Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot

Alec J. Jeffreys* and Rita Neumann

Department of Genetics,

University of Leicester,

Leicester LE1 7RH,

UK

* to whom correspondence should be addressed.

tel. +44 116 2523435

fax. +44 116 2523378

email ajj@le.ac.uk

ABSTRACT

Little is known about factors that influence the frequency and distribution of meiotic recombination events within human crossover hotspots. We now describe the detailed analysis of sperm recombination in the *NID*1 hotspot. Like the neighbouring MS32 hotspot, the *NID*1 hotspot is associated with a minisatellite, suggesting that hotspots predispose DNA to tandem repetition. Unlike MS32, crossover resolution breakpoints in *NID*1 avoid the minisatellite, producing a cold spot within the hotspot. This avoidance may be related to the palindromic nature of the minisatellite interfering with the generation and/or processing of recombination intermediates. The *NID*1 hotspot also contains a single nucleotide polymorphism (SNP) close to the centre which appears to directly influence the frequency of crossover initiation. Quantitative gene conversion assays show that this SNP affects the frequency of gene conversion and crossover to a very similar extent, providing evidence that conversions and crossovers are triggered by the same recombination initiating events. The recombination-suppressing allele is over-transmitted to recombinant progeny, and provides the most dramatic example to date of recombination-mediated meiotic drive, of a magnitude sufficient to virtually guarantee that the recombination suppressor will eventually replace the more active allele in human populations.

INTRODUCTION

There is substantial interest in understanding the fine-scale distribution of meiotic recombination events in the human genome, to understand the processes involved in recombination as well as how patterns and processes of recombination can impact on haplotype diversity (1-3). Direct analysis of these fine-scale distributions is not feasible by pedigree analysis but is possible using single molecule PCR methods to recover recombinant DNA molecules directly from sperm (4,5). This approach has shown, for all regions tested, that meiotic crossovers are far from randomly distributed, clustering instead into narrow hotspots 1-2 kb wide within which the great majority of crossovers are resolved (4-7). Limited surveys in the class II region of the major histocompatibility complex (5) and around minisatellite MS32 on chromosome 1 (8) have shown that these hotspots occur roughly every 30 kb, at locations that appear to be unpredictable from primary DNA sequence. They have a major impact on patterns of haplotype diversity, organising markers into blocks of strong association (haplotype blocks) separated by recombination hotspots. More extensive surveys of patterns of human haplotype diversity, analysed in particular by coalescent approaches to estimate local rates of historical recombination, have shown that these hotspots are likely to be a general feature of the human genome (9,10). DNA diversity evidence also suggests that hotspots might be rather ephemeral during evolution, appearing and disappearing on the timescale of recent primate evolution (11-14). This raises major questions about how hotspots come into existence, how long they persist and what causes them to go extinct.

Human crossover hotspots are active not only in crossover but also in gene conversion (15). This suggests that hotspots mark sites of recombination initiation, with initiating events, most likely double-strand DNA breaks, arising within a localised zone at

the centre of the hotspot (16). Limited data to date on two MHC hotspots and a hotspot in the pseudoautosomal pairing region PAR1 on the sex chromosomes indicate that recombination events are preferentially resolved as conversions without crossover (15), though it is not known whether this is a general feature of human hotspots.

Little is known about factors that might influence the types of DNA sequences associated with hotspots or the shape of a hotspot. All hotspots characterised to date show sperm crossover exchange points smoothly and symmetrically distributed across the hotspot centre (4-7). Such distributions would be expected if the processing of recombination intermediates (resection, strand invasion, subsequent branch migration at Holliday junctions and resolution) occurs smoothly and symmetrically in both directions from sites of initiation and is not influenced by local DNA sequences. The first hotspot to be defined at high resolution by sperm analysis is centred close to the highly variable minisatellite MS32 (*D1S8*) and appears to trigger both crossovers and conversions within the repeat array that can change the length of the minisatellite (6,17). However, the distribution of crossovers within and outside MS32 remains symmetric (6), suggesting that this minisatellite does not perturb the processing of recombination intermediates. Other minisatellites also show recombination-based instability (18-21) and, in at least one case, substantial crossover activity both in and near the repeat array (22). However, it is not known how these exchanges are distributed relative to the repeat DNA nor whether minisatellites are in general associated with nearby hotspots. The converse is not true – none of the hotspots characterised to date in the MHC and PAR1 contains a minisatellite (5,7).

Different recombination hotspots in humans show a wide range of sperm crossover frequencies with peak activities ranging from 0.4 to 300 cM/Mb, compared with a background activity outside hotspots of very approximately 0.04 cM/Mb (5,7,8). The causes of this huge rate variation remain completely unknown, but most likely reflect differences in the efficiency of recombination initiation at different hotspots. There is also evidence for polymorphism between men in crossover frequency within hotspots. The MS32 hotspot shows a strongly-suppressed version, preferentially found in Africans, that is associated with and probably caused by a single base change in the hotspot (6). A similar phenomenon has been seen in hotspot *DNA*2 in the MHC, where haplotype analysis has identified a single SNP very close to the hotspot centre that appears to influence directly the frequency of initiation (23). The result is that reciprocal crossovers in active/suppressed heterozygotes map to different locations (reciprocal crossover asymmetry), with the recombination-suppressing allele being preferentially transmitted to crossover progeny (biased gene conversion accompanying crossover). This meiotic drive in favour of the recombination suppressor at hotspot *DNA*2 is however weak at the population level and only modestly increases the likelihood that the suppressor will eventually sweep to fixation. It is unclear whether such recombination-based meiotic drive is a rarity or instead a common feature of human recombination hotspots. It is also not known whether this meiotic drive influences gene conversion events without exchange as well as crossovers.

To address these issues, we now describe a detailed analysis of hotspot morphology, DNA sequence features and recombination frequency variation in the *NID*1 crossover hotspot. This hotspot was discovered in a survey of LD patterns and sperm

recombination hotspots in a region around the minisatellite MS32 hotspot and is the nearest hotspot centromeric to the MS32 hotspot (8). It is also very active, with a mean peak crossover activity of 70 cM/Mb similar to that seen at the MS32 hotspot.

RESULTS

Location and morphology of the *NID*1 hotspot

The *NID*1 recombination hotspot is located 58 kb upstream from the minisatellite MS32 hotspot (8) and is centred in intron 4 of the *NID* (nidogen, entactin) (24) gene (Fig. 1A). Sperm DNAs from two men (man 1, man 2) heterozygous for multiple single nucleotide polymorphism markers (SNPs) in and near the hotspot were analysed using repulsion-phase allele-specific PCR to selectively amplify recombinant DNA molecules (Fig. 1B). Exchange breakpoint mapping of reciprocal crossover products (Fig. 1C) showed that the hotspot was centred close to a 330 bp long AT-rich tract of DNA (91% AT) consisting of a degenerate AT repeat without clear higher-order repeat structure. This tract is preceded by a short minisatellite (MSNID) consisting, in the reference human genome sequence, of 7 repeats of a 34 bp AT-rich sequence that appears to have evolved from the 330 bp tract. Unlike other human hotspots (5-7), the distribution of exchange points across hotspot *NID*1 does not follow a normal distribution but instead appears to be skewed, with an excess of exchanges mapping well to the left of the hotspot centre.

Exchange points avoid minisatellite MSNID

Highly variable GC-rich minisatellites often generate variant repeat arrays by meiotic recombination-based processes (17-21) and, in two cases at least, are associated with high levels of crossover activity (6,22). The location of minisatellite MSNID entirely within a very active hotspot suggests that this AT-rich minisatellite (Fig. 2A), which shows a strong palindromic structure (Fig. 2B), might also be highly variable as a result of repeat array instability driven by recombination.

Sequence analysis of MSNID showed that alleles were short (2-7 repeats, with some showing a duplication of a different AT-rich motif adjacent to the repeat array) and in one case very similar to the chimp orthologue which also contains two repeats. Patterns of variant repeats revealed only a modest repertoire of different human alleles (Fig. 2C). The heterozygosity at MSNID is low (~0.58) and predicts a mutation rate to new length alleles of only $3x10^{-5}$ per gamete, considerably lower than the crossover frequency in hotspot *NID*1 ($3x10^{-4}$ per sperm). Analysis of flanking haplotypes around the most common MSNID allele (frequency ~0.6) provided additional evidence that this minisatellite might be largely immune to the effects of recombination (Fig. 2D). Identical allele structures were found on a wide variety of different haplotypes, with evidence of historical exchanges, between SNP markers M-58.9 and M-58.3 immediately flanking the minisatellite, that have created all four SNP haplotypes but without altering minisatellite structure.

To test directly the recombinational activity of MSNID, we analysed sperm crossovers in three MSNID heterozygotes (men 1 and 2, plus an additional man 3) and sequenced exchanges that mapped to the M-58.9/M-58.3 interval to locate crossover points relative to MSNID (Fig. 3). Of 41 such exchanges, only one mapped within MSNID compared to 11 expected if exchanges were randomly distributed in the M-58.9/M-58.3 interval ($P < 0.001$). This single exchange was between MSNID alleles aligned at their 5' ends (Fig. 3) and created a recombinant allele, without change in repeat copy number, that was identical to an allele already seen in the diversity survey. The male crossover frequency within MSNID (1 exchange seen in 1.3×10^6 sperm) is even lower than the mutation rate inferred from diversity data (3×10^{-5}). Diversity surveys and direct sperm analysis both therefore show that recombination exchange points in the *NID*1 hotspot strongly avoid minisatellite MSNID which thus constitutes a cold spot, splitting the hotspot in two (Fig. 3).

Crossover frequency polymorphism and reciprocal crossover asymmetry

Sperm crossover analysis showed that man 3 had a significantly higher recombination frequency than man 1 or man 2 (Fig. 3), indicating polymorphism in exchange rates at hotspot *NID*1. To investigate this polymorphism in more detail, we analysed four additional men selected to include a range of haplotypes in and around the hotspot. Each man was analysed for reciprocal crossovers (A-type and B-type; see Fig. 1B) as well as for overall crossover frequency (Fig. 4).

All men showed indistinguishable cumulative frequencies of combined A+B crossovers across the hotspot (Fig. 4A). However, analysis of A and B crossovers separately revealed significant reciprocal crossover asymmetry in five of the men, with B crossover distributions consistently displaced on average 440 bp away from A crossovers. Such asymmetry leads to transmission distortion of hotspot alleles into crossover progeny (Fig. 4C). This distortion was strongest for marker M-57.8C/T located just 70 bp from the centre of the hotspot and was consistently in favour of allele T. Crossovers in C/T heterozygotes showed on average a 74:26 transmission ratio of allele T vs. C into crossovers, corresponding to a conversion bias of 2.8:1 in favour of T during crossover. Weaker transmission distortion was also seen at the other hotspot marker M-58.3T/C located further from the centre (410 bp). However, this distortion was only seen in M-57.8C/T heterozygotes and the identity of the over-transmitted M-58.3 allele varied according to which allele was linked to M-57.8T. Marker M-57.8C/T was the only SNP in and around the hotspot that was heterozygous in all men showing asymmetry but homozygous, for either allele M-57.8C or M-57.8T, in the two men with no evidence of asymmetry as judged by transmission of marker M-58.3T/C. The simplest interpretation is that heterozygosity at M-57.8C/T is both necessary and sufficient to trigger asymmetry and biased gene conversion at both hotspot markers.

As discussed previously (23), reciprocal crossover asymmetry can arise either through biased heteroduplex repair (specifically affecting in this case marker M-57.8 in favour of the T allele) or through disparity in the frequency of crossover initiation on the two haplotypes in an individual. In the latter case, markers from the initiating haplotype will tend to be converted by markers from the other haplotype, allowing active and

suppressed haplotypes to be identified in heterozygotes (Fig. 4B). Under this model, the over-transmitted allele M-57.8T must cause a suppression of initiation relative to allele M-57.8C, by a factor of at least 2.8-fold to give the level of transmission distortion observed. Consistent with this initiation model, man 3 homozygous for the active allele C shows no significant asymmetry and the highest crossover frequency, while man 7 homozygous for the suppressing allele T shows the second from lowest frequency, again without asymmetry. Rank order analysis of crossover frequencies in all men shows that this correlation with M-57.8 status is significant (P = 0.048), providing evidence that asymmetry and transmission distortion arise from effects of this SNP on the efficiency of initiation rather than from biased repair.

Marker M-57.8 shows a corresponding bias in the frequency of gene conversion without exchange

Reciprocal crossover asymmetry provides a powerful rate-independent method for detecting disparity in the frequencies of gene conversion accompanying crossover (23). If gene conversion events without crossover result from the same events that initiate crossover, then the frequency of gene conversional transfer, without crossover, of M-57.8 allele C to chromosomes carrying allele T should be lower, by a factor of ~2.8-fold, than the frequency of transfer of T into C chromosomes (Fig. 5A). We tested this prediction by directly measuring gene conversion frequencies in sperm at marker M-57.8.

Gene conversions were detected by hybridisation enrichment of sperm DNA (15,25) from man 4 to select for molecules carrying either M-57.8C or M-57.8T. Allele-specific PCR directed to sites either side of the hotspot on the depleted haplotype was then used to selectively amplify recombinant molecules plus any remaining molecules of the depleted haplotype (Fig. 5B). This enrichment procedure greatly increases the signal from recombinants. It also allows conversions and crossovers to be detected simultaneously (Fig. 5C) and validates exchange events (25). Analysis of the other hotspot marker M-58.3C/T was not possible as this SNP lies in the AT-rich region (Fig. 1A) and gave extremely poor yields of DNA during enrichment.

Conversion and crossover data are summarised in Fig. 5D. The crossover frequency ($5.0x10^{-4}$ per sperm) was indistinguishable in the two enrichments and was the same as that established by normal crossover analysis ($5.3x10^{-4}$; Fig. 4B). In addition to crossovers, 100 conversion events involving M-57.8 were also detected. Almost all affected M-57.8 alone, with only one showing a conversion tract that included more than one marker (markers M-57.8 plus M-58.3, the only other SNP marker within the hotspot). This establishes that conversion tracts are short (one tract 0.4-3.7 kb long and all others shorter than 1.2 kb), though marker density was insufficient to estimate tract lengths more precisely. The mean conversion frequency was $1.3x10^{-4}$, about 4-fold lower than the crossover frequency. As predicted, there was a significant disparity in conversion frequencies of M-57.8C→T and T→C, with the former occurring ~2.5-fold more frequently than the latter when estimated either from the numbers of molecules analysed or by comparing the numbers of conversions and crossovers detected in each assay; the latter measure is independent of the numbers of molecules tested. The disparity of

conversions in favour of M-57.8T (~2.5-fold) is similar to that predicted from the strength of reciprocal crossover asymmetry seen in M-57.8T/C heterozygotes (2.8-fold both in man 4 and averaged over all five heterozygotes).

DISCUSSION

Identifying features that influence hotspot activity in humans is dependent on sequences and variants available in human populations. At first glance, the *NID*1 hotspot is entirely unremarkable, with a peak activity (70 cM/Mb) and width (1.5 kb) fully within the range seen at other human hotspots (4-8). Detailed analysis has however revealed several unusual features. The first is the presence of minisatellite MSNID located entirely within the hotspot. The only other minisatellite with repeat length >8 bp within a 140 kb region around MS32 is the MS32 minisatellite itself, again associated with a hotspot (6). This provides evidence for a preferential association between minisatellites and hotspots and suggests that hotspot activity may actively promote tandem repetition of reasonably long repeat units (29 bp for MS32, 34 bp for MSNID). This is further supported by the existence of a second polymorphic duplication of 35 bp next to MSNID (Fig. 2C).

However, MS32 and MSNID have very different impacts on their host hotspots. Minisatellite MS32 is 62% GC-rich and extremely variable (26). It is not palindromic and appears not to influence the distribution of crossover exchange points across the hotspot, which is centred upstream of the minisatellite and extends only into the beginning of the repeat array (6). In contrast, MSNID shows modest variability, is very GC-poor (20%

GC) and has a palindromic sequence. Crossovers are rarely resolved within this minisatellite and it thus constitutes a cold spot within the *NID*1 hotspot. This effect cannot be attributed to its AT-richness, as the neighbouring AT-rich domain, which has less palindromic structure than MSNID (not shown), appears to be fully active in crossover (Fig. 1C). The cold-spot phenomenon must therefore depend on the palindromic or tandem repetitive nature of MSNID. Perhaps palindromes cause fold-back of resected 3' ends that happen to terminate in MSNID, preventing strand invasion into the homologous chromosome, establishment of a recombination complex and the generation of exchanges that map within the minisatellite. Another possibility is that heteroduplex DNA is formed between MSNID alleles but that subsequent mismatch repair fully restores the hybrid DNA to one or other parental state.

There is however an alternative interpretation for the MSNID cold spot. Crossover resolution sites appear not to be symmetrically distributed across the *NID*1 hotspot, but show an excess of exchanges that map upstream of the interval containing MSNID (Fig. 1C). However, if this interval is shortened by the length of MSNID (214-248 bp), then this excess largely disappears and a better fit to the symmetric, normal distribution seen at all other characterised hotspots is obtained ($P < 0.001$ and $P = 0.11$ before and after shortening respectively; data not shown). The implication therefore is that exchanges that should resolve within MSNID are not being lost but instead are being displaced further away from the hotspot centre. The cause of this is unclear but could for example involve MSNID accelerating branch migration at Holliday junctions.

Another possibility is that the *NID*1 hotspot is in fact two hotspots separated by MSNID, with each hotspot containing its own zone of recombination initiation. Hotspot

clusters have indeed been seen in the MHC (5) and near MS32 (8). However, the hotspot upstream of MSNID would have to be narrower (0.6 kb) than any seen previously (1.0-2.1 kb, refs. 4-8). Further, the ratio of exchanges upstream and downstream of MSNID remains constant in men with different crossover frequencies, and both types of exchange show reciprocal crossover asymmetry (Fig. 4A, B). This is consistent with exchanges at a single but not a double hotspot at *NID*1.

Hotspot *NID*1 provides the second example of a common single nucleotide change in a hotspot (current allele frequency of 0.49 in north Europeans) that, when heterozygous, appears to cause non-mendelian transmission of hotspot markers to crossover progeny. The parallels with the MHC hotspot *DNA*2, which shows the same phenomenon (23), are striking. Both involve the SNP closest to the centre of the hotspot (~70 bp away from the centre for *NID*1, ~5 bp for *DNA*2) and both are simple transitional polymorphisms (C/T at *NID*1, A/G at *DNA*2) located in single-copy DNA. The allele over-transmitted to crossover progeny is T:A at *NID*1 and G:C at *DNA*2, suggesting no common pattern as to which allele is favoured by recombinational meiotic drive. In both cases, great ape comparisons show that the over-transmitted allele is the derived state and has therefore recently evolved from the under-transmitted allele, consistent with the direction of meiotic drive. The strength of over-transmission to crossover progeny is almost identical in both hotspots (mean 74:26 at *NID*1, 76:24 at *DNA*2) and the accompanying level of reciprocal crossover asymmetry is also virtually identical (A and B crossover distributions displaced on average by 440 bp at *NID*1, 430 bp at *DNA*2). However, comparison of DNA sequences around these SNPs failed to reveal any obvious

sequence similarities which could help explain how these sequence differences can trigger meiotic drive.

In addition to *NID*1 and *DNA*2, there is good evidence for a similar drive phenomenon at the MS32 hotspot near *NID*1, though detailed analysis of reciprocal crossover asymmetry has not proved possible (6,23). Thus, 2-3 out of 15 human recombination hotspots analysed to date show meiotic drive. This is a minimum estimate since hotspot crossover surveys to date have been limited to small numbers of men and may have missed heterozygotes at markers that can cause crossover disparity and meiotic drive. It is therefore likely that meiotic drive within human crossover hotspots is a common phenomenon.

For both *NID*1 and *DNA*2, crossover frequency evidence suggests that meiotic drive results from allelic influences on the rate of crossover initiation, with the suppressing allele being over-transmitted due to repair of the initiating chromosome with information from the other chromosome. The involvement of biased mismatch repair cannot however be excluded, though is a less likely mechanism (23). The present study also shows that conversion without crossover in hotspot *NID*1 also shows meiotic drive at a level quantitatively indistinguishable from that seen in crossovers. This correlation is expected if crossovers and conversions arise from the same recombination-initiating events, the frequency of which are influenced by SNP M-57.8C/T. These biased conversions also add to the overall level of meiotic drive at *NID*1. However, the increase is unexpectedly low given the low ratio of conversions to crossovers (1:4, compared to 2.7:1 for the MHC hotspot *DNA*3; ref. 15). This 11-fold discrepancy cannot be readily explained by differences between the hotspots in the numbers of short marker-less

conversion tracts that cannot be detected (15), since the distance between the hotspot centre and the closest marker assayed for conversion is similar for both hotspots (~70 bp for *NID*1, ~85 bp for *DNA*3). It therefore appears that the balance between conversions and crossovers can vary between hotspots, as has also been seen in yeast (27). It is not known what controls this shifting bias in crossover outcome (3), but highlights the problem of deducing recombination initiation rates from crossover frequencies.

As noted previously (23), transmission bias in favour of the recombination-suppressing variant at hotspot *DNA*2 results in weak meiotic drive at the population level, due to the low recombination frequency at this hotspot ($3.7x10^{-5}$ per sperm for crossovers, conversion frequency unknown). This in turn only modestly affects the likelihood that the recombination-suppressing variant will achieve eventual fixation in a population. For hotspot *NID*1, the situation is very different. The mean crossover frequency in M-57.8C/T heterozygotes is much higher ($5.5x10^{-4}$ per sperm) and the conversion frequency of $1.3x10^{-4}$ at M-57.8 contributes further, though modestly, to the level of meiotic drive. The observed bias in favour of M-57.8T in crossovers and conversions would lead to a non-mendelian transmission ratio of 50.009:49.991 of T:C from one generation to the next, a degree of distortion at the population level 18-fold more intense than estimated from crossovers at hotspot *DNA*2. Population simulations showed that this level of meiotic drive has a major effect, increasing the likelihood of eventual fixation of the suppressing T allele, currently at a frequency of 0.49 in Europeans, from 49% for a non-driven allele to >95%, even if the hotspot is only active in male meiosis. Thus variants in strong hotspots that suppress initiation, or are favoured by biased repair, are very likely to sweep to eventual fixation. As noted previously, these

levels of drive do not however substantially accelerate the fixation process. For M-57.8T/C, fixation of T will be achieved in on average $1.2N_e$ generations (where N_e is the effective population size), compared to $2.8N_e$ generations for an undriven allele.

Hotspot *NID*1 thus provides further evidence for processes involved in the attenuation and elimination of recombination hotspots and strengthens the "hotspot paradox" (28,29), namely how hotspots can exist in the face of a systematic bias towards fixation of recombination-suppressing alleles. What remains wholly unclear is how hotspots arise in the first place and how long they can persist in the face of meiotic drive of recombination suppressors before they eventually become extinct.

MATERIALS AND METHODS

Sperm DNAs and genotyping

Semen and blood samples were collected, with informed consent and approval from the Leicestershire Health Authority Research Ethics Committee, from 200 UK men of north European descent including volunteers and men attending fertility clinics. Eighty of these men showing good sperm DNA yields were selected for analysis. Sperm DNAs were whole-genome amplified by MDA (30) and genotyped by subsequent PCR amplification of appropriate 3-8 kb targets followed by allele-specific oligonucleotide (ASO) hybridisation to dot blots of PCR products (4). Details of SNPs, ASOs and genotypes are provided at http://www.le.ac.uk/ge/ajj/MS32/. SNPs are named according to their location relative to the centre of minisatellite MS32 (e.g. SNP M-61.6T/C is located 61.6 kb upstream of the minisatellite). Minisatellite MSNID structures were determined by separating haplotypes from men heterozygous at SNP M-55.5T/C, using allele-specific PCR directed to this site, then PCR amplifying the minisatellite interval from each haplotype and sequencing it using BigDye™ Terminators (ABI) on an ABI 377 Automated Sequencer. The mutation rate μ at MSNID was estimated from the heterozygosity H assuming mutation/drift equilibrium and an infinite allele model, under which $H = 4N_e\mu/(1+4N_e\mu)$, where N_e is the effective population size (assumed to be 10,000; ref. 31).

Sperm crossover assays

Men were identified who were heterozygous for markers 5' and 3' to the *NID*1 hotspot; in most cases these markers used for crossover recovery were M-61.6T/C, M-61.5T/C and M-55.4C/T, M-55.5T/C. Allele-specific primers (ASPs) were designed for each SNP and tested by PCR on genomic DNA from individuals homozygous for the correct or incorrect allele, to identify primers that showed excellent efficiency and allele specificity and to determine optimum annealing temperatures. Linkage phase between 5' and 3' SNPs was determined by PCR amplification of genomic DNA using these ASPs. Sperm DNAs for recombination analysis were prepared as described previously under conditions designed to minimise the risk of contamination (4). To selectively amplify crossover molecules, multiple batches of sperm DNA each containing, depending on crossover frequency, 300-7000 amplifiable molecules of each progenitor haplotype (3.6-84 ng DNA containing 0.4-2 crossover molecules, with 50% efficiency of amplifying a single DNA molecule) were amplified in 96-well plates by long PCR using ASPs in repulsion phase; these ASP pairs were selected for compatible annealing temperatures. These primary PCR products were digested with S1 nuclease to remove single-stranded DNA and PCR artifacts, and re-amplified using nested internal ASPs in repulsion phase. These secondary PCRs were analysed by agarose gel electrophoresis and DNA visualised by staining with ethidium bromide to identify crossover-positive reactions. Secondary PCRs were re-amplified using nested non-allele-specific primers and crossover exchange points mapped by dot-blot hybridisation of these tertiary PCR products with ^{32}P-labelled ASOs. All crossover assays included multiple aliquots of blood DNA; no examples of crossovers were detected in these negative controls. Full details of long PCR, S1 digestion, crossover

mapping and Poisson correction for more than one crossover per PCR are given elsewhere (4,5). Details of ASPs and crossover assay conditions are provided at http://www.le.ac.uk/ge/ajj/MS32/.

DNA enrichment for gene conversion assays

To enrich for DNA molecules carrying allele M-57.8C, 40 μg sperm DNA from man 4 were digested with *Avr*II to release the target region on a DNA fragment 6.6 kb long, then subjected to four rounds of denaturation and hybridisation at 47.5°C with a 5'-biotinylated oligonucleotide specific for M-57.8C (5' bio-TTAAATCACCCCCCACCC 3') in the presence of excess non-biotinylated competitor ASO (5' TTAAATCACCTCCCACCC 3'). Hybrids at each round were captured on Dynabeads™ M-280 Streptavidin (Dynal Biotech) and single-stranded target DNA recovered by thermal denaturation of the bio-oligonucleotide/genomic DNA hybrid. Enrichment for M-57.8T was similarly performed using 5' bio-TTAAATCACCTCCCACCC 3' and competitor 5' TTAAATCACCCCCCACCC 3'. Full details of enrichment procedures are provided elsewhere (15,25). Single-stranded DNAs from all four rounds of enrichment were pooled and assayed for the number of haplotype 1 and 2 molecules (Fig. 5) by amplifying dilution series of starting and enriched genomic DNA, using a reverse allele-specific primer directed to M-55.5C or M-55.5T in combination with the universal (not allele-specific) forward primer M-60.3F2. Primer sequences are provided at http://www.le.ac.uk/ge/ajj/MS32/. Yields of PCR products indicated that selection with

M-57.8C had yielded 36% of target molecules enriched 50-fold over the other haplotype. Similarly, M-57.8T enrichment gave a 30% yield and 80-fold enrichment.

Assaying gene conversions

Each of the enriched DNA fractions was assayed for conversions as described previously (15,25). Briefly, multiple aliquots of the M-57.8C-enriched pool (376 reactions in total, each containing 540-1090 amplifiable DNA molecules) were amplified using a forward primer specific for allele M-61.6C on the depleted haplotype and the reverse universal primer M-55.4R (experiment 1, Fig. 5B). Primary PCR products were diluted 20-fold with water and 0.5 μl aliquots used to seed secondary PCRs which were amplified using a M-61.5T-specific forward primer plus the reverse universal primer M-55.5R. Secondary PCR products were analysed by dot-blot hybridisation with ^{32}P-labelled ASOs. Controls (conversion events not seen at unselected SNPs, conversions showing a signal appropriate for the complexity of the pool being analysed, detection of crossovers switched at multiple SNP sites as an internal control for exchange rates) are described in detail elsewhere (15,25). The other conversion experiments, performed with minor modifications of cycling conditions to maximise efficiency and allele-specificity, were: experiment 2, M-57.8C-enriched DNA, primary PCR with forward universal primer M-61.5F and reverse allele-specific primer M-55.4T, secondary PCR with forward universal primer M-61.5bF and reverse allele-specific primer M-55.5C; experiment 3, M-57.8T-enriched DNA, primary PCR with M-61.6T plus M-55.4R, secondary PCR with M-61.5G plus M-55.5R; experiment 4, M-57.8T-enriched DNA, primary PCR with M-61.5F plus

M-55.4C, secondary PCR with M-61.5bF plus M-55.5T. PCR cycling conditions are available on request, and primer sequences are provided at http://www.le.ac.uk/ge/ajj/MS32/.

The number of amplifiable molecules in each enriched fraction was assayed by diluting the enriched single-stranded DNA with 5 mM Tris-HCl (pH 7.5) plus 5 μg/ml double-stranded herring DNA carrier and 5 μg/ml single-stranded herring DNA carrier, then amplifying as above, over 240 PCR reactions with inputs of ~1 molecule/PCR, but using primers specific to alleles on the enriched, not depleted, haplotype. Numbers of amplifiable molecules were estimated by Poisson analysis of the numbers of positive secondary PCR reactions.

Crossovers and conversions detected in these recombination assays were corrected for PCRs containing more than one recombinant molecule and for reactions containing a conversion concealed by a crossover (15). Confidence intervals for conversion frequencies were established by full maximum likelihood analysis, systematically varying parameters (numbers of molecules in each enriched pool, crossover frequencies and M-57.8C→T, T→C conversion frequencies) to identify the combination that maximised the likelihood of obtaining the entire data set. Confidence intervals were identified from the most extreme values that gave a likelihood no less than 5% of the maximum likelihood.

Population simulations

Computer simulations used to estimate the probability of population fixation of an allele subject to meiotic drive were performed as described previously (15), with transmission

between generations being modified to reflect known rates of crossover and transmission distortion in crossovers, plus known rates of gene conversion. The effective population size was assumed to be 10,000 (31); larger contemporary population sizes will increase the likelihood of fixation of a driven allele.

ACKNOWLEDGEMENTS

We thank J. Blower and volunteers for providing semen and blood samples, S. Mistry for oligonucleotide synthesis and assistance with DNA sequencing, and colleagues for helpful discussions. This work was supported by grants to AJJ from the Medical Research Council, the Royal Society and the Louis-Jeantet Foundation.

REFERENCES

1. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225-2229.
2. The International HapMap Consortium. (2003) The International HapMap Project. *Nature*, **426**, 789-796.
3. Kauppi, L., Jeffreys, A.J. and Keeney, S. (2004) Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.*, **5**, 413-424.
4. Jeffreys, A.J., Ritchie, A. and Neumann, R. (2000) High-resolution analysis of haplotype diversity and meiotic crossover in the human *TAP2* recombination hotspot. *Hum. Mol. Genet.*, **9**, 725-733.
5. Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.*, **29**, 217-222.
6. Jeffreys, A.J., Murray, J. and Neumann, R. (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell*, **2**, 267-273.
7. May, C.A., Shone, A.C., Kalaydjieva, L., Sajantila, A. and Jeffreys, A.J. (2002) Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene *SHOX*. *Nat. Genet.*, **31**, 272-275.

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

9. Crawford, D.C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M.J., Nickerson, D.A. and Stephens M. (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.*, **36**, 700-706.

10. McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R. and Donnelly, P (2004). The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581-584.

11. Wall, J.D., Frisse, L.A., Hudson, R.R. and Di Rienzo, A. (2003) Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. *Am. J. Hum. Genet.*, **73**, 1330-1340.

12. Ptak, S.E., Roeder, A.D., Stephens, M., Gilad, Y., Paabo, S. and Przeworski, M. (2004) Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS Biol.*, **2**, 849-855.

13. Winckler, W, Myers, S.R., Richter, D.J., Onofrio, R.C., McDonald, G.J., Bontrop, R.E., McVean G,A., Gabriel, S.B., Reich, D., Donnelly, P. and Altshuler, D. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*, **308**, 107-111 (2005).

14. Ptak, S.E., Hinds, D.A., Koehler, K., Nickel, B., Patil, N., Ballinger, D.G., Przeworski, M., Frazer, K.A. and Paabo, S. Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.*, **37**, 429-434 (2005).

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

16. Qin, J., Richardson, L.L., Jasin, M., Handel, M.A. and Arnheim, N. (2004) Mouse strains with an active H2-Ea meiotic recombination hot spot exhibit increased levels of H2-Ea-specific DNA breaks in testicular germ cells. *Mol. Cell. Biol.*, **24**, 1655-1666.

17. Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.*, **6**, 136-145.

18. Buard, J. and Vergnaud, G. (1994) Complex recombination events at the hypermutable minisatellite CEB1 (*D2S90*). *EMBO J.*, **13**, 3203-3210.

19. Tamaki, K., May, C.A., Dubrova, Y.E. and Jeffreys A.J. (1999) Extremely complex repeat shuffling during germline mutation at human minisatellite B6.7. *Hum. Mol. Genet.*, **8**, 879-888.

20. Stead, J.D.H. and Jeffreys, A.J. (2000) Allele diversity and germline mutation at the insulin minisatellite. *Hum. Mol. Genet.*, **9**, 713-723.

21. Berg, I., Neumann, R., Cederberg, H., Rannug, U. and Jeffreys, A.J. (2003) Two modes of germline instability at human minisatellite MS1 (locus *D1S7*): complex rearrangements and paradoxical hyperdeletion. *Am. J. Hum. Genet.*, **72**, 1436-1447.

22. Buard, J., Shone, A.C. and 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.

23. Jeffreys, A.J. and Neumann, R. (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hotspot. *Nat. Genet.*, **31**, 267-271.

24. Zimmermann, K., Hoischen, S., Hafner, M. and Nischt, R. (1995) Genomic sequences and structural organization of the human nidogen gene (*NID*). *Genomics*, **27**, 245-250.

25. Jeffreys, A.J. and May, C.A. (2003) DNA enrichment by allele-specific hybridization (DEASH): a novel method for haplotyping and for detecting low-frequency base substitutional variants and recombinant DNA molecules. *Genome Res.*, **13**, 2316-2324.

26. Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. and Monckton, D.G. (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature*, **354**, 204-209.

27. Fogel, S., Mortimer, R. K. and Lusnak, K. (1981) in *The Molecular Biology of the Yeast Saccharomyces* (eds. Strathern, J.N., Jones, E.W. and Broach, J.R.) pp.289–339 (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

28. Boulton, A., Myers, R.S. and Redfield, R.J. (1997) The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci. USA*, **94**, 8058-8063.

29. Pineda-Krch, M. & Redfield, R. Persistence and loss of meiotic recombination hotspots. *Genetics*, **169**, 2319-2333 (2005).

30. Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J. *et al.* (2002) Comprehensive human genome amplification

using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA*, **99**, 5261-5266.

31. Morton, N.E. (1982) *Outline of genetic epidemiology* (Karger, Basel).

FIGURE LEGENDS

Figure 1. Location and morphology of the *NID*1 recombination hotspot. **(A)** Location in *NID* intron 4, with *Alu* repeats shown as hatched boxes and the AT-rich region next to minisatellite MSNID indicated in grey. The locations of SNPs are shown below. **(B)** Strategy for recovering crossover molecules from sperm DNA. Two rounds of nested allele-specific PCR using primers in repulsion phase specific to the black or white haplotype (arrows) are used to selectively amplify either A-type crossovers (black to white) or B-type crossovers (white to black) from sperm DNA. Crossover breakpoints in these recombinant DNA molecules are mapped by typing internal informative SNPs. **(C)** Sperm crossover distribution across the *NID*1 hotspot. A and B crossovers recovered from sperm DNA from two men (man 1, 51 A crossovers, 67 B crossovers, each recovered from 180,000 amplifiable molecules of each progenitor haplotype; man 2, 58 A crossovers, 78 B crossovers, each recovered from 310,000 molecules) were combined to determine the numbers of crossovers (in italics) in each interval between heterozygous SNPs, from which the crossover activity in cM/Mb was estimated. Man 1 and man 2 showed similar crossover frequencies (3.3×10^{-4}, 2.2×10^{-4} per sperm respectively) and indistinguishable crossover distributions (not shown). The solid line shows the least-squares best-fit distribution assuming that crossovers are normally distributed across the hotspot (5). The fit is poor ($P < 0.001$), with evidence that the true distribution is skewed to the left (distribution symmetry test, $z = 1.86$, two-tailed $P = 0.06$).

Figure 2. Properties of minisatellite MSNID. **(A)** Repeat sequence of MSNID with variant repeats indicated below. **(B)** Predicted foldback structure, illustrated using 2.5 repeats. **(C)** Allelic diversity in human MSNID from a survey of 24 fully-resequenced alleles, with repeats coded as in (A). The adjacent 35 bp sequence ATAAATATAT TACATATATG TAATATATAT TAAAT (dark grey rectangle) is perfectly duplicated on some alleles. The chimpanzee MSNID structure was deduced from the draft chimp genome sequence. **(D)** Flanking haplotype diversity among fifteen identical human MSNID alleles, with SNP alleles coded black (first allele, e.g. M-61.6T) and white (second allele, e.g. M-61.6C). SNP locations are not to scale and LD blocks flanking the hotspot are shown below as black bars.

Figure 3. Location of sperm crossovers in hotspot *NID*1 relative to minisatellite MSNID. Three men were typed for crossovers. Data for men 1 and 2 are from Fig. 1C. Man 3 yielded 140 crossovers from 88,000 progenitor molecules of each haplotype; the shorter test interval is due to homozygosity at the M-61.5T/G selector site, forcing the use of SNP M-59.7C/G for crossover recovery (see Fig. 4B). All exchanges mapping to the interval containing MSNID were sequenced. Minisatellite structures so determined were compared with the progenitor alleles in each man (shown at right) to locate each exchange point. Only the region between the barred lines for each man is informative for locating minisatellite exchanges. Only one exchange, in man 1, mapped within MSNID;

its structure and the deduced location of the exchange is shown at right. The location of MSNID (hatched) and its associated AT-rich DNA (grey box) is indicated below.

Figure 4. Reciprocal crossover asymmetry and meiotic drive in hotspot *NID*1. **(A)** Cumulative frequencies of crossovers across the hotspot in seven men. A- and B-type crossovers were recovered and mapped from sperm from each man and are shown combined (left) and separately (right). Markers within the hotspot are arrowed and minisatellite MSNID shown as a box below. **(B)** Haplotypes, crossover frequency and the existence or otherwise of asymmetry in each man analysed. Markers in haplotypes are not shown to scale, and alleles are coded black or white as in Fig. 2C. If asymmetry arises from differences between haplotypes in recombination initiation rate, then the haplotype asterisked in each donor showing asymmetry must be the more active in initiation. Haplotype blocks flanking the hotspot (8) are indicated below by black bars. **(C)** Transmission frequencies of hotspot alleles to crossover products. The mean frequencies, plus 95% C.I.s, of transmission of alleles M-57.8T and M-58.3T to equal numbers of A- and B-type crossovers are shown for each informative man. The red line shows the mean level of transmission of M-57.8T in M-57.8T/C heterozygotes. All data were determined from the following numbers of A- and B-type crossovers analysed/numbers of amplifiable progenitor molecules of each haplotype screened: man 1, A 51/180,000, B 67/180,000; man 2, A 57/310,000, B 137/540,000; man 3, A 140/78,000, B 140/97,000; man 4, A 152/270,000, B 80/160,000; man 5, A 32/46,000, B 65/75,000; man 6, A 108/250,000, B 120/250,000; man 7, A 101/290,000, B 95/330,000.

Figure 5. Disparity in gene conversion frequencies between M-57.8C and M-57.8T. **(A)** Relative frequencies of sperm conversions without exchange involving marker M-57.8, predicted from the mean strength of reciprocal crossover asymmetry seen in Fig. 4. **(B)** Procedure for assaying crossovers and conversions involving exchanges at M-57.8C/T in man 4. Sperm DNA containing a low frequency of conversion (C) and crossover (X) molecules was restricted and fractionated by hybridisation with a biotinylated ASO specific for M-57.8C or M-57.8T to remove the bulk of haplotype 2 or haplotype 1 molecules respectively (i). Enriched single-stranded DNA was then amplified using allele-specific primers directed to markers either 5' (experiments 1, 3) or 3' (experiments 2, 4) to the hotspot and specific for the depleted haplotype. PCR products will be derived from any remaining molecules of the depleted progenitor haplotype, together with any conversions and crossovers that include an exchange between M-57.8 and the allele-specific primer site. Note that experiments 1 and 2 should give the same conversion frequency but different crossover frequencies since not all crossovers will be detected (for example, experiment 1 will only yield crossovers that map 5' to M-57.8). The same is true for experiments 3 and 4. **(C)** Examples of PCRs from experiment 3 analysed by dot-blot hybridisation with ASOs specific for the enriched haplotype 2. Each PCR contained 940 amplifiable molecules of haplotype 2 plus ~12 remaining molecules of haplotype 1; the former will not be amplified and a recombinant molecule, if present, will constitute ~10% of the PCR products. Three conversions exchanged at M-57.8 only, plus six crossovers each exchanged in the interval M-58.3/M-57.8, were detected. M, control 10:1

mixtures of haplotype 1 : haplotype 2. **(D)** Summary of recombinants seen in experiments 1-4. The numbers of molecules screened were estimated by Poisson analysis of multiple limiting dilutions of each enriched DNA fraction. Crossover frequencies were corrected for the proportion of exchanges that would be missed in each experiment, using data on A- and B-type crossover distributions from Fig. 5A. The relative frequencies and confidence intervals for M-58.3T→C and C→T conversions shown below were determined by maximum likelihood analysis of the entire data set.

Fig. 1

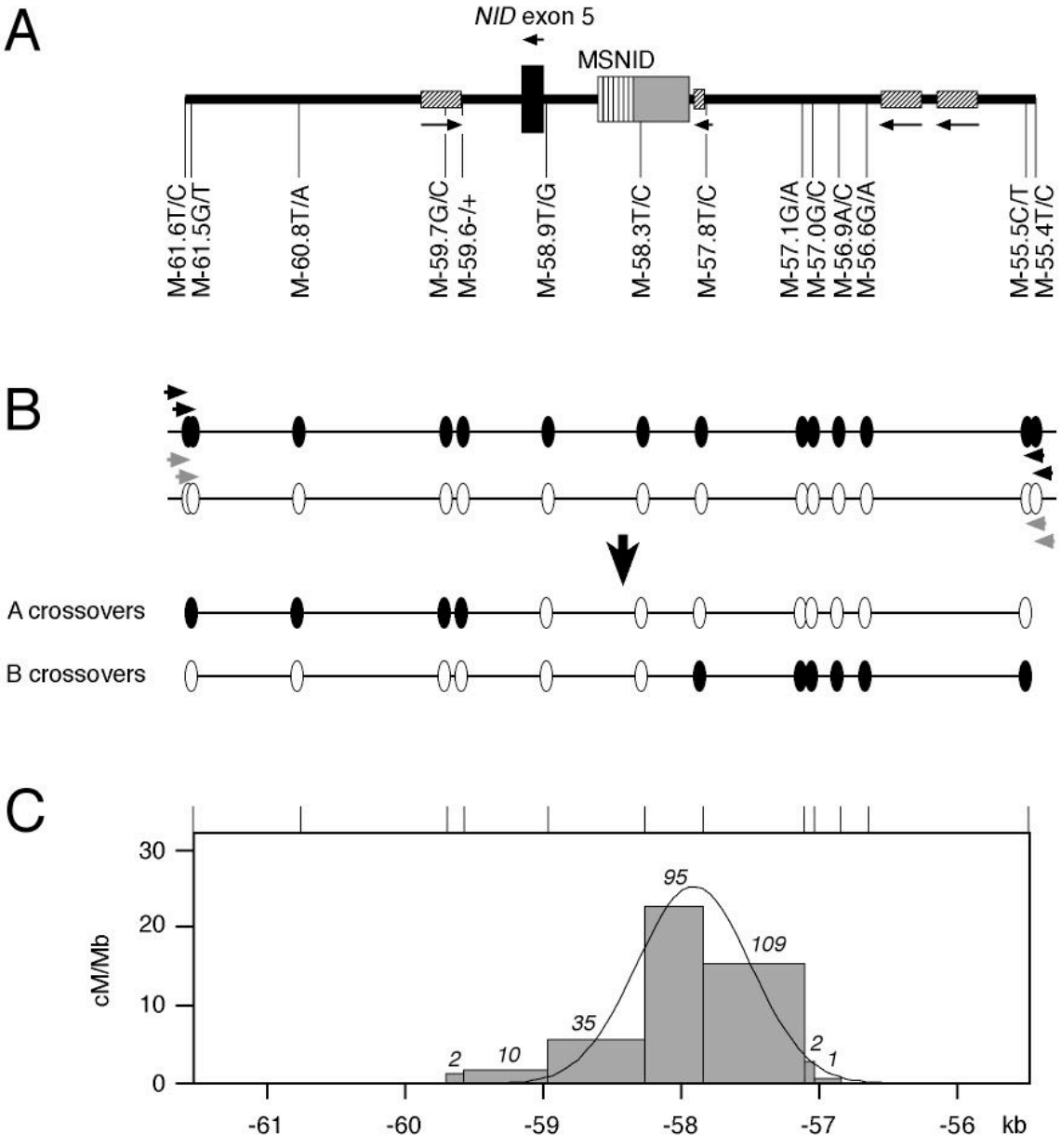

A
NID exon 5
MSNID
M-61.6T/C
M-61.5G/T
M-60.8T/A
M-59.7G/C
M-59.6-/+
M-58.9T/G
M-58.3T/C
M-57.8T/C
M-57.1G/A
M-57.0G/C
M-56.9A/C
M-56.6G/A
M-55.5C/T
M-55.4T/C
B
A crossovers
B crossovers
C
cM/Mb
30
20
10
0
95
109
35
2
10
2
1
-61
-60
-59
-58
-57
-56
kb

Fig. 2

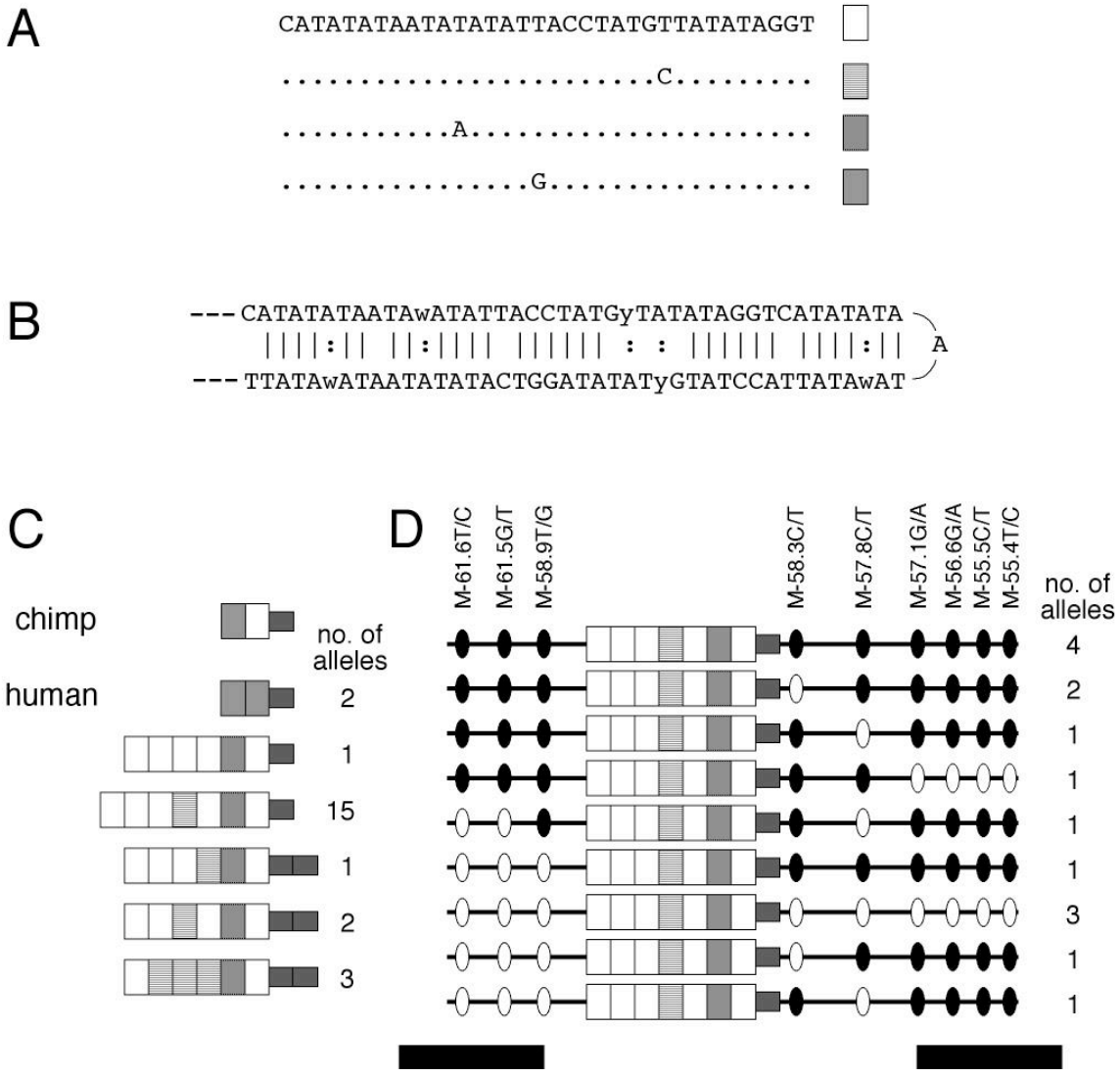

A
CATATATAATATATATTACCTATGTTATATAGGT
.....C.....
.....A.....
.....G.....
B
---CATATATAATAwATATTACCTATGyTATATAGGTCATATATA
A
---TTATAwATAATATATACTGGATATATyGTATCCATTATAwAT
C
chimp
human
no. of alleles
2
1
15
1
2
3
D
M-61.6T/C
M-61.5G/T
M-58.9T/G
M-58.3C/T
M-57.8C/T
M-57.1G/A
M-56.6G/A
M-55.5C/T
M-55.4T/C
no. of alleles
4
2
1
1
1
1
3
1
1

Fig. 3

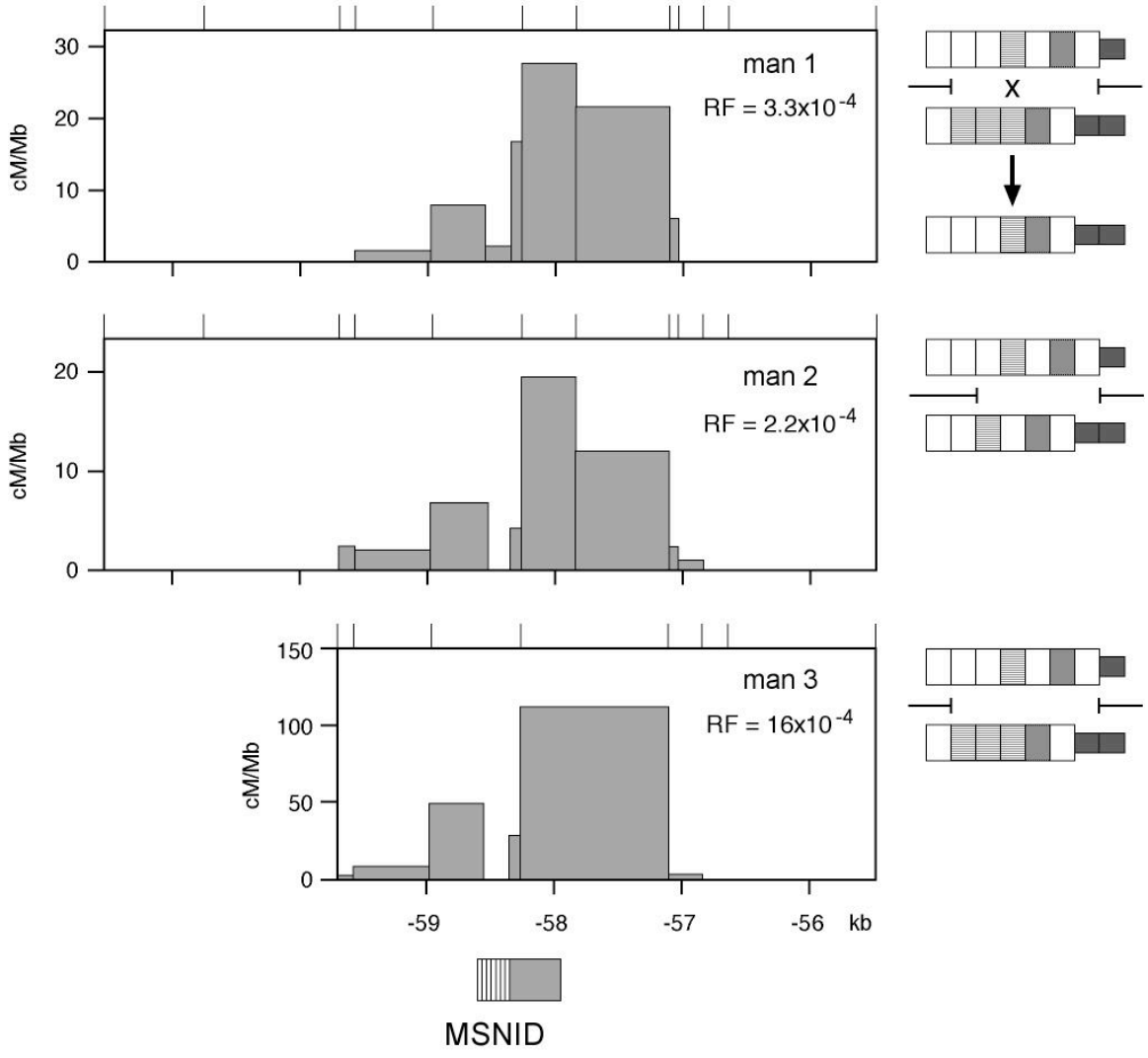

man 1
RF = 3.3x10^{-4}
man 2
RF = 2.2x10^{-4}
man 3
RF = 16x10^{-4}
cM/Mb
30
20
10
0
150
100
50
-59
-58
-57
-56
kb
X
MSNID

Fig. 4

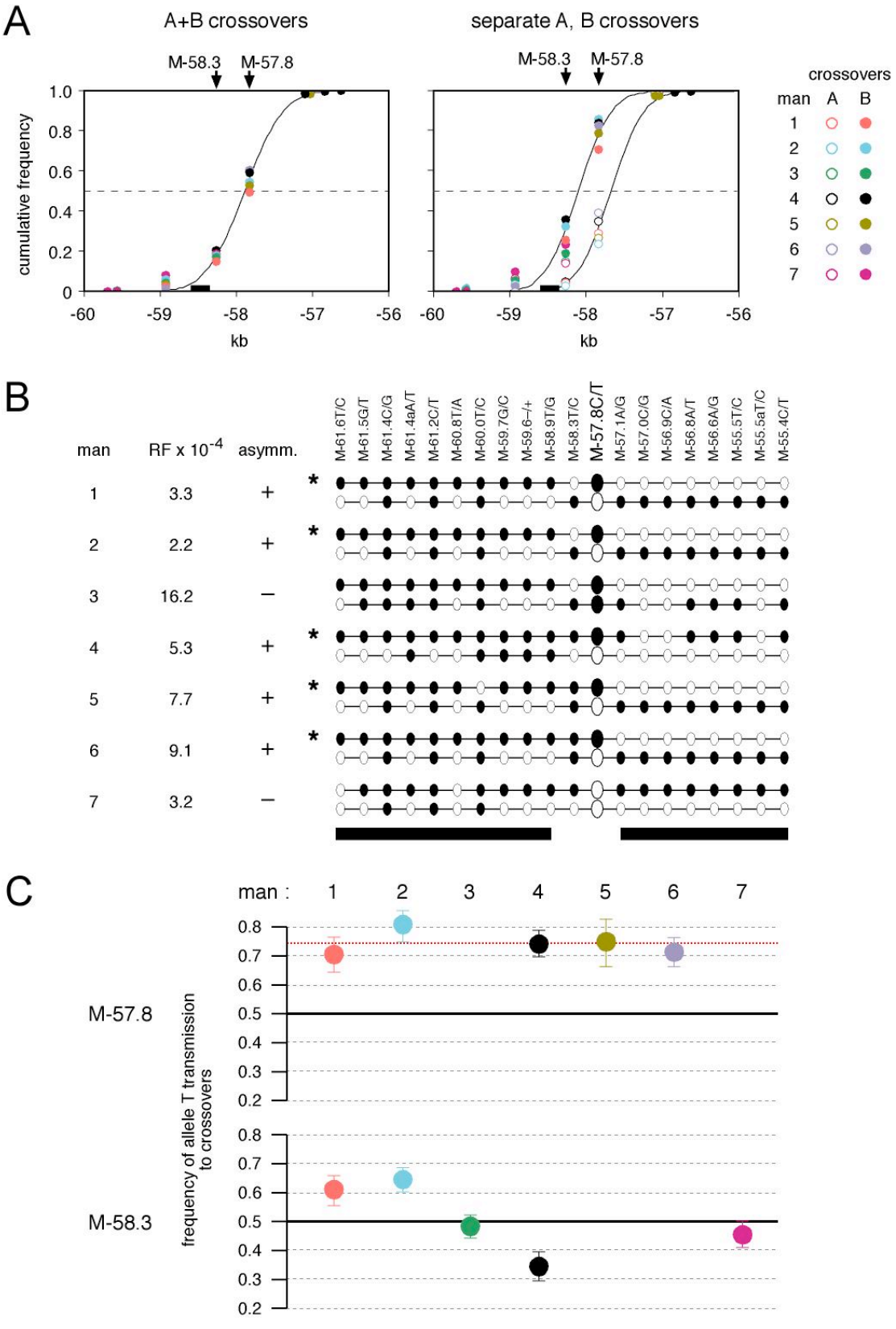
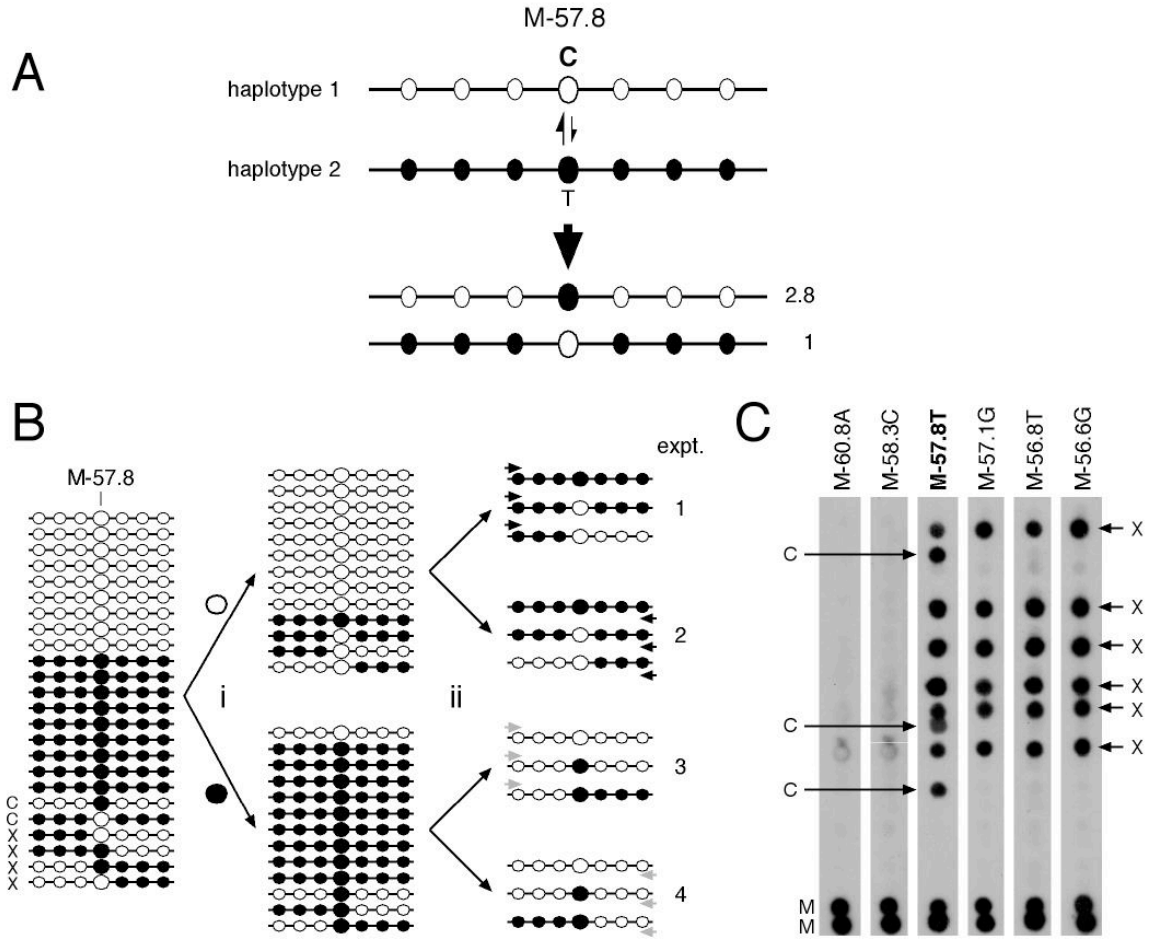

A
A+B crossovers
separate A, B crossovers
M-58.3
M-57.8
cumulative frequency
kb
crossovers
man A B
B
man
RF x 10-4
asymm.
M-61.6T/C
M-61.5G/T
M-61.4C/G
M-61.4aA/T
M-61.2C/T
M-60.8T/A
M-60.0T/C
M-59.7G/C
M-59.6–/+
M-58.9T/G
M-58.3T/C
M-57.8C/T
M-57.1A/G
M-57.0C/G
M-56.9C/A
M-56.8A/T
M-56.6A/G
M-55.5T/C
M-55.5aT/C
M-55.4C/T
C
man :
M-57.8
M-58.3
frequency of allele T transmission to crossovers

Fig. 5



D

expt.	M-57.8 allele selected	molecules screened	no. conversions	no. crossovers	rate x 10^{-4} conversion	rate x 10^{-4} crossover	conversion: crossover
1	C	336,000	27	59	0.9	6.0	0.15
2	C	336,000	21	31	0.7	5.1	0.14
3	T	175,000	31	60	2.1	5.0	0.42
4	T	175,000	21	42	1.4	4.0	0.36

Relative conversion rate T : C = **2.3 : 1** C.I. 1.2-4.2

Relative conversion rate per crossover T : C = **2.7 : 1** C.I. 1.4-5.4