

A major recombination hotspot in the XqYq pseudoautosomal region gives new insight into processing of human gene conversion events

Shriparna Sarbajna^{1,3}, Matthew Denniff^{1,4}, Alec J Jeffreys¹, Rita Neumann¹,
María Soler Artigas², Amelia Veselis¹ & Celia A May^{1,*}

¹ Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

² Department of Health Sciences, University of Leicester, Leicester LE1 7RH, UK

³ Present address: Cancer Research UK, London Research Institute, Clare Hall, South Mimms, EN6 3LD, UK

⁴ Present address: Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK

* Correspondence:

Dr Celia A May

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

email: cam5@le.ac.uk

tel: +44(0) 116 252 3032

fax: +44(0) 116 252 3378

Abstract (234 of 250 limit)

Recombination plays a fundamental role in meiosis. Non-exchange gene conversion (non-crossover, NCO) may facilitate homologue pairing whilst reciprocal crossover (CO) physically connects homologues so they orientate appropriately on the meiotic spindle. In males, X-Y homologous pairing and exchange occurs within the two pseudoautosomal regions (PARs) together comprising <5% of the human sex chromosomes. Successful meiosis depends on an obligatory CO within PAR1, while the nature and role of exchange within PAR2 is unclear. Here we describe the identification and characterization of a typical ~1 kb-wide recombination hotspot within PAR2. We find that both COs and NCOs are strongly modulated in *trans* by the presumed chromatin remodelling protein PRDM9, and in *cis* by a SNP located at the hotspot centre that appears to influence recombination initiation and which causes biased gene conversion in SNP heterozygotes. This, the largest survey to date of human NCOs, reveals for the first time substantial inter-individual variation in the NCO:CO ratio. Although the extent of biased transmission at the central marker in COs is similar across men, it is highly variable amongst NCO recombinants. This suggests that *cis*-effects are mediated not only through recombination initiation frequencies varying between haplotypes but also through subsequent processing, with the potential to significantly intensify meiotic drive of hotspot-suppressing alleles. The NCO:CO ratio and extent of transmission distortion (TD) amongst NCOs appear to be inter-related, suggesting the existence of two NCO pathways in humans.

Introduction

Our understanding of the fundamental process of meiotic recombination in humans has greatly advanced through complementary genome-wide analyses and focused fine-scale sperm DNA approaches. Most crossover events (COs) in our genome cluster into narrow hotspots 1-2 kb in width^{1,2}, and many of these hotspots contain degenerate GC-rich sequence motifs which appear to influence the initiation of recombination³.

Recently, the meiosis-specific protein PRDM9 has been identified as a major specifier of hotspots in humans and mice^{4,11}. PRDM9 contains a zinc-finger (ZnF) domain, encoded by a variable minisatellite that is capable of binding hotspot sequences, plus a SET domain implicated in histone methylation and thus perhaps activation of hotspots via chromatin remodelling. The analysis of *de novo* events in sperm DNA has shown that hotspot activity is highly sensitive to variation in the ZnF array of PRDM9⁵, to the extent that individuals with differing genotypes can use substantially different sets of hotspots^{4,11}. Sperm DNA typing has also revealed *cis*-acting effects within hotspots that down-regulate CO activity^{12,13}, and has been used to explore a second class of meiotic recombination event, highly localised non-reciprocal gene conversions or non-crossovers (NCOs) between homologues¹⁴.

The peak numbers of COs and NCOs at individual human and mouse hotspots co-localise¹⁴⁻¹⁷ suggesting that these events are alternative outcomes of the same recombination-initiating lesions, programmed double strand breaks (DSBs) induced by the endonuclease Spo11 at leptotene¹⁸. However, mouse studies have shown that CO and NCO have different genetic requirements indicating that they likely result from distinct pathways^{19,20}, possibly comparable to those in *Saccharomyces cerevisiae* that diverge shortly after recombination initiation^{21,22}. It is widely held that in the budding

yeast, COs are formed through resolution of double Holliday junction (dHJ) intermediates whilst NCOs are predominantly formed earlier via single-end invasion intermediates by synthesis-dependent strand annealing (SDSA)²³.

Observed as chiasmata during diplotene, the physical linkages that remain between homologues following CO are required for appropriate orientation on the first meiotic spindle and thus correct disjunction of chromosome pairs. Cytological data indicate that only ~10% of mammalian DSBs genome-wide are repaired to generate COs¹⁹, suggesting the remainder are repaired as NCOs. It has been proposed that these homology-dependent interactions might directly enforce proper homologous pairing²⁴⁻²⁶ and/or be the indirect consequence of induction of sufficient DSBs to ensure properly regulated placement of COs¹⁷. To date, observed relative rates of CO and NCO show major variation between human hotspots¹⁵ yet we have little insight into the factors that influence this balance at the hotspot level either in mouse or man.

The critical role of CO is evident from the behaviour of the sex chromosomes during male meiosis. In humans, normal pairing and CO between the X and Y is mainly limited to the 2.7 Mb region of shared homology at the tips of short arms, the major pseudoautosomal region or PAR1^{27,28}. Family data are consistent with a single obligatory PAR1-mediated CO per male meiosis²⁹, and diminished recombination in PAR1 has been associated with an increased frequency of 24, XY aneuploid sperm (*i.e.* a single copy of each autosome, but both an X and a Y chromosome)^{30,31}. The human sex chromosomes also harbour a second much smaller (~330-kb) pseudoautosomal region, PAR2, at the ends of the long arms of X and Y³². PAR2 is human-specific, having resulted from an L1-mediated ectopic recombination event that transferred the subtelomeric region of the X onto the Y chromosome after the human and chimpanzee

lineages diverged³³⁻³⁵. Cytological studies have shown that ~50% of early pachytene spermatocytes synapse at PAR2, but curiously these associations disappear by mid-pachytene such that chiasmata have only been observed at PAR1³⁶⁻³⁹. This suggests that the two human PARs are not functionally equivalent and raises questions about the nature of genetic exchanges in PAR2 – are they genuine CO events or atypically long non-reciprocal NCOs³³? Very low-resolution molecular studies have identified a few *de novo* paternal exchanges within families, at least some of which are consistent with reciprocal CO^{40,41}. However, as noted by Flaquer and colleagues^{42,43}, PAR2 has been largely neglected in recent marker-dense linkage studies. To address this, we have used population diversity data from the International HapMap project⁴⁴ and sperm recombination assays⁴⁵ to examine genetic exchange in PAR2 at high resolution. We demonstrate that exchange is dominated by a single highly-active hotspot previously shown on a subset of data to be regulated by PRDM9⁵. This hotspot has proved to be highly informative for analysing recombination events and for the first time has allowed us to explore in detail inter-individual variation in the ratio of COs to NCOs.

Results

Crossover activity in PAR2: identification of the SPRY3 hotspot

Published historical CO rates for PAR2, estimated by coalescent analysis of CEU HapMap data^{44,46}, showed major clustering of COs with two clear peaks of activity in the first 100 kb of PAR2, the more extreme coinciding with the 5' region of the most proximal gene *SPRY3* some 66 kb from the PAR2 boundary^{34,35} (Figure 1A). All exchange events previously identified in CEPH families mapped to a 145-kb interval entirely encompassing these two peaks of historical recombination^{40,41} (Figure 1B). We attempted to refine the mapping of these COs in five of these six key CEPH families by SNP typing. In one family we could not identify a recombinant child, while for four families we could map the events to short intervals (0.9–30.7 kb) that all overlapped each other and the extreme peak in historical recombination (Figure 1B).

We homed in on a 15-kb interval centred on the candidate hotspot upstream of *SPRY3* and genotyped 97 North European semen donors for all available SNP markers (Supplementary Tables 1 & 2). Linkage disequilibrium (LD) mapping⁴⁷ confirmed a single interval of LD breakdown ~2-kb wide (Figure 1C). A similar LD profile was seen in 74 African semen donors, suggesting that this LD hotspot has been active in both African and non-African populations. Corresponding analysis of the second more distal peak of activity (Figure 1A) failed to identify an LD hotspot (CAM, unpublished data). Also, this peak disappeared when using coalescent analysis on HapMap genotype data filtered for markers unambiguously typed in >70% of individuals, suggesting that this HapMap peak is an artefact.

We designed long-range PCR assays spanning the major LD hotspot to detect *de novo* CO molecules in sperm DNA⁴⁵ (Supplementary Table 2). Seventeen men (9 north Europeans and 8 Africans) were screened, and 7–587 COs were identified per man (median 94). Sperm CO frequencies showed ~50-fold variation amongst these semen donors (0.017% to 0.845%), with the highest activity comparable to that of the most active autosomal hotspots described to date⁴⁸. Reciprocal assays performed on seven of the men showed no significant difference in CO frequency between orientations as expected for classical reciprocal CO.

All CO events were mapped by typing intervening SNPs. More than 99% of recombinants involved a single switch from one haplotype to the other. These simple exchange points were clustered into a single 1 kb–wide recombination hotspot in both African and European men, with CO breakpoints apparently normally distributed across the hotspot (Figure 1D).

CO asymmetry triggered by heterozygosity at a hotspot SNP

SNP rs700442 (hereafter known as “0442”) is the most central marker in the hotspot, located ~45 bp away from the centre⁵. All 0442C/T heterozygotes assayed in both orientations showed highly significant ($P = 1 \times 10^{-7}$ – 5.5×10^{-11}) over-transmission of the C allele to recombinant sperm, with on average 66% of COs inheriting the C allele (Figure 2A). Similar though less intense transmission distortion was also seen at other heterozygous SNPs within the hotspot in two of these men. In contrast, 0442C and 0442T homozygotes showed no significant transmission distortion at any hotspot SNP (Figure 2B), suggesting that this biased gene conversion is specifically triggered by

heterozygosity at 0442. As a consequence of distortion, reciprocal events in 0442 heterozygotes map to different locations and the offset of the centre points gives an indication of the average conversion tract associated with CO¹⁹. A mean displacement of 329 bp was seen at the *SPRY3* hotspot, very similar to ~350 bp shifts seen at autosomal hotspots displaying CO asymmetry^{12,13,48,49}.

CO asymmetry is most readily explained by differences in the frequency of recombination-initiating DSBs between haplotypes, with 0442C marking the less active haplotype¹². Gene conversion accompanying CO ensues as the overhangs generated by subsequent resection are returned to the double-stranded state using information from the other intact chromosome. This implies that 0442T-bearing haplotypes experience more DSBs than 0442C haplotypes and consistent with this, 0442T homozygotes tend to have higher overall CO frequencies than 0442C homozygotes amongst the 17 men analysed (on average ~3-fold higher).

SNP 0442 falls within a near-perfect match to the CCNCCNTNNCCNC motif found in 40% of European LD hotspots³ that likely serves as a binding site for the common “A” form of PRDM9. Consistent with this, the PRDM9 A variant specifically activates this hotspot, as previously shown on a subset of the present data (hotspot “PAR2” in ref. 5), and the derived recombination-suppressing allele 0442C disrupts the motif (Figure 2C).

High levels of inter-individual variation in CO frequency

PRDM9 and SNP 0442 have major influences on *SPRY3* hotspot activity but cannot explain all the variation in activity between men. For example, the seven men tested

who were homozygous for the *PRDM9* A variant and heterozygous at 0442 showed ~100-fold variation in CO frequency (Supplementary Table 3). We quantified the combined effects of the *PRDM9* ZnF array and 0442 status on CO frequency using a linear regression model on data from all 29 men analysed. Assuming an additive effect for each, the proportion of the variance explained by these two factors was found to be 29%. This increased modestly to 34% when a term was included to model the biological interaction between these two co-variables. In relative terms, the *trans*-acting factor *PRDM9* was found to account for 1.4 times more of the variance in CO frequency than the *cis*-acting SNP 0442. We did attempt to identify other factors that might be influencing CO frequencies at the *SPRY3* hotspot, including additional variants within the hotspot as well as *trans*-acting loci reported to influence overall crossover frequencies in the human genome^{50,51} (Supplementary Tables 3-5, Supplemental Note). However, once we took *PRDM9* and 0442 status into account, we lacked statistical power to detect or exclude influences of these other candidate factors on CO frequency.

NCO events are also prevalent and regulated by PRDM9

To look at interhomologue recombination without exchange, we designed assays capable of detecting NCO as well as CO molecules in sperm DNA⁴⁵. Analysis of two 0442 heterozygotes showed comparable CO frequencies and the same transmission distortion (TD) as seen in regular CO assays, thus validating the approach. Data on 14 men (7 north Europeans and 7 Africans) revealed NCO frequencies ranging from 0.01 to 0.49%, with ~85% of events encompassing only a single SNP site. True NCO frequencies will be higher since markerless conversion tracts cannot be detected.

Indeed mean tract lengths estimated across all men analysed were heavily affected by differences in marker distribution (see Supplementary Table 6). Nonetheless, detectable conversions appear to have short tract lengths (shortest 1–99 bp in length, longest 613–2910 bp) and relatively steep gradients extending from the hotspot centre, and are likely consistent with those seen at other human hotspots¹⁴.

For direct comparisons between men, we focussed on NCO frequencies at the central SNP 0442 in thirteen 0442 heterozygotes (366 NCO events detected in total). NCO frequencies ranged from 0–0.35% and, as seen for COs at this hotspot, appear to be regulated by PRDM9, with the highest activities seen in men homozygous for the activating PRDM9 A variant (Figure 3A). As with COs, NCOs showed significantly biased gene conversion, with most NCOs showing transfer of the 0442C allele (plus linked hotspot alleles) to the 0442T haplotype and not *vice versa* (Figure 3B). These findings are consistent with the PRDM9 A variant acting upstream of the CO/NCO decision, with both types of recombinant being different outcomes of the same initiating events. However, there was only a marginally significant correlation between NCO and CO frequencies, with NCO:CO ratios varying from ~1:10 to 3.8:1 between men and with no evidence that PRDM9 influences the NCO/CO decision (Figure 4).

Qualitative as well as quantitative NCO differences between men

Not only did the NCO:CO ratio differ between men but also the level of biased gene conversion seen at SNP 0442 in COs vs. NCOs, with COs showing a fairly constant 2:1 C:T bias across all men but with NCOs showing biases ranging from ~1:1 to >85:1 (Figure 5A). Curiously, men with the highest NCO:CO ratios tended to show the

strongest conversion bias specifically in NCOs (Figure 5B). For example, man 3 showed similar 0442 bias in NCOs and COs and approximately 2.5 times fewer NCOs than COs, whilst man 22 showed extreme 0442 bias in NCOs and around 3.5 times more NCOs than COs. Comparison of conversion tracts in these two men revealed further differences: 23% of NCOs (13/57) had co-converted at SNP rs56750753 (“753”) in man 3, versus only 4% (3/78) in man 22 (Figure 5C, Fisher Exact test, $P = 0.001$). Interestingly, co-conversion at SNPs 0442 and 753 implies a tract length of at least 313 bp, very similar to mean conversion tract lengths seen in COs (see above). Despite these differences between NCOs, the distribution of COs between these two men was nonetheless comparable (Fisher Exact test, $P = 0.792$).

Discussion

We describe the first in-depth analysis of recombination in the minor human pseudoautosomal region, PAR2, more than 20 years after its discovery³². European population diversity data together with high-resolution mapping of familial recombinants showed that most CO events in PAR2 co-localize to an interval encompassing the promoter region of the most proximal gene, *SPRY3*. Sperm DNA typing, on more than 1 million haploid genome equivalents from 29 men, resulted in characterization of 3000 recombinant gametes and localised recombination to a ~1kb-wide hotspot active in both CO and NCO. The tight clustering of both types of event is similar to that seen at autosomal hotspots^{1,48,52} and at a hotspot characterized in PAR1⁵³. These findings allay early concerns that conventional COs might not occur in this human-specific pseudoautosomal region³³. The activity of the *SPRY3* recombination hotspot is similar to that of the hottest human hotspots, but nonetheless, the maximum CO frequency seen in this study (0.8%) is entirely compatible with the more limited cytological studies of the XY bivalent that have failed to reveal definitive signs of CO such as chiasmata or MLH1 foci in PAR2.

Kauppi *et al.* (2011) have recently shown that the appearance of RAD51/DMC1 foci in the mouse pseudoautosomal region, which mark the DSBs that mediate pairing and successful exchange, are temporally delayed compared with the rest of the genome, possibly being predominantly induced by Spo11 α rather than the conventional Spo11 β isoform⁵⁴. They speculated that this conserved splice form might also be responsible for DSBs in the human major pseudoautosomal region, PAR1. Cytological studies have shown that human X-Y pairing is initiated at PAR1, and that it can go beyond the region of homology and even lead to pairing at PAR2³⁶⁻³⁹. It is therefore possible that

DSBs within PAR2 might also be induced predominantly by Spo11 α rather than Spo11 β . Our data cannot address this issue but do suggest that recombination in PAR2 is subject to the same types of *cis*- and *trans*-acting factors as the autosomes. For example, both CO and NCO frequencies at the *SPRY3* hotspot are strongly influenced by *PRDM9* status, suggesting similar hotspot activation mechanisms (possibly chromatin remodelling) in PAR2 and the rest of the genome. In addition, we have shown that recombination at the *SPRY3* hotspot is also modulated in *cis* by a motif-disrupting SNP that appears to down-regulate CO frequency. The scale of this suppression, and the resulting displacement in the locations of reciprocal crossovers, is very similar to that seen at autosomal hotspots, strongly suggesting that downstream processing of PAR2 and autosomal DSBs is likely to be the same too, at least for COs.

Motif-disrupting variants like 0442C are systematically over-transmitted to crossover progeny at other human hotspots too^{12,13,49}. This constitutes a form of meiotic drive that is predicted to lead to the demise of the hotspots as the recombination-suppressing variants become fixed in the population⁵⁵. While the degree of overtransmission to COs at the *SPRY3* hotspot is relatively modest (66:34 in favour of the motif-disrupting allele), it is significantly enhanced to 74:26 by the stronger bias seen in NCOs. The very high activity of this hotspot predicts a 0442 C:T gametic ratio of 50.13 : 49.87 and thus the most extreme level of meiotic drive seen to date at a human hotspot (Table 1). Population simulations indicate that the recombination-suppressing C allele is not only guaranteed to go to fixation ($P > 0.999$), leading to hotspot attenuation, but will do so rapidly, within $\sim 60,000 \pm 16,000$ years, about 10% the time required for an undriven allele. Thus NCOs can play a significant role in determining the lifespan of a recombination hotspot¹⁷.

This study provides the largest survey to date of NCOs at a human recombination hotspot. The ratio of NCOs to COs at *SPRY3* is surprisingly variable, and is comparable to that seen between other human hotspots surveyed in just one or two men. Until now, these differences in ratio between hotspots have been interpreted as different hotspot-specific biases in resolution of recombination-initiation events¹⁵. However, it is possible that this apparent variation between hotspots reflects variation between men at each hotspot, as seen at *SPRY3*. This can be tested by more extensive surveys of NCO activity at autosomal hotspots.

If variable NCO:CO ratios prove to be unique to the *SPRY3* hotspot, then this would provide evidence for a PAR2-specific process of recombination. One possibility is that successful reciprocal exchange in PAR1 on the short arms of X and Y confers strong CO interference and diverts the repair of DSBs in PAR2 down a NCO route. If so, then men who exhibit more COs than NCOs at *SPRY3* may be less proficient at exchange within PAR1 than those that show the opposite trend at this hotspot. Thus, the balance of CO and NCO events at *SPRY3* for a given man might be related to successful reciprocal exchange in the major pseudoautosomal region. Unfortunately our existing sperm data from PAR1 do not allow us to test this as they are on limited numbers of men who do not overlap the men tested at *SPRY3*⁵³.

Variation in the NCO:CO ratio at *SPRY3* appears to be associated with differences in the degree of TD at SNP 0442, particularly in NCOs. This could reflect both quantitative and qualitative variation in the biochemical behaviour of the NCO pathway, or alternatively indicate that NCOs are produced by two distinct pathways. Since there is no significant difference in TD between men amongst COs, we assume that this distortion most closely reflects the initiation bias caused by the different 0442

SNP alleles. Men with low NCO:CO ratios show similar levels of bias in NCOs and COs, plus relatively long conversion tracts in NCOs apparently comparable to those seen in COs. These data are compatible with COs and NCOs in these men being produced from the same recombination intermediate, for example by alternative resolution of dHJs mediated by different resolvase complexes⁵⁶. Both types of event would then have to be subject to the same heteroduplex repair bias to generate the same level of TD seen in COs and NCOs. In contrast, men with a high NCO:CO ratio show greater transmission bias and shorter conversion tracts in NCOs but not apparently in COs. This might reflect a predominant and distinct NCO pathway that is variable in activity between men and differentially repairs heteroduplex DNA compared with the CO pathway, leading to greater TD. It is tempting to speculate that the SDSA pathway²³ might fulfill this role. Interestingly, a recent genome-wide analysis of recombination intermediates challenged current thinking by indicating that at least two separate pathways are frequently used to generate NCO events in budding yeast⁵⁷.

We have presented data that for the first time implicate differential processing in the formation of NCOs in humans. The factors that influence the frequency and nature of such events have yet to be elucidated, but may come to light when NCOs are similarly investigated in detail at other human hotspots.

Materials and Methods

DNA samples. North European semen samples were collected with informed consent and approval from Leicestershire Health Authority Research Ethics Committee.

Collection of African samples is detailed in ref 58. DNAs were extracted and handled to minimize contamination risk⁴⁵, and concentrations quantified on a NanoDrop 1000 spectrophotometer. Aliquots (40ng) of each DNA were whole-genome-amplified using the GenomiPhi HY DNA amplification kit (GE Healthcare Bio-Sciences) and the resulting products used for routine genotyping. Recombination assays were performed on high molecular weight total genomic DNA.

SNP genotyping. Genotypes were established by ASO hybridization against dot blots of suitable PCR products⁴⁵. The PAR2 markers typed in this study are listed in Supplementary Table 1 and details of the primers and PCR conditions are listed in Supplementary Table 2.

Population analysis. Historical recombination activity was estimated by coalescent-based analysis of genotype data using LDhat⁴⁶. Historical hotspot locations were inferred from metric LD maps generated using LDMAP⁴⁷.

Sperm CO assays. Semen donors with suitable SNP heterozygosities flanking the region of LD breakdown were chosen for analysis. Allele-specific primers for rs5940571, rs17653343 and rs9645286 were designed and initially tested on DNA from men homozygous for one or the other SNP allele at various annealing temperatures to achieve maximum discrimination and PCR efficiency. For each man, linkage phases were established by ASO hybridization against allele-specific PCR products. CO molecules were selectively amplified by repulsion-phase allele-specific long PCR (Supplementary Table 2). Specifically, for a given man and orientation of COs, between 94 and 190 PCR reactions each containing typically 2.5ng of DNA were amplified using the buffer described previously⁵⁹ supplemented with 12mM Tris base,

0.2μM of each primer, 0.03U/μl *Taq* polymerase (KAPA Biosystems) and 0.0015U/μl cloned *Pfu* polymerase (Agilent Technologies). Secondary allele-specific PCRs were seeded with 1/86th vol primary PCR products, and CO-positive PCRs identified by agarose gel electrophoresis. CO breakpoints were mapped by reamplifying the secondary PCR products using internal universal primers and typing the resulting products by ASO hybridization. The numbers of each type of CO molecule observed were Poisson-adjusted to account for PCR reactions containing more than one CO, and a single-molecule PCR efficiency of 50% (one amplifiable molecule of each haplotype per 12pg sperm DNA) was assumed when calculating CO frequencies⁴⁵.

Detecting sperm NCOs. For a given man, molecules derived from one parental haplotype were amplified from between 187 and 457 pools of sperm DNA containing typically 38 amplifiable molecules each (range ~20-50), using two rounds of PCR with allele-specific primers directed to SNPs on one side of the hotspot in conjunction with universal primers on the other side of the hotspot (Supplementary Table 2). COs and NCOs were simultaneously detected against the background of non-recombinant molecules by ASO hybridization using probes corresponding to the opposite parental haplotype from that selectively amplified in the PCR reactions. Single-molecule PCR efficiencies and Poisson corrections were as for COs.

Assigning proportion of variance in recombination activity to PRDM9 and 0442.

Linear regression was undertaken on inverse-normal transformed CO frequencies using *PRDM9* and 0442 status as covariates. An additive effect was assumed for each *PRDM9* A allele, and similarly for each 0442C allele.

Hotspot sequencing. Separated haplotypes generated using allele-specific primers (see Supplementary Table 2) from each of the men in this study were used in standard BigDye Terminator v3.1 Cycle Sequencing reactions (Applied Biosystems) with primers 5' GGCTGAGTAGATTGGTAGA 3' and 5' AGTGGCGTCATCCAGATGA

3'. Extension products were purified using Performa Gel Filtration columns (EdgeBiosystems) and run on a 3730 DNA Analyzer (Applied Biosystems). Potential additional *cis*-acting effects were evaluated by rank order analysis controlling for *PRDM9* and 0442 status.

Meiotic drive simulations. Population simulations were carried out as described previously¹² taking into account the biased transmission of 0442 in both COs and NCOs averaged over five *PRDM9* A/A homozygotes. The chances of fixation of the favoured allele and mean time to fixation were estimated assuming a starting population frequency of 50%, an effective human population size N_e of 10,000 and a generation time of 20 years. The same parameters were also estimated assuming no drive of the favoured allele.

Acknowledgements

We thank J.Wetton and colleagues for helpful discussions. This work was supported by the Medical Research Council [grant number G0601068].

Conflict of Interest Statement

None declared.

References

1. 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.
2. Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science*, **310**, 321-324.
3. Myers, S., Freeman, C., Auton, A., Donnelly, P., and McVean, G. (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.*, **40**, 1124-1129.
4. Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, **327**, 836-840.

5. Berg, I.L., Neumann, R., Lam, K-W., Sarbajna, S., Odenthal-Hesse, L., May, C.A., and Jeffreys, A.J. (2010) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.*, **42**, 859-863.
6. Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G., and Donnelly, P. (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, **327**, 876-879.
7. Parvanov, E.D., Petkov, P.M., and Paigen, K. (2010) Prdm9 controls activation of mammalian recombination hotspots. *Science*, **327**, 835.
8. Paigen, K., and Petkov, P.M. (2010) Mammalian recombination hot spots: Properties, control and evolution. *Nat. Rev. Genet.*, **11**, 221-233.
9. Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G.B., Jonasdottir, A., Gylfason, A., Kristinsson, K.T. *et al.* (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, **467**, 1099-1103.
10. Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., Akylbekova, E.L. *et al.* (2011) The landscape of recombination in African Americans. *Nature*, **476**, 170-175.
11. Berg, I.L., Neumann, R., Sarbajna, S., Odenthal-Hesse, L., Butler, N.J., and Jeffreys, A.J. (2011) Variants of the protein PRDM9 differentially regulate a set of

human meiotic recombination hotspots highly active in African populations. *Proc. Natl. Acad. Sci. USA*, **108**, 12378-12383.

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

13. Jeffreys, A.J., and Neumann, R. (2005) Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.*, **14**, 2277-2287.

14. 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.

15. Holloway, K., Lawson, V.E., and Jeffreys, A.J. (2006) Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. *Hum. Mol. Genet.*, **15**, 1099-1111.

16. Guillon, H., and de Massy, B. (2002) An initiation site for meiotic crossing-over and gene conversion in the mouse. *Nat. Genet.*, **32**, 296-299.

17. Cole, F., Keeney, S., and Jason, M. (2010) Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. *Mol. Cell*, **39**, 700-710.

18. Keeney, S. (2001) Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.*, **52**, 1-53.

19. Guillon, H., Baudat, F., Grey, C., Liskay, R.M., and de Massy, B. (2005) Crossover and noncrossover pathways in mouse meiosis. *Mol. Cell*, **20**, 563-573.
20. Baudat, F., and de Massy, B. (2007) Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Res.*, **15**, 565-577.
21. Allers, T., and Lichten, M. (2001) Intermediates of yeast meiotic recombination contain heteroduplex DNA. *Mol. Cell* **8**, 225-231.
22. Hunter, N., and Kleckner, N. (2001) The single-end invasion: an asymmetric intermediate at the double-strand break to double-Holliday junction transition of meiotic recombination. *Cell*, **106**, 59-70.
23. Pâques, F., and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.*, **63**, 349-404.
24. Carpenter, A.T.C. (1987) Gene conversion, recombination nodules, and the initiation of meiotic synapsis. *Bioessays*, **6**, 232-236.
25. Zickler, D., and Kleckner, N. (1998) The leptotene-zygotene transition of meiosis. *Ann. Rev. Genet.*, **32**, 619-697.

26. Goldman, A.S., and Lichten, M. (2000) Restriction of ectopic recombination by interhomolog interactions during *Saccharomyces cerevisiae* meiosis. *Proc. Natl. Acad. Sci. USA*, **97**, 9537-9542.
27. Cooke, H.J., Brown, W.R.A., and Rappold, G.A. (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature*, **317**, 687-692.
28. Simmler, M.C., Rouyer, F., Vergnaud, G., Nystrom-Lahti, M., Ngo, K.Y., de la Chapelle, A., and Weissenbach, J. (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature*, **317**, 692-697.
29. Rouyer, F., Simmler, M.C., Johnsson, C., Vergnaud, G., Cooke, H.J., and Weissenbach, J. (1986) A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature*, **319**, 291-295.
30. Hassold, T.J., Sherman, S.L., Pettay, D., Page, D.C., and Jacobs, P.A. (1991) XY chromosome nondisjunction in man is associated with diminished recombination in the pseudoautosomal region. *Am. J. Hum. Genet.*, **49**, 253-260.
31. Shi, Q., Spriggs, E., Field, L.L., Ko, E., Barclay, L., and Martin, R.H. (2001) Single sperm typing demonstrates that reduced recombination is associated with the production of aneuploid 24, XY human sperm. *Am. J. Med. Genet.*, **99**, 34-38.

32. Bickmore, W., and Cooke, H.J. (1987) Evolution of homologous sequences on the human X and Y chromosomes, outside of the meiotic pairing segment. *Nucleic Acids Res.*, **15**, 6261-6271.
33. Kvaløy, K., Galagni, F., and Brown, W.R.A. (1994) The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. *Hum. Mol. Genet.*, **3**, 771-778.
34. Ciccodicola, A., D'Esposito, M., Esposito, T., Gianfrancesco, F., Migliaccio, C., Miano, M.G., Matarazzo, M.R., Vacca, M., Franzè, A., Cuccurese, M. *et al.* (2000) Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. *Hum. Mol. Genet.*, **9**, 395-401.
35. Charchar, F.J., Svartman, M., El-Mogharbel, N., Ventura, M., Kirby, P., Matarazzo, M.R., Ciccodicola, A., Rocchi, M., D'Esposito, M., and Graves, J.A. (2003) Complex events in the evolution of the human pseudoautosomal region 2 (PAR2) *Genome Res.*, **13**, 281-286.
36. Chandley, A.C., Goetz, P., Hargreave, T.B., Joseph, A.M., and Speed, R.M. (1984) On the nature and extent of XY pairing at meiotic prophase in man. *Cytogenet. Cell. Genet.*, **38**, 241-247.
37. Chandley, A.C., Hargreave, T.B., McBeath, S., Mitchell, A.R., and Speed, R.M. (1987) Ring XY bivalent, a new phenomenon at metaphase I of meiosis in man. *J. Med. Genet.*, **24**, 101-106.

38. Speed, R.M., and Chandley, A.C. (1990) Prophase of meiosis in human spermatocytes analysed by EM microspreading in infertile men and their controls and comparisons with human oocytes. *Hum. Genet.*, **84**, 547-554.
39. Armstrong, S.J., Kirkham, A.J., and Hultén, M.A. (1994) XY chromosome behaviour in the germ-line of the human male, a FISH analysis of spatial orientation, chromatin condensation and pairing. *Chromosome Res.*, **2**, 445-452.
40. Freije, D., Helms, C., Watson, M.S., and Donis-Keller, H. (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science*, **258**, 1784-1787.
41. Li, L., and Hamer, D.H. (1995) Recombination and allelic association in the Xq/Yq homology region. *Hum. Mol. Genet.*, **4**, 2013-2016.
42. Flaquer, A., Rappold, G.A., Wienker, T.F., and Fischer, C. (2008) The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur. J. Hum. Genet.*, **16**, 771-779.
43. Flaquer, A., Fischer, C., and Wienker, T.F. (2009) A new sex-specific map of the human pseudoautosomal regions (PAR1 and PAR2). *Hum. Hered.*, **68**, 192-200.
44. International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851-862.

45. Kauppi, L., May, C.A., and Jeffreys, A.J. (2009) Analysis of meiotic recombination products from human sperm. *Methods Mol. Biol.*, **557**, 323-355.
46. 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.
47. Maniatis, N., Collins, A., Xu, C.F., McCarthy, L.C., Hewett, D.R., Tapper, W., Ennis, S., Ke, X., and Morton, N.E. (2002) The first linkage disequilibrium (LD) Maps:delineation of hot and cold blocks by diplotype analysis. *Proc. Natl. Acad. Sci. USA*, **99**, 2228-2233.
48. Webb, A.J., Berg, I.L., and Jeffreys, A. (2008) Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proc. Natl. Acad. Sci. USA*, **105**, 10471-10476.
49. Jeffreys, A.J., and Neumann, R. (2009) The rise and fall of a human recombination hotspot. *Nat. Genet.*, **41**, 625-629.
50. Kong, A., Thorleifsson, G., Stefansson, H., Masson, G., Helgason, A., Gudbjartsson, D.F., Jonsdottir, G.M., Gudjonsson, S.A., Sverrisson, S., Thorlacius, T. *et al.* (2008) Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science*, **319**, 1398-1401.

51. Chowdhury, R., Bois, P.R.J., Feingold, E., Sherman, S.L., and Cheung, V.G. (2009) Genetic analysis of variation in human meiotic recombination. *PLoS Genet.*, **5**, e1000648.
52. Jeffreys, A.J., Neumann, R., Panayi, M., Myers, S., and Donnelly, P. (2005) Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.*, **37**, 601-606.
53. 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.
54. Kauppi, L., Barchi, M., Baudat, F., Romanienko, P.J., Keeney, S., and Jasin, M. (2011) Distinct properties of the XY pseudoautosomal region crucial for male meiosis. *Science*, **331**, 916-920.
55. Coop, G., and Myers, S.R. (2007) Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.*, **3**, e35.
56. Svendsen, J.M., and Harper, J.W. (2010) GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. *Genes Dev.*, **24**, 521-536.
57. Martini, E., Borde, V., Legendre, M., Audic, S., Regault, B., Soubigou, G., Dujon, B. and Llorente, B. (2011) Genome-wide analysis of heteroduplex DNA in mismatch repair-deficient yeast cells reveals novel properties of meiotic recombination pathways. *PLoS Genet.*, **9**, e1002305.

58. Monckton, D.G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A., and Jeffreys, A.J. (1994) Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nat. Genet.*, **8**, 162-170.

59. Jeffreys, A.J., Wilson, V., Neumann, R., and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction, towards DNA fingerprinting of single cells. *Nucleic Acids Res.*, **16**, 10953-10971.

Legends to Figures

Figure 1. Recombination activity in PAR2 inferred from population and pedigree data.

(A) Published historical recombination map of PAR2 as estimated by coalescent analysis of CEU genotypes from Phase II HapMap, with peaks of historical activity arrowed. This map does not cover the proximal and distal regions of PAR2 (shaded). Coding sequences and the PAR2 boundary PAB2 are indicated below. (B) Mapping *de novo* recombination events identified in CEPH pedigrees, with individual identifiers to the right. All four events had been previously mapped to the interval between markers JXYQ and sDF-1^{40,41} shown by thin lines between triangles. Further mapping localized the events to the regions marked by black boxes. Partial pedigree data for the two key markers for family 1332 are shown beneath; the recombinant child 133206 (asterisk) receives his father's Y-allele at rs700442 but his father's X-allele at rs71190340.

[Note: despite a previous report⁴¹ we were unable to identify a recombinant child within CEPH pedigree 1416]. (C) Metric LD maps over a 15-kb interval encompassing the 5' end of *SPRY3*, established from high-resolution SNP genotype data from 97 men of North European descent (filled circles) and 74 men of Southern East-African descent (open circles). Co-ordinates are given for the X chromosome (NCBI36/hg18). (D) Cumulative frequency distributions of sperm COs from seven men are shown for the assay interval (0 kb= chrX:154,644,494), with filled symbols representing Europeans and open symbols Africans. The least-squares best-fit cumulative normal distribution is shown by the black curve. Note that despite an apparent double LD hotspot in Africans (Figure 1C), there is only one active hotspot in the African men tested over this interval.

Figure 2. Transmission distortion at the hotspot centre is associated with heterozygosity at SNP rs700442. (A) Transmission frequencies, with 95% confidence intervals, of SNP 0442 allele C and linked alleles into normalised numbers of reciprocal sperm COs. Data are shown for two 0442 heterozygotes each informative for an additional marker within the hotspot (indicated by grey shading). Significant transmission distortion is seen for all hotspot markers. (B) Hotspot markers carried by men homozygous for 0442 show no significant deviation from 50% transmission regardless of their location. Man 10 is heterozygous for both rs306882 and rs56750753, markers that show distortion in 0442 heterozygous men. (C) Sequence context of SNP 0442 showing the close match to the hotspot motif CCNCCNTNNCCNC³.

Figure 3. *Trans* and *cis* factors influence NCO frequencies at SNP 0442 in the *SPRY3* hotspot. (A) Variation in NCO frequency, with 95% confidence intervals, in 0442C/T heterozygotes carrying two *PRDM9* A alleles (A/A, black), one A allele (A/N, grey) and two non-A alleles (N/N, white). The median conversion frequency per group is indicated by a dotted line. The low ranking of both N/N men amongst all men tested is significant (Mann-Whitney test, $P < 0.05$). (B) Preferential over-transmission of the SNP 0442C allele amongst NCOs. Data are displayed for man 8 with the upper graph showing events detected by amplifying the parental haplotype bearing the 0442T allele (white circles) and screening for the presence of markers from the parental haplotype bearing the 0442C allele (black circles), and *vice versa* for the lower graph. The hotspot region is shaded in grey.

Figure 4. Variation in the frequency of NCOs and COs between men. Frequencies are shown with associated 95% confidence intervals and the *PRDM9* status of each man is

indicated as in Figure 3. Using arcsine-transformed data, there is a weak correlation between NCO and CO frequency but of only borderline significance (PMCC $r = 0.553$, $P = 0.0502$). There is no correspondence between *PRDM9* homozygosity and the ratio of NCO to CO in these men (Mann Whitney U test, $P > 0.05$).

Figure 5. Transmission distortion at SNP 0442 and the CO/NCO decision. (A) Variation in the SNP 0442 C:T transmission ratio amongst sperm NCOs and COs in different men. Circle size reflects the number of each type of recombination event scored for a given man, as shown in the key below. Men showing significant differences in C:T ratio in COs vs. NCOs are shown in black. For clarity, 95% CIs are shown only for men 3 and 22. Man 22 showed extreme bias, with all NCO events involving transfer of 0442C to the 0442T haplotype; we conservatively plotted this assuming one NCO event in the opposite direction. (B) Relationship between NCO:CO ratios and the relative C:T transmission bias in NCOs vs. COs in different men. Extreme TD in NCOs is significantly associated with higher NCO:CO ratios (Mann-Whitney U test, $P < 0.01$). (C) Men 3 and 22 show a significant difference in the structures of NCO events involving 0442. Both men are also heterozygous at hotspot marker rs56750753 (“753”), allowing single and co-conversion events to be scored. The structures of the four types of NCO events are shown with marker alleles from the parental haplotype bearing 0442C in black. The numbers of convertants of each type, together with the numbers of sperm molecules screened, are indicated to the right. Man 22 was also heterozygous for an intervening SNP rs306882; this marker was ignored when scoring single-site conversions at SNP 0442 (hatched circles). No NCOs were scored in one orientation for man 22, though CO frequencies estimated from this assay were indistinguishable between the two orientations (9/17,865 vs. 9/10,729, Fisher’s exact test, $P = 0.33$).

Tables

Table 1. Comparison of the nature, extent and consequence at the population level of TD seen amongst COs at human recombination hotspots.

hotspot ^a	TD at central SNP	P value ^b	central SNP	approx. dist. from centre	over-transmitted allele ^c	mean CO freq	gametic ratio ^d
<i>NID1</i>	0.740	<0.001	Y	70 bp	T(der)	0.03%	50.009 : 49.991
<i>DNA2</i>	0.870	<0.001	R	5 bp	G(der)	0.004%	50.00049 : 49.99951
B	0.810	<0.001	S	68 bp	C(der)	0.04%	50.012 : 49.988
J1	0.850	<0.001	R	63 bp	A(anc)	0.04%	50.014 : 49.986
S2	0.743	<0.001	R	60 bp	A(anc)	0.08%	50.020 : 49.980
<i>SPRY3</i>	0.660	<0.0001	Y	45 bp	C(der)	0.31%	50.049 : 49.950
<i>SPRY 3 including NCOs:</i>							50.132 : 49.870

^a Data for the five autosomal hotspots were taken from refs. 12,13,48,49 & AJJ, unpublished data.

^b Typical probabilities in each man tested that alleles at the central SNP showed 50:50 transmission to reciprocal crossovers (COs).

^c “der”, derived; “anc”, ancestral

^d Gametic ratios in favour of the over-transmitted allele. The gametic ratio for the *SPRY3* hotspot taking NCOs as well as COs into account is shown in bold.

Abbreviations

CO	crossover
dHJ	double Holliday junction
DSB	double strand break
LD	linkage disequilibrium
NCO	non-crossover
PAR1	XpYp major pseudoautosomal region
PAR2	XqYq minor pseudoautosomal region
PMCC	Pearson product moment correlation coefficient
SDSA	synthesis-dependent strand annealing
SNP	single nucleotide polymorphism
TD	transmission distortion
ZnF	zinc finger