation between hotspots, with most showing a historical RF of ~ 0.2%.

**Sperm crossover profiles at strong LD hotspots.** Sperm DNA was assayed for crossover molecules across each of the 16 LD hotspots, with each hotspot analysed in three different informative men to test for crossover frequency variation (6).
Reciprocal crossovers (aB and Ab crossovers in a man heterozygous for haplotypes AB and ab) were separately analysed in each man to verify crossover frequency estimates, which proved to be very reproducible (see Materials and Methods), and to test for reciprocal crossover asymmetry (19). Excluding a few instances of unusually low crossover frequencies, ~230 crossovers were typically analysed per hotspot per man. This survey yielded in total 11,200 crossovers, all of which were mapped to locate crossover breakpoints.

Each test interval showed a highly non-random distribution of crossovers (Fig. 2A). In most cases (13 out of 16), these distributions were consistent with a single crossover hotspot. Targets C, G and J each showed a broader distribution that, in men with sufficient marker density, could be seen to resolve into two hotspots in each case separated by 2 kb. This survey therefore yielded a total of 19 sperm-verified crossover hotspots.

Some strong LD hotspots are very active in sperm crossover. Despite the similar estimates of LDU step size and historical activity at these hotspots, they showed a wide range of sperm crossover frequencies, varying from 0.015% to 1.0% (median 0.13%) (Fig. 2B). These activities are considerably higher than at previously characterised human autosomal hotspots (0.0005-0.14%, median 0.011%) (5-11, 13), with eight of the hotspots being the most active yet described. This elevated activity at the present hotspots is significant (Mann-Whitney test, $P = 0.0005$), indicating that our strategy of identifying very active hotspots from intervals of extreme LD breakdown was successful. While these hotspots are very active, their flanking DNA is as suppressed for recombination as at previously reported hotspots (only 55
crossovers mapped outside hotspots, giving a mean RF of 0.16% / Mb, consistent with previous estimates (6, 8, 13)).

**LD profiles are an imperfect surrogate for crossover profiles.** Sperm crossover profiles can, with sufficient SNP marker density, locate hotspot centres with considerable accuracy, within ±50 bp or less (6). Comparison of sperm profiles with LD maps and historical rate profiles (Fig. 2A) showed an erratic relationship between DNA diversity and crossover. In some cases (e.g. hotspot P), sperm and diversity profiles were very closely congruent. In other cases (e.g. hotspot D), there were significant discrepancies with the hotspot centre mis-located by up to 1 kb. Coalescent analysis using LDhat (17) did not have the resolution to resolve any of the double hotspots, while LD mapping only succeeded in detecting a doublet at hotspot J. Thus diversity analyses would interpret at least two of these regions (C, G) as a single broad hotspot centred within a region between sperm hotspots that is in fact recombinationally suppressed. Finally, LD mapping identified a second putative hotspot at 12 kb in target N (LDU step = 0.74). However, only three crossovers were seen in this region in 240,000 molecules screened, indicating that this hotspot, if it exists in the men tested, is extremely weak (RF 0.001%).

The quantitative relationship between sperm crossover frequency and both LDU step size and historical rates estimated by coalescent analysis was far from perfect (Fig. 3). Both population estimates appeared to plateau above 0.2% RF in sperm, presumably due to free association arising through saturation of LD breakdown. Below 0.2%, the overall trend for sperm crossover frequencies followed that predicted from historical rates. There was however substantial scatter, with individual hotspots showing
anywhere from 30-fold less sperm activity, compared with historical rates, to a 5-fold excess. It is unclear whether these discrepancies reflect imperfections in historical rate estimation by coalescent analysis, or instead arise through individual- and sex-specific differences in crossover activity at specific hotspots (10) together with rapid evolutionary turnover of hotspots (8, 9, 13, 19, 21-23). In any event, neither LD mapping nor historical rate estimation is a good predictor of current sperm crossover activity.

**Properties of sperm crossover hotspots.** All hotspots showed fairly uniform widths of 1.2-1.9 kb (mean 1.45 kb, Fig. 2C), very similar to the widths of previously characterised but less active hotspots (5-7, 9-13). Thirteen were intergenic while the remaining six were in introns, with hotspots L and N located in the same gene (*CCBE1*) just 84 kb apart. Variants of the motif CCTCCCTNNCCAC, reported as being associated with LD hotspots (16), were found within 70 bp of the centre of hotspots C2, D, F, K, L, M and P (37% of hotspots), in two cases (D, K) mapping exactly at the centre defined by sperm typing. While each of these associations was significant (*P < 0.05*), such motifs were not predictive of hotspot location, with examples of equally, if not better, matched motifs being present in the recombinationally inert regions flanking all of these hotspots. In summary, the properties of these very active hotspots in terms of width, genomic location and sequence motifs are very similar to those of previously characterised but less active hotspots (5-13).

**Complex crossovers.** As with previously characterised sperm crossovers (5-13), the great majority of exchanges were simple, with each crossover mapping to a single
interval between heterozygous SNPs. However, 37 of the 11,200 mapped crossovers (0.33%) were more complex, showing haplotype switching near the site of exchange presumably arising through patchy repair of heteroduplex DNA generated during recombination (Fig. 4). Most of these events were located within the hotspot, but 22% of them (8 out of 37) extended beyond the hotspot, reminiscent of the complex crossovers seen in mismatch-repair deficient mice (28). Given that different hotspots were analysed in different men, it was not possible to test whether these complex events were preferentially restricted to specific men, nor whether they might associate with impaired mismatch repair functions. These complex events were however not randomly distributed across hotspots ($P < 0.001$), with the most active hotspots E, F and R showing the highest proportion of complex crossovers (1.1%, compared with 0.15% at the remaining hotspots). The reason for this correlation is unclear, but may in part be due to the higher mean number of markers within these very active hotspots (3.6 per hotspot per man vs. 2.2 in the remaining hotspots) allowing short conversion patches to be more readily detected.

**Polymorphism of hotspot activity.** Four hotspots (A, B, C1, J1) showed major (> 10-fold) variation in crossover frequency between the men tested (Fig. 2B). Three other hotspots (C2, G1, G2) showed minor (2-4 fold) inter-individual variation in activity. These variations in RF are significant ($P < 0.001$), but we cannot exclude the possibility that such minor variation might arise from subtle shifts in DNA quality and/or PCR efficiency affecting crossover recovery. Hotspots B and J1 showed the most extreme variation, with one man in each case showing no crossovers mapping to the hotspot region, indicating substantial or complete suppression of crossover activity (suppression > 90 fold and > 30 fold at B and J1 respectively, $P = 0.95$). Only
one example of presence/absence polymorphism of a hotspot has been previously reported (13).

Curiously, variation in crossover activity was restricted to the less active hotspots showing < 0.1% RF (Fig. 2B). This restriction is significant (Mann-Whitney test, $P = 0.0047$ if marginally polymorphic hotspots are scored as negative or $P = 0.018$ if positive). It therefore appears that hotspot polymorphism associates preferentially with weaker hotspots. Data on the current hotspots plus previously characterised hotspots (5-13) indicate that these polymorphisms are common, with about 55% of weaker hotspots (12 out of 22) showing evidence of crossover frequency variation between men.

It is puzzling that strong hotspots show less variation in activity between individuals. Intense crossover activity should accelerate the spread and fixation of recombination-suppressing mutations within the hotspot through preferential over-transmission of such variants to recombinant progeny (8, 19). This predicts that the most intense hotspots should be the most ephemeral in human populations and thus the most likely to show activity variation between individuals (20). These hotspots showed no unusual features and we therefore have no explanation for this apparent paradox, but any model of how hotspots arise and persist in the face of meiotic drive in favour of recombination suppressors will need to take this phenomenon into account.

**Activity polymorphism and biased gene conversion.** If the two haplotypes in a man differ in their frequency of recombination initiation, then disparities between reciprocal crossover distributions will arise, with systematic over-transmission of
markers from the suppressed haplotype into crossover progeny (19). Two hotspots (B, J1) each showed very strong disparity in at least one man, with markers close to the hotspot centre showing strongly distorted transmission ratios (83:17) into crossover progeny (Fig. 5A). Similar levels of segregation distortion have been seen at other hotspots (8, 19). Hotspots B and J1 were also the two hotspots that showed presence/absence polymorphism between men (Fig. 2B). This association is significant ($P = 0.006$) and provides further evidence that strong transmission distortion is the result of men with the hotspot carrying both active and suppressed haplotypes (with inactive men homozygous for suppressed haplotypes), and is not solely due to strong mismatch repair biases acting on heteroduplex DNA generated during recombination (19).

A second, more subtle, type of transmission distortion was also observed. Twenty-three crossover assays at hotspots excluding B and J1 had markers sufficiently close to the hotspot centre (< 100 bp) to allow asymmetry to be tested. Of these assays, five showed significant asymmetry ($P < 0.05$ in each case; $P = 0.0002$ for the entire data set) but with very minor biases in transmission ratios (56:44) (Fig. 5B). These biases were seen in one or two men at hotspots D, E, P and Q and preferentially affected markers at the hotspot centre. Such biases could arise through subtle disparities between haplotypes in recombination initiation rates – a difference of just 20% would be sufficient to create these biases but would be impossible to detect on the basis of the very minor (1.2-fold) predicted variation in sperm crossover frequency between men. Alternatively they could arise through subtle mismatch repair biases leading to biased gene conversion in crossovers. While the effect is subtle, the population consequences could be very significant, particularly at the most
active hotspots; if meiotic drive affects a specific variant in the hotspot, then the population level of transmission distortion (up to 50.015:49.985 for the very active hotspot E) is strong enough to essentially guarantee eventual fixation of the over-transmitted variant (8, 19). Even more subtle biased gene conversion might occur during crossing over, but much larger surveys of crossover molecules would be needed to detect it.

**Summary.** This survey has more than doubled the number of human meiotic recombination hotspots characterised by direct sperm analysis. Features previously established from studies of less active hotspots, such as narrow (1-2 kb) width, symmetric and quasi-normal distribution of crossover exchange points, generally simple exchange events and hotspot clustering, fully apply to these highly active hotspots. Comparisons of crossover profiles with LD patterns show that population diversity data are excellent predictors of hotspots and can generally locate them with reasonable accuracy, but are poor at predicting hotspot intensity. This in turn implies that the present survey has not detected all of the most intense hotspots in the human genome. It is likely that a full survey will only be possible through information gained, not from population level DNA diversity data, but from very high resolution genome-wide mapping of crossovers detected by extensive family analyses. However, current surveys lack the numbers of meioses and marker density required to reliably locate even the most active hotspots at the kilobase level (2).

The new hotspots in this survey establish that variation between men in hotspot activity is common, particularly at less active hotspots, and provide new examples of on/off polymorphisms suitable for in-depth analysis to determine the causes of
hotspot appearance and disappearance (9, 13, 19, 21-23). They have further revealed extreme and also more subtle biased gene conversion processes operating during crossing over, both of which could substantially influence the population behaviour of markers within hotspots. These very active hotspots will provide a valuable resource for characterising non-exchange gene conversion events within crossover hotspots (29), and also for beginning to explore recombination initiation and the processing of recombination intermediates in human meiosis.

Materials and Methods

DNA Samples. Blood and semen samples were collected with approval from the Leicestershire Health Authority Research Ethics Committee and with informed consent. DNAs were extracted and manipulated under conditions designed to minimise the risk of contamination as described elsewhere (7). Routine genotyping, optimisation of allele-specific primers and linkage phasing were performed on genomic DNAs that had been whole genome amplified using a GenomiPhi HY DNA amplification kit (GE Healthcare Bio-Sciences, Little Chalfont, Bucks, UK).

LD mapping and coalescent analysis. LDMAP (http://cedar.genetics.soton.ac.uk/pub/PROGRAMS/LDMAP) was used to create genome-wide LD maps in non-overlapping 500 SNP sections using 2,382,648 SNPs from 60 unrelated CEU individuals from the non-redundant filtered Phase II HapMap dataset (release 20) (http://www.hapmap.org). LD hotspots A-D were selected from LD maps constructed from early releases of Phase II HapMap data for chromosomes 6 and 21. Hotspots E-Q were selected from Phase II HapMap LD maps covering all autosomes combined with Phase I coalescent analysis information on historical
crossover activity available through HapMap. Strong LD hotspots were identified on the basis of a major (> 3) LDU step and high historical activity within a 5 kb window, then ranked by the product of LDU step and historical rate. Hotspots E-Q were chosen from the top 19, with others rejected due to low SNP marker density flanking the LD hotspot. LD hotspots A-D were ranked > 500th, 87th, 25th and 10th respectively on this list. More recent data available from Phase II HapMap (16) showed little difference between Phase I and II historical rates at these hotspots. Coalescent analysis was carried out on targets re-genotyped in semen donors using LDhat (17) (http://www.stats.ox.ac.uk/~mcvean/LDhat/) with 3,000,000 iterations, sampling every 2000th iteration and discarding the initial 10% of iterations as the burn-in period. Historical RFs were estimated assuming an effective population size of 10,000.

**Re-genotyping semen donors.** A 15 kb interval spanning each LD hotspot was amplified by long PCR as three overlapping amplicons and re-genotyped on a panel of 94 unrelated semen donors of north European origin by allele-specific oligonucleotide (ASO) hybridisation as described elsewhere (10). All Phase II HapMap SNPs were typed, plus some additional SNPs in dbSNP, yielding on average 26 HapMap SNPs plus 12 additional informative SNPs per target. Target sequence information and donor genotypes are available at http://www.le.ac.uk/genetics/ajj/superhotspots.

**Sperm crossover assays.** For each LD hotspot, three semen donors were chosen for analysis with suitable SNP heterozygosities flanking the hotspot that allowed crossover molecules to be recovered by repulsion-phase allele-specific long PCR (10),
plus sufficient internal markers to allow crossover breakpoints to be mapped. In total, 34 men were assayed for crossovers, at 1-3 LD hotspots per man. Assay intervals were 5.5–9.6 kb long depending on marker location. Allele-specific primers were optimised by PCR amplifying DNA from men homozygous for the correct or incorrect allele at varying annealing temperatures. SNP markers across the LD hotspot were phased in each man by allele-specific PCR amplification between selector SNPs upstream and downstream of the hotspot.

Crossover DNA molecules were selectively amplified from sperm DNA by nested repulsion-phase allele-specific long PCR across the hotspot using methods described elsewhere (10). PCR products were analysed by agarose gel electrophoresis and staining with ethidium bromide, allowing each PCR reaction to be scored as positive or negative for a crossover molecule. Crossover breakpoints were mapped by re-amplifying PCR products using PCR primers located internally to the allele-specific primer sites and typing these PCR products by ASO hybridisation (10). Primer sequences and crossover assay conditions for each hotspot are available at http://www.le.ac.uk/genetics/ajj/superhotspots.

The crossover frequency at each hotspot was initially estimated by a pilot assay on pools of sperm DNA containing 50-1600 amplifiable molecules of each haplotype per PCR. Assay specificity was verified by parallel analysis of blood DNA; no mitotic exchanges were seen at any of the hotspots. The full-scale (96-well) assay used the pilot crossover frequency estimate to amplify pools of sperm DNA of various size containing 0.7-1.8 crossover molecules per PCR, yielding typically 110 crossovers per plate. Poisson correction of the numbers of crossovers of each type, identified by
mapping exchange points, to take into account DNA pools containing more than one crossover molecule is described elsewhere (10). A single molecule PCR efficiency of 50% (one amplifiable DNA molecule of each haplotype per 12 pg sperm DNA) was assumed throughout, based on extensive previous data on single-molecule long PCR (6, 10).

Reciprocal $a$ and $b$ crossover assays using independent sets of allele-specific primers (Fig. 5) gave very similar RFs with a median difference of 1.2-fold (range 1.0–1.7 over 42 different reciprocal assays). After Bonferroni correction, none of the differences was significant ($P > 0.05$), establishing that RF estimates were robust. Reproducibility was also tested by re-assaying three men for crossovers using different sperm DNA preparations; again, indistinguishable RF estimates were obtained (1.1–1.3-fold differences between assays).

**Screening crossover hotspots for sequence motifs.** The sequence of each hotspot interval assayed for crossovers was progressively scanned for motifs that showed 0, 1, 2, 3… mismatches with the sequence CCTCCCTNNCCAC reported to be associated with LD hotspots (16), and at each level of mismatch the distance between the sperm crossover hotspot centre and the nearest motif was measured. The probability that any of the motifs detected at each level would map by chance within this distance of the hotspot centre was then estimated and a significant association was declared at $P < 0.05$. These motifs are shown on sequences available at http://www.le.ac.uk/genetics/ajj/superhotspots.
ACKNOWLEDGMENTS. We thank Jane Blower and volunteers for providing blood and semen samples, Rita Neumann, Joan Sutherland and Julie Turner for technical assistance and colleagues for helpful discussions. This work was supported by grants to AJJ from the Medical Research Council, the Wellcome Trust (ref. 081227/Z/06/Z), the Royal Society and the Louis-Jeantet Foundation.
References


Fig. 1. Identifying strong candidate recombination hotspots from genotype data. (A) Metric LD map of a 500 kb region of chromosome 8 determined from Phase II HapMap genotypes of 60 unrelated CEU individuals, compared with the historical recombination map estimated by coalescent analysis of the same data and shown below (data in cM taken from HapMap). The strong LD hotspot E is marked. (B) Similar analysis of a 100 kb interval on chromosome 20 containing the strongest LD hotspot reported by HapMap (16). (C) LDU step sizes and historical recombination activity at the 15 strong LD hotspots A-Q selected for sperm crossover analysis, estimated from Phase II HapMap data and from genotypes of 94 north European semen donors. HapMap and donor LDU steps were both estimated across the same 15 kb interval genotyped in donors. LD hotspot R was omitted due to uncertainties in historical activity (B) as well as lack of markers preventing HapMap LDU step estimation in the 15 kb interval.

Fig. 2. Crossover hotspots identified by sperm typing. (A) The cumulative frequency of sperm crossovers across each test interval (A-R) established for three informative men (black, red, green circles) per target. Each man was assayed for reciprocal crossovers and the results combined. Only two men were available for typing at target B; one showed a very low crossover frequency and is not shown. Best-fit cumulative normal distributions (6) are shown in black, or in different colours for men showing unusual distributions. These distributions are compared with LDU profiles and historical recombination profiles determined from genotypes of 94 semen donors and normalised to 1 for each test interval. Double hotspots are arrowed. (B) Sperm RF per hotspot for each man, colour coded as in (A), plus mean RF (circle). Hotspots are ranked by mean activity. Hotspots labelled in red showed major variation in activity
between men, while those marked in green showed weaker variation with a 2-4 fold difference in activity between the most and least active man. (C) Hotspot widths within which 95% of crossovers occur, estimated from best-fit normal distributions, together with chromosomal coordinates of hotspot centres in the NCBI 36 assembly of the human genome.

Fig. 3. Relationship between DNA diversity and sperm crossover activity. Historical crossover activity and LDU step size were estimated from semen donor genotype data for each of the 16 strong LD hotspot regions (in red), with double hotspots combined, and for 15 previously characterised sperm hotspots genotyped in the same panel of semen donors (in black) (5-13). The dotted line shows the expected relationship between sperm and historical activity, corrected for the greater mean crossover frequency seen in female meiosis (27).

Fig. 4. Complex crossover molecules recovered from sperm DNA. SNPs are indicated by circles, filled for one progenitor haplotype or empty for the other. Each hotspot is indicated on the left, the man of origin of each crossover is colour coded as in Fig. 2A, and the number of crossovers of each class shown at right. Crossovers are aligned with respect to the hotspot centres, with the mean hotspot width of 1.45 kb indicated by shading. Crossovers marked with an asterisk show haplotype switching extending outside the hotspot.

Fig. 5. Biased transmission of hotspot markers to crossovers. (A) Strong transmission distortion in hotspot B. (i) The two haplotypes in the man assayed, plus typical crossovers in orientation a and b. (ii) Crossover activity in each marker interval, with
the best-fit normal distribution for $a$ plus $b$ crossovers indicated in red. (iii) Frequency of alleles from the black haplotype in reciprocal crossovers, normalised to equal numbers of $a$ plus $b$ crossovers and with 95% confidence intervals indicated. (B) Weak transmission distortion in hotspot P. (i) Best-fit normal distribution of $a$ plus $b$ crossovers (Fig. 2A). (ii) Transmission frequencies in combined $a$ plus $b$ crossovers in two men, colour coded as in Fig. 2A.
Fig. 1
Fig. 2

A

B

C

D

E

F

G

H

J

K

L

M

N

P

Q

R

kb

sperm
crossovers
LDU profile
historical RF

B

C

hotspot

chr.

centre

hotspot

28
Fig. 3

![Graph showing data points and trend lines.](image-url)
Fig. 4

[Diagram of various hotspots labeled with different letters (F, R, E, N, K, D, P, H, L, B, A). Each hotspot is marked with symbols indicating presence or absence. A yellow shaded area highlights certain regions within the kb scale ranging from -4 to +4.]
Fig. 5

A

\[ a \quad b \]

\[ ii \]

hotspot B

\[ iii \]

B

\[ i \]

hotspot P

\[ ii \]