Published in final edited form as:
Science. 2019 March 22; 363(6433): . doi:10.1126/science.aau8861.

Factors Influencing Meiotic Recombination Revealed by Whole Genome Sequencing of Single Sperm♦

Anjali Gupta Hinch[#1], **Gang Zhang**[#1], **Philipp W Becker**[1], **Daniela Moralli**[1], **Robert Hinch**[1,2], **Benjamin Davies**[1], **Rory Bowden**[1], and **Peter Donnelly**[1,3,*]

[1]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

[2]Big Data Institute, University of Oxford, Oxford, UK

[3]Department of Statistics, University of Oxford, Oxford, UK

[#] These authors contributed equally to this work.

Abstract

Recombination is critical to meiosis and evolution, yet many aspects of the physical exchange of DNA via crossovers remain poorly understood. We report an approach for single-cell whole-genome DNA sequencing and sequence 217 individual hybrid mouse sperm, providing a kilobase-resolution genome-wide map of crossovers. Combining this map with molecular assays measuring stages of recombination, we identify factors that affect crossover probability, including PRDM9 binding on the non-initiating template homologue and telomere proximity. These factors also influence the time for sites of recombination-initiating DNA double-strand breaks to find and engage their homologues, with rapidly-engaging sites more likely to form crossovers. We show that chromatin environment on the template homologue affects positioning of crossover breakpoints and provide insights into recombination in the pseudoautosomal region.

Recombination is a fundamental component of meiosis, the process that creates gametes in sexually reproducing organisms, and ensures the correct segregation of homologous chromosomes into daughter cells (1). Along with mutation, it shapes patterns of genetic variation in populations, providing the substrate for natural selection.

[*]Corresponding author. donnelly@well.ox.ac.uk.

Author Contributions: P.D. designed the study. B.D. bred the mice. G.Z. developed the single-cell sequencing protocol with assistance from R.B. and performed the single-cell sequencing experiments. P.W.B. performed MNase ChIP-seq. D.M. performed cytological analysis. A.G.H. analyzed the data. R.H. contributed to the H3K4me3 peak caller. A.G.H. and P.D. wrote the paper with input from G.Z., P.W.B, R.H., and B.D.

Competing interests: PD is founder and CEO of Genomics plc, and a partner in Peptide Groove LLP. G.Z., R.B. and P.D. are listed as co-inventors on a patent application for the single-cell DNA amplification and sequencing protocol. P.W.B is now an employee of GeneFirst Ltd.

Data and materials availability: Raw and processed data are available on the Gene Expression Omnibus website under SuperSeries accession GSE125327, comprising GSE125326 (sperm sequencing) and GSE124991 (DMC1). Code is available on Zenodo (DOI 10.5281/zenodo.2540356).

Europe PMC Funders Author Manuscripts

In many species, recombination events occur mainly in narrow regions of the genome called recombination hotspots (2). In humans, mice, cattle, and likely many other vertebrates (3), an early step in recombination is the binding of DNA by the histone methyl transferase PRDM9 (2). A subset of PRDM9 binding sites are subject to the formation of programmed double-strand breaks (DSBs). These breaks are repaired by a specialized pathway, which involves the meiosis-specific protein DMC1, and uses the homologous chromosome as the template for repairing the break (1). How the correct homologous sequence is located efficiently amongst the bulk of chromatin-embedded nuclear DNA remains unclear (4). A subset of the breaks repaired via the homologue become crossovers, whilst the majority resolve without a crossover (5). Any remaining DSBs are likely repaired using the sister chromatid as template (6). Despite its fundamental importance, critical aspects of the meiotic recombination process remain poorly understood.

Most mammals make only a few crossovers per chromosome (7), even though the number of DSBs is substantially greater (8). This raises the question of how the cell determines which DSBs will be repaired as crossovers. While it is clear that not all DSBs are equally likely to resolve as a crossover (9–13), the factors affecting this decision remain largely unknown. Improper crossing-over leads to aneuploidy, which affects 20-30% of human eggs and 1-8% of human sperm (14).

There are currently two major challenges to understanding crossover formation. Firstly, pedigree-based maps in humans and mice only localize crossovers within tens to hundreds of kilobases (15–17). Cytological assays are informative about the staging of meiotic events (18), but crossovers can only be placed within large domains containing dozens if not hundreds of hotspots. An alternative approach, identifying recombinant molecules in sperm at targeted sites, has high precision but is limited to a small number of selected hotspots (10, 19, 20). Whole genome sequencing of single sperm can identify crossovers genome-wide, however existing methods have low resolution (21, 22). Secondly, analyses of genetic maps can be complicated by allelic variation in *Prdm9*, which leads to distinct sets of hotspots, within and between populations (17, 23, 24). This makes it difficult to connect the initiation of recombination with the final outcome of crossovers.

A single-cell DNA sequencing method to identify crossovers in individual sperm

We have developed a method for amplifying and sequencing DNA from single cells, and applied it to sperm to identify crossovers with high resolution (25). We isolate a cell mechanically and amplify its DNA using RNA random priming and Klenow fragment synthesis (Fig. 1A, (25)). Our method achieves uniform coverage, with regions missed randomly, rather than systematically due to genomic features (Figs. S1-2).

This approach was applied to 217 sperm cells from a single adult F1 hybrid mouse, derived from a cross between the C57BL/6J (henceforth B6) and CAST/EiJ (henceforth CAST) inbred strains. The B6 mother is *Mus musculus domesticus* and is genetically altered at the *Prdm9* gene to be homozygous for an allele carrying the zinc-finger domain found in many human populations (26), which we refer to as $Prdm9^{HUM}$. The CAST father is *Mus*

musculus castaneus and homozygous for the mouse wild-type *Prdm9* allele we call $Prdm9^{CAST}$. The F1 mouse (hereafter, hybrid) is thus heterozygous at *Prdm9*, which allows us to compare the properties of hotspots associated with the two *Prdm9* alleles. Since $Prdm9^{HUM}$ is not found in mice, we can separate biological properties of recombination from species-specific evolutionary effects. We also chose this design for the high sequence divergence between the parental strains (27), which improves localization of crossovers, and often allows us to assign events specifically to one or other of the two homologous chromosomes.

We sequenced the individual sperm from the hybrid mouse to a median depth of 6.3X, which yielded a median genome coverage of 62% (Fig. 1B). The genome coverage was stable, with 90% of sperm having coverage between 48% and 70% of the genome (Fig. S1). All sperm were euploid and there were no significant differences between the number of sperm carrying the X (108) or the Y (109) chromosome.

Each chromosome in the sperm is expected to comprise one or more segments of B6 and CAST genomes, with transitions between the haplotypes representing crossovers. We developed a Hidden-Markov model-based computational approach that maps sequencing reads to the B6 or the CAST haplotype (25, 27), and identifies the most likely haplotypes taking sequencing error into account (Fig. 1C). We applied this approach to identify 2649 crossovers in our 217 sperm samples. The median resolution of crossovers is 916 bp, with 386 crossovers localized within 250 bp (Fig. S3). This large study of crossovers localized at a fine-scale in mammals provides a resource for understanding crossover formation genome-wide.

Molecular Assays

Recombination is a multi-stage process (2). We draw attention here to aspects of five of those stages. PRDM9 binds DNA in a sequence-specific manner (stage 1) and places an H3K4me3 mark on nearby nucleosomes (stage 2). SPO11 makes double-strand breaks (stage 3), which are resected to form single-stranded DNA (ssDNA) decorated with the meiosis-specific strand-exchange protein DMC1 and other proteins. The ssDNA covered with DMC1 undergoes a search for its homologous sequence (stage 4) and invades the homologous chromosome. This results in the formation of joint molecules, a subset of which are resolved as crossovers (stage 5).

To identify the factors affecting this process, we used data measuring H3K4me3 (stage 2) (28) and performed assays for DMC1 bound to ssDNA (stage 4) in testes. We focus on processes after the SPO11-induced DSB. In analyses below in the B6 mouse we can compare counts of DSBs, as measured directly using the SPO11-oligos produced with each DSB (29), with downstream properties. However, SPO11-oligo measures require impractically large numbers of mice and are not available in our hybrid. H3K4me3 levels and SPO11-oligos have high biological correlation ($r \approx 0.83$, Fig. S4, (25)). Therefore, where necessary in the hybrid mouse, we use measures of the H3K4me3 mark at hotspots as a surrogate for DSB counts.

Two distinct factors affect ChIP-seq measures for DMC1 on ssDNA: the number of breaks; and how long the ssDNA remains unpaired, which leads to the persistence of DMC1 near the break-site (25). Our peak calling algorithm (26) identified 24,586 peaks for DMC1. We also called peaks in H3K4me3 (25). In addition to hotspots, H3K4me3 is found at transcription start sites, and other functional elements, due to its role in the regulation of gene expression.

Properties of Crossovers

Amongst the 2649 genome-wide crossovers identified, 2615 crossovers are autosomal, corresponding to an autosomal map length of 12.1M, similar to previous work (16, 30). We confirm robust crossover assurance on the autosomes (Fig. 2A). The number of crossovers per cell is compatible with random segregation of homologous chromosomes and sister chromatids, with no evidence of systematic variation in the number of crossovers between gametes (Figs. S5, S6).

Most crossovers overlapped hotspots identified by DMC1 ChIP-seq (92%) or H3K4me3 ChIP-seq (94%). Nearly all crossovers (96%) overlapped at least one of these two sets of hotspots, which is unlikely to happen by chance (25). The expected number and localization of crossovers to known hotspots provides evidence that our single-sperm sequencing approach is effective. Conversely, this also shows that nearly all crossovers happen in hotspots, with little recombination in the rest of the genome.

Hotspots with greater H3K4me3 have more crossovers ($p<10^{-15}$, test for Pearson's correlation), as expected. Total H3K4me3 in hotspots is also a good predictor of the number of crossovers per chromosome ((25), Fig. S7). However, crossovers are seen disproportionately more frequently in hotspots with higher H3K4me3 (Fig. 2B). The most active hotspots have a 5-fold number of crossovers compared to a larger set of less active hotspots with the same total level of H3K4me3 (Fig. S8).

Whilst crossovers overlapped 1634 distinct autosomal hotspots in total, several hotspots showed a high concentration of crossovers (Fig. S9). Each of 17 specific hotspots had crossovers in over 5% of meioses (95% CI=[2%-11%]). One sub-telomeric hotspot on chromosome 19 exhibited crossovers in more than 9% (95% CI=[4%-15%]) of meioses (Fig. 2C). These appear to be the most active hotspots for crossover identified in a mammal to date (10, 19, 20, 31).

PRDM9 variants exhibit unexpected dominance and often bind asymmetrically to homologous sites in hotspots

Recombination hotspots in the hybrid mouse consist of hotspots that are activated by $PRDM9^{CAST}$ or $PRDM9^{HUM}$ and those that are PRDM9-independent (32). We can identify, in most cases, which PRDM9 variant activates a hotspot (25), and the DNA sequence motif to which it binds (Fig. S10) (33). Amongst autosomal crossovers that overlapped hotspots, 2309 (92%) overlapped a single DMC1 hotspot, of which 1377 (60%) overlapped $PRDM9^{CAST}$ and 784 (34%) overlapped $PRDM9^{HUM}$ hotspots. The remaining 148

crossovers (6%) could not be confidently assigned to an allele, including some (29, or 1.3%) where PRDM9-dependent and independent hotspots overlapped. We saw no instances of crossovers in the autosomes that could definitively be assigned to a PRDM9-independent hotspot. PRDM9^{CAST} is dominant over PRDM9^{HUM} for crossovers (64:36, 95% CI=62%-66%), as it is for H3K4me3 (62:38) and DMC1 (68:32) (also observed independently in (28)).

Hotspots in hybrid mice can vary in their activity on the two homologues if sequence differences cause differences in PRDM9 binding affinity at that position (26, 32, 34). Some such sequence differences are random polymorphisms. Others result from degradation of PRDM9-binding sites for evolutionary reasons: as a species evolves with a particular *Prdm9* allele, the best binding sites for that allele are lost from the host genome due to meiotic drive favoring hotspot-disrupting mutations (35). As a result, a particular PRDM9 variant will bind chromosomes from another genome in preference to its own. For example, PRDM9^{CAST} binding sites are lost on CAST chromosomes, with no systematic loss on B6 chromosomes. In the hybrid, this leads to a spectrum of "asymmetry" in PRDM9^{CAST} binding, with reduced binding on the CAST chromosome in many hotspots (Fig. S11). Informally, asymmetry is a measure of the extent to which PRDM9 preferentially binds one of the two homologues at a particular hotspot (Figs. 2D,E). As expected, PRDM9^{HUM} shows asymmetry in a smaller fraction of hotspots, with no overall bias towards either chromosome (Fig. S11): evolutionary loss of binding sites has not occurred for the engineered *Prdm9*^{HUM} allele and the observed asymmetry is due to stochastic variation in DNA sequence, which affects both homologues equally on average.

The dominance of PRDM9^{CAST} over PRDM9^{HUM} despite loss of its binding sites is surprising. Both alleles appear to have similar overall levels of expression (Fig. S12), but different distributions of H3K4me3 across hotspots, with PRDM9^{CAST} hotspots skewed towards greater H3K4me3 (Fig. S13). This suggests a functional difference between the alleles, such as a greater affinity of the PRDM9^{CAST} zinc finger domain for its binding sites.

PRDM9 binding on the non-initiating, template, homologue boosts resolution as a crossover

Crossovers require engagement of the two homologous chromosomes, so it is natural to ask whether PRDM9 binding on both chromosomes influences crossover formation. To check this, we first compare the asymmetry of hotspots with their crossover resolution probability, informally the probability that a particular DSB resolves as crossover, which we estimate using H3K4me3 as the measure of recombination initiation (25). We find that asymmetry correlates with a decrease in crossover resolution probability (Fig. 3A), an effect also observed independently in a mouse pedigree (28). DSBs in very asymmetric hotspots are only 31% (95% CI=[19%–52%]) as likely to resolve as a crossover as those in symmetric hotspots ($p=5x10^{-6}$, (25)). We observe comparable effects for PRDM9^{HUM} and PRDM9^{CAST} hotspots (Fig. S14). The strong observed excess of DMC1 in asymmetric hotspots (Fig. S15) rules out the possibility that these hotspots simply have fewer DSBs.

A previous study (32) noted that hotspots that form crossovers in F1 mice tend to have a lower density of polymorphisms than other hotspots, and proposed that local DNA sequence heterozygosity caused reduced crossover formation. We find a slight effect of heterozygosity ($p=0.06$, (25)), but a significant effect with asymmetry ($p=2\times10^{-13}$, (25)). Since asymmetry and the presence of polymorphisms are correlated ($p=1.6\times10^{-280}$, test for Pearson's correlation), and polymorphisms within the PRDM9-binding motif are usually the cause of asymmetry (26), we checked whether the effect of heterozygosity could be due to its impact on asymmetry. This is the case: there is no effect of heterozygosity once asymmetry is taken into account ($p=1$, Fig. 3B, Table S1).

Asymmetric hotspots are less likely to have both homologues bound by PRDM9 (Figs. 2D,E). Therefore, a possible explanation may be that PRDM9 binding on the template chromosome (hereafter, template) increases the chance of a DSB being repaired as a crossover. To investigate further, we assessed how frequently crossovers arise on the "less-bound" and "more-bound" homologues in asymmetric hotspots. We selected asymmetric hotspots with the property that the more-bound homologue was 20-fold more likely to be bound by PRDM9 than the less-bound homologue on average. We identified 60 crossovers occurring at such asymmetric hotspots for which we could also infer the initiating chromosome (25), that is the chromosome on which the break occurred. If the likelihood of repairing a break as a crossover were equal between the homologues, we would simply expect to find that 20-fold more of these crossovers initiated on the more-bound than the less-bound homologue. Analysis of these 60 cases, however, revealed that significantly fewer than expected initiated on the more-bound homologue (47/60, $p=5\times10^{-6}$, binomial test). The crossover resolution probability shows a clear directional effect (Fig. 3C): when a DSB does occur on the less-bound homologue it is 6 times more likely to form a crossover than a DSB which occurs on the more-bound homologue ($p<10^{-4}$, (25)). The more asymmetric the hotspot, the greater is the difference in crossover resolution between the homologues (Fig. S16). Differences in DNA sequence between the homologues cannot explain this directional effect within hotspots. We also rule out the possibility that the increase in crossovers on the less-bound homologue is driven by an increase in DSBs, as there is less DMC1 than expected on it (Fig. 3D). Our data strongly suggest a model in which it is PRDM9 binding on the template which promotes crossover formation: when a DSB occurs on the less-bound homologue, it is likely that PRDM9 will have bound the template (Fig. 3E). Conversely, when the more-bound homologue is cut, the template will not often be bound by PRDM9 (Fig. 3E). In fact, DSBs initiating on the less-bound homologue are almost 2.5 times more likely than those at symmetric hotspots to resolve as a crossover despite the greater heterozygosity at asymmetric hotspots ($p=0.006$, (25)).

If our model is correct, PRDM9-binding on the template should increase crossover resolution in all hotspots, regardless of whether they are asymmetric or not. Consistent with this, the rate of crossovers originating from a fixed amount of H3K4me3 on the initiating chromosome increases with H3K4me3 on the template chromosome (Fig. 3F). Hotspots with the greatest PRDM9 binding on the template are 4-fold more likely to have crossovers than hotspots with the lowest binding on the template. This effect also helps explain the observation above that more active hotspots have a disproportionately greater number of crossovers (25).

Taken together, these lines of evidence lead us to conclude that PRDM9 binding on the template chromosome increases the chance that a DSB is resolved as a crossover. Our data also imply that PRDM9 binding on the template is probably not essential for crossover formation (Fig. 3A).

PRDM9 allele, GC-content, and proximity to the distal telomere strongly influence crossover resolution

To identify further, independent, effects on crossover formation, we created two sets of otherwise-matched hotspots by pairing each hotspot that has a crossover with a hotspot that does not have a crossover (25). The paired hotspots were matched on their PRDM9-dependent H3K4me3 enrichment on both homologues and chosen to be on the same chromosome (1592 pairs). We then asked whether the hotspot sets differed in additional features.

While $PRDM9^{CAST}$ is dominant over $PRDM9^{HUM}$ overall, $PRDM9^{HUM}$ hotspots have significantly more crossovers than $PRDM9^{CAST}$ hotspots matched for the same level of H3K4me3 and asymmetry (Table S2, odds ratio=1.32, $p=2.7x10^{-4}$, Fisher's exact test). The difference between the variants is greater for more active hotspots than for less active hotspots (Table S3).

We then matched hotspots for the allele, in addition to the criteria mentioned above, and find that hotspots with crossovers have significantly greater GC-content within 500 bp of the hotspot center (Fig. 4A, $p=1.2\times10^{-14}$, paired t-test). This is true of both $PRDM9^{CAST}$ and $PRDM9^{HUM}$ hotspots separately (Fig. S17), so cannot be explained by historical GC-biased gene conversion in $PRDM9^{CAST}$ hotspots. The possible reasons for this observation are either an increase in DSBs, or a greater likelihood for a DSB to resolve as a crossover, in very local regions of higher GC. Previous data for SPO11-oligos in B6 (29) indicates that there is no significant effect of local GC-content on the number of breaks in hotspots (Fig. S18). Therefore, we conclude that greater GC-content is conducive to the repair of DSBs as crossovers.

Mouse chromosomes are acrocentric, and a well-known effect in male mice is a greater number of crossovers near distal telomeres (16), although the reason for this is not known. Crossover counts combine two different effects: the rate at which DSBs occur and the probability that a particular DSB resolves as a crossover. Analysis of SPO11-oligos in B6 establishes that the rate of DSBs within hotspots does not show spatial variation along a chromosome (Fig. S19), although the number of hotspots over broad scales may vary (36). On the other hand, our data show that the probability of DSB resolution as a crossover depends on the chromosomal location of a hotspot, increasing 5-fold from the centromere to the distal telomere (Figs. 4B, S20, S21). This chromosome-wide effect, which is strongest near the distal-telomere, cannot be explained solely by suppression of crossovers at the centromere (11, 12).

The GC-and telomere effects are both observed with and without accounting for H3K4me3 (Figs. S22, S23). Additional analyses show that the four effects, namely, PRDM9 binding on

the template, proximity to the distal telomere, PRDM9 allele and local GC-content, are distinct (Figs. S24-S26, (25)).

Factors that boost crossover probability also lead to faster homologue engagement

Each DSB results in a pair of long-lived SPO11 oligos, therefore quantitative sequencing of these oligos provides a direct measure of the number of DSBs (29). In contrast, the assay for DMC1 measures its *transient* association with the ssDNA near each DSB (37). As a result, it depends on both the number of DSBs and on how long DMC1 remains bound to the ssDNA. Therefore, comparison of DMC1 with SPO11-oligos allows an assessment of the time until DMC1 is no longer associated with ssDNA (36). Since this happens after successful strand invasion takes place and the homologues become locally engaged near the DSB site, we refer to the ratio of DMC1 to SPO11-oligos as a measure of "homologue-engagement time" (25).

The B6 mouse is the only one in which assays of H3K4me3, SPO11, and DMC1 are all currently available. From these data we find that, across hotspots, homologue-engagement time decreases as H3K4me3 levels increase, with DSBs at the most active hotspots engaging the fastest (Fig. 5A). Therefore, we can conclude that homologue-engagement time is affected by PRDM9-binding on the template chromosome (Fig. S27). Interestingly, even on the non-pseudoautosomal region of the X chromosome, where the sister chromatid is thought to be used as the template (6, 38), repair is faster for hotspots with the highest levels of H3K4me3 (Fig. S28).

It is known that DMC1 relative to SPO11 is lower in the 5Mb adjacent to the centromere-distal telomere relative to the rest of the chromosome in B6 (36). Extending this finding, we show that the average DMC1 per hotspot increases further from the centromere-distal telomere (Fig. S29), although the rate of DSBs and the width of DMC1 loading near break-sites remain stable (Figs. S19, S30). Indeed, we find that homologue-engagement time increases continuously as a function of distance from the distal telomere (Fig. 5B), with engagement time for breaks furthest from the distal telomere 25% longer than for the nearest ones. Homologue-engagement time also decreases with increase in local GC-content in B6 (Fig. 5C).

Measurements of SPO11 are not available in our, or indeed in any, hybrid mouse, so to see whether the results from B6 extend to our hybrid, we approximate homologue-engagement time by using H3K4me3 in lieu of SPO11-oligos. Recapitulating the B6 findings, estimated homologue-engagement time in our hybrid mouse also decreases with increasing H3K4me3 on the template (Figs. S15, S31). In asymmetric hotspots, DSBs on the more-bound homologue take almost four times as long to repair as DSBs on the less-bound homologue (Fig. 5D). Similar results on the impact of PRDM9 on DMC1 have been seen in several F1 mice (26). As in B6, estimated homologue-engagement time in the hybrid increases with distance from the telomere ($p=10^{-31}$,(25)) and decreases with local GC-content ($p=2x10^{-15}$, (25)). Finally, for $PRDM9^{HUM}$ hotspots, estimated homologue-engagement time is 18% lower on average relative to $PRDM9^{CAST}$ hotspots (Fig. S32).

In summary, four factors influence homologue-engagement time and in each case they strongly and consistently influence crossover probability (25).

Crossover breakpoints are modulated by the chromatin environment on the template chromosome

Crossover breakpoints, which are the points at which sperm DNA switches from one parental chromosome to the other, have been shown to be contained within the extent of H3K4me3 modification (39) and of DMC1 binding (29) in a small number of hotspots. However, detailed knowledge of breakpoints has been elusive. Our data allow a genome-wide examination of the fine-scale distribution of crossover breakpoints.

For crossovers in symmetric PRDM9^{CAST} hotspots, we see a strongly multi-modal pattern of breakpoints (Fig. 6A). Breakpoints appear to flank positions occupied by nucleosomes around the PRDM9 binding site, with clear peaks in the first, second, and third nucleosome depleted regions (NDRs). In asymmetric hotspots, we also see a multi-modal pattern; however, it is shifted from that of symmetric hotspots (Fig. 6B). For hotspots that are not particularly symmetric or asymmetric, the peaks merge into a more continuous distribution (Fig. S33), as might be expected from a mix of both situations.

Nucleosome positions are known to exhibit a phase shift concomitant with PRDM9 binding (39). The homologue on which the DSB occurs is bound by PRDM9, regardless of whether the hotspot is symmetric or asymmetric (Figs. 6C,D). However, the template is much more likely to have been bound by PRDM9 in symmetric, compared to asymmetric, hotspots, and thus have a different nucleosome profile. The shift in crossover breakpoints that we observe between symmetric and asymmetric hotspots is consistent with the shift in nucleosomes between bound and unbound sites, and with a model in which crossover breakpoints avoid nucleosome positions (Figs. 6A,B, Fig. S34, (25)). We conclude that crossover resolution is modulated by nucleosome positioning on the template chromosome.

Crossover breakpoints also avoid nucleosomes in symmetric PRDM9^{HUM} hotspots (Fig. S35) although, in contrast with symmetric PRDM9^{CAST} hotspots, there does not seem to be a peak in the first NDR from the motif site (Figs. 6A, S36). This may be due to PRDM9^{CAST} binding the template more strongly or for longer than PRDM9^{HUM}, thereby, for example, creating a greater barrier to Holliday junction migration. Alternatively, this may point to differences in the *a priori* histone binding energies in regions that each allele prefers to bind, making nucleosomes more or less difficult to evict. Indeed, while PRDM9^{CAST} preferentially binds sites that are occupied by a nucleosome *a priori*, PRDM9^{HUM} preferentially binds sites that are depleted in nucleosomes (Fig. S37). The overall differences in crossover breakpoints between the two alleles (Fig. S38) reflect both differences at symmetric hotspots, and the different proportions of symmetric and asymmetric hotspots for each allele.

Crossovers in the Pseudoautosomal Region

The pseudoautosomal region (PAR) is a short region of homology between the X and Y chromosomes, which must have a crossover in males for successful segregation of these chromosomes during meiosis. The precise PAR region varies in mouse subspecies – it is ~700 kb long in B6 and 430 kb longer in CAST (40) (Fig. S39). A crossover in the PAR is achieved partly by an increased DSB rate, which is thought to be the result of a disproportionally long axis in this region (41). However, it is not known whether these biological properties of the PAR are determined by *cis*- or *trans*-acting factors. Specifically, it is not clear if the PARs on both chromosomes in the hybrid behave differently, retaining the properties of their parental strains, or if one of the parental strains is dominant.

We compare the DMC1 signal in hotspots in the region that is pseudoautosomal in CAST but not B6 (henceforth het-PAR, Fig. S39). Most of these hotspots have an excess of DMC1 on the CAST relative to the B6 chromosome, with 7-fold greater DMC1 on the CAST chromosome on average (Fig. 7A, $p<10^{-4}$,(25)). This is not explained by any artefactual differences in sequence mapping between the haplotypes (Fig. 7A, p=0.85,(25)). The effect could be explained by a greater number of DSBs on the CAST chromosome or by DSBs initiating on the CAST chromosome taking longer to engage their homologue, or both. Either way, it follows that the CAST and B6 regions behave differently, which implies that the PAR is determined by factors that can distinguish between them, likely cis-acting factors.

We identified 34 PAR crossovers in 217 sperm, all of which are in the het-PAR (Fig. S40). These crossovers demonstrate that any potential structural differences between the two homologues in this region do not preclude reciprocal exchange between them. The number of crossovers we identified is roughly in line with the proportional size of the het-PAR within the whole PAR (34 out of 108.5, which is expected if there is one crossover per meiosis). We do not have power to detect crossovers outside the het-PAR due to lack of adequate sequence assembly.

While previous research has shown the co-existence of PRDM9-dependent and independent hotspots near the PAR (42), their relative importance in crossover formation within the PAR remains unclear (42, 43). We found that 19 crossovers overlapped PRDM9-independent hotspots, 4 overlapped $PRDM9^{CAST}$, and none overlapped $PRDM9^{HUM}$ hotspots. PRDM9-independent hotspots have 53% of the DMC1 signal among het-PAR hotspots, yet the concentration of crossovers in them is substantially greater (83%, p=0.016, (25)). This suggests differences in the timing or processing of DSBs in PRDM9-independent hotspots. The dominance of PRDM9-independent hotspots over $PRDM9^{HUM}$ hotspots also proves that it is not simply a consequence of the evolutionary erosion of PRDM9 binding motifs in this region. Genome-wide the hotspot with the greatest number of crossovers is in the het-PAR, and PRDM9-independent, with crossovers in 11% of meioses (Fig. 7B). While the mechanism controlling PRDM9-independent hotspots is not currently known, we note that the number of DSBs in these hotspots is disproportionately elevated genome-wide in $Atm^{-/-}$ mice (Fig. S41). This shows a role, either direct or indirect, for the ATM pathway in modulating the use of PRDM9-independent hotspots.

Discussion

Recombination via the formation of crossovers is a central part of meiosis. We have identified four distinct factors that affect the probability that a particular DSB is resolved as a crossover: (1) whether PRDM9 is or has bound at the same position on the homologous chromosome; (2) distance from the centromere-distal telomere; (3) local GC-content around the DSB; and (4) whether $PRDM9^{HUM}$ or $PRDM9^{CAST}$ bound the hotspot where the DSB occurred. Our work uniquely separates upstream effects (numbers of DSBs) from those downstream of the breaks, and implicates each of these four factors in an increase in the preferential use of the crossover pathway for DSB repair. The effect of these factors appears to be cumulative, so that hotspots with multiple favourable conditions are most likely to form crossovers (Fig. S42). Equally, the effect of an unfavorable condition in one factor may be mitigated by a favorable condition in another: for example, while breaks in asymmetric hotspots are less likely to resolve as crossovers overall, those in telomere-proximal asymmetric hotspots are more likely to do so than breaks in telomere-distal symmetric hotspots (Fig. S42).

We further show that the same four factors that increase the probability that a particular DSB is resolved as a crossover also decrease homologue-engagement time, namely the time until successful strand invasion takes place and DMC1 is no longer associated with ssDNA. The relative impact of these factors is also consistent for both, with the biggest effect being PRDM9 binding on the homologue, followed by telomere proximity, GC-content, and PRDM9 variant (25). Note that multiple lines of evidence establish that it is PRDM9 binding on the homologue, rather than polymorphisms or hotspot asymmetry *per se*, which affects outcomes for DSBs. The relative effect sizes are consistent with the presence of additional factors impacting the spatial localization of crossovers within a chromosome (25).

Each meiotic cell has to solve, for each DSB, the seemingly intractable problem of finding the homologous sequence amongst billions of bases of DNA (4). The factors we have identified, by virtue of their impact on homologue-engagement time, suggest potential mechanisms that affect this process. A natural explanation for the effect of PRDM9 binding on the homologue is that it facilitates homology search, either directly or indirectly. Possible mechanisms for this include its effect on the local chromatin environment (44), a role in bringing the template homologue to the chromosome axis (45) (thereby reducing the search space) or direct interaction between PRDM9 molecules at the DSB site and the template. Telomeres have distinct properties in meiosis which may facilitate homology search: they are physically bound to the nuclear envelope (46), and may thus be closer to each other *a priori* (47). They also engage in active movements during the phase of meiosis when the search for the homologue is taking place (48). The effect of GC-content could be mediated by its influence on the local chromatin environment (49).

Why are rapidly engaging breaks more likely to become crossovers? A compelling explanation is that delay in finding and engaging the homologue itself is a causal factor. There are several classes of mechanism, which are not mutually exclusive, with this property. In the first class, the earlier a DSB engages its homologue the more likely it is to be resolved as a crossover. For example, sites of early-engaging breaks may be more likely

to appropriate and stabilize protein complexes that are essential for crossover formation (50). This view is consistent with cytological findings that crossover sites are correlated with those where formation of the synaptonemal complex nucleates (18, 51, 52). A second class of model posits a window of opportunity during which DSB sites can acquire the necessary protein complexes and become crossover-proficient (crossover licensing (53)): early-engaging breaks may resolve as crossovers more often by virtue of having found their homologue prior to the end of this period. If multiple breaks on a particular chromosome engage their homologue during this window, other factors may determine which will become crossovers (18). One possibility is that several (or all) breaks might initially proceed down the crossover repair pathway (53, 54) but in the event of a surfeit in prospective crossovers, a subset of them could be re-designated down an alternative repair pathway (crossover designation (53, 55)). A third class of model is that highly-delayed breaks, which may have failed to engage the homologue, are repaired from the sister, for example via a cut-off mechanism after which the cell switches from homologue-mediated to sister-mediated repair of the remaining breaks (6, 38).

Previous research has shown the impact of nucleosomes on the initiating chromosome on strand-resection (56). We have shown that the distribution of crossover breakpoints differs depending on whether PRDM9 has bound the template and is affected by the template's chromatin environment. This suggests that PRDM9 often remains bound (and actively maintaining the local nucleosome environment) on the template until at least strand invasion and perhaps until Holliday junction resolution. Finally, our work sheds new light on how crossover is achieved in the PAR.

Methods Summary

We harvested and isolated 217 sperm from an adult B6xCAST mouse, which has the *Prdm9* alleles $Prdm9^{HUM}$ (26) and $Prdm9^{CAST}$ (32). We developed a protocol for whole-genome amplification and DNA sequencing of single cells (25), which we applied to the sperm. Bulk sperm from the same animal was sequenced at high-depth and the DNA sequence was used to call variants *de novo.* We developed a computational approach to identify the most likely sequence of CAST and B6 haplotypes in each chromosome in each sperm (25). DMC1 ChIP-seq (37) was performed using testis tissue from the same animal, and hotspots were called using our previously published peak-calling algorithm (26). The PRDM9 variants activating hotspots (25) and the PRDM9-binding DNA sequence motifs in them (33) were identified. We developed a method for assessing the evidence of enrichment in H3K4me3 from ChIP-seq data (25). MNase-seq and H3K4me3 Mnase ChIP-seq were performed in testes from another animal with the same genetic background (25).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank R. Li and C. Green for helpful discussions, and the High-Throughput Genomics team at the Wellcome Centre for Human Genetics for all sequencing work. We are grateful to E. Hatton for analytical work in the early development of our single sperm sequencing protocol.

Funding: Wellcome Trust grants 095552/Z/11/Z to PD and grants 090532/Z/09/Z and 20314/Z/16/Z as core support for the Wellcome Centre for Human Genetics. RH is supported by Wellcome Trust grant 106130/Z/14/Z.

References

1. Cole F, et al. Homeostatic control of recombination is implemented progressively in mouse meiosis. Nat Cell Biol. 2012; 14:424–430. [PubMed: 22388890]
2. Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. Nat Rev Genet. 2013; 14:794–806. [PubMed: 24136506]
3. Baker Z, et al. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. Elife. 2017; 6:e24133. [PubMed: 28590247]
4. Renkawitz J, Lademann CA, Jentsch S. Mechanisms and principles of homology search during recombination. Nat Rev Mol Cell Biol. 2014; 15:369–383. [PubMed: 24824069]
5. Cole F, Keeney S, Jasin M. Preaching about the converted: how meiotic gene conversion influences genomic diversity. Ann N Y Acad Sci. 2012; 1267:95–102. [PubMed: 22954222]
6. Lao JP, Hunter N. Trying to avoid your sister. PLoS Biol. 2010; 8:e1000519. [PubMed: 20976046]
7. Dumont BL. Variation and Evolution of the Meiotic Requirement for Crossing Over in Mammals. Genetics. 2017; 205:155–168. [PubMed: 27838628]
8. Martinez-Perez E, Colaiácovo MP. Distribution of meiotic recombination events: talking to your neighbors. Curr Opin Genet Dev. 2009; 19:105–112. [PubMed: 19328674]
9. Sarbajna S, et al. A major recombination hotspot in the XqYq pseudoautosomal region gives new insight into processing of human gene conversion events. Hum Mol Genet. 2012; doi: 10.1093/hmg/dds019
10. de Boer E, Jasin M, Keeney S. Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. Genes Dev. 2015; 29:1721–1733. [PubMed: 26251527]
11. Serrentino ME, Chaplais E, Sommermeyer V, Borde V. Differential association of the conserved SUMO ligase Zip3 with meiotic double-strand break sites reveals regional variations in the outcome of meiotic recombination. PLoS Genet. 2013; 9:e1003416. [PubMed: 23593021]
12. Vincenten N, et al. The kinetochore prevents centromere-proximal crossover recombination during meiosis. Elife. 2015; 4:923.
13. Fowler KR, Sasaki M, Milman N, Keeney S, Smith GR. Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. Genome Res. 2014; 24:1650–1664. [PubMed: 25024163]
14. Capalbo A, Hoffmann ER, Cimadomo D, Ubaldi FM, Rienzi L. Human female meiosis revised: new insights into the mechanisms of chromosome segregation and aneuploidies from advanced genomics and time-lapse imaging. Hum Reprod Update. 2017; 23:706–722. [PubMed: 28961822]
15. Bhérer C, Campbell CL, Auton A. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nat Commun. 2017; 8:14994. [PubMed: 28440270]
16. Liu EY, et al. High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. Genetics. 2014; 197:91–106. [PubMed: 24578350]
17. Kong A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010; 467:1099–1103. [PubMed: 20981099]
18. Hunter N. Meiotic Recombination: The Essence of Heredity. Cold Spring Harb Perspect Biol. 2015; 7:a016618. [PubMed: 26511629]
19. Cole F, Keeney S, Jasin M. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. Mol Cell. 2010; 39:700–710. [PubMed: 20832722]

20. Webb AJ, Berg IL, Jeffreys A. Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. Proceedings of the National Academy of Sciences. 2008; 105:10471–10476.
21. Lu S, et al. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science. 2012; 338:1627–1630. [PubMed: 23258895]
22. Wang J, Fan HC, Behr B, Quake SR. Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm. Cell. 2012; 150:402–412. [PubMed: 22817899]
23. Berg IL, et al. *PRDM9* variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nat Genet. 2010; 42:859–863. [PubMed: 20818382]
24. Hinch AG, et al. The landscape of recombination in African Americans. Nature. 2011; 476:170–175. [PubMed: 21775986]
25. More information on materials, methods, and analyses are available as supplementary materials
26. Davies B, et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature. 2016; 530:171–176. [PubMed: 26840484]
27. Keane TM, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011; 477:289–294. [PubMed: 21921910]
28. Li R, et al. A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. biorxiv.org.
29. Lange J, et al. The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. Cell. 2016; 167:695–708.e16. [PubMed: 27745971]
30. Froenicke L, Anderson LK, Wienberg J, Ashley T. Male mouse recombination maps for each autosome identified by chromosome painting. Am J Hum Genet. 2002; 71:1353–1368. [PubMed: 12432495]
31. Cole F, et al. Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. Nat Genet. 2014; 46:1072–1080. [PubMed: 25151354]
32. Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. The evolutionary turnover of recombination hot spots contributes to speciation in mice. Genes Dev. 2016; 30:266–280. [PubMed: 26833728]
33. Altemose N, et al. A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. Elife. 2017; 6:e28383. [PubMed: 29072575]
34. Baker CL, et al. PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination. PLoS Genet. 2015; 11:e1004916. [PubMed: 25568937]
35. Coop G, Myers SR. Live hot, die young: transmission distortion in recombination hotspots. PLoS Genet. 2007; 3:e35. [PubMed: 17352536]
36. Yamada S, et al. Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice. Cell Cycle. 2017; 16:1870–1884. [PubMed: 28820351]
37. Khil PP, Smagulova F, Brick KM, Camerini-Otero RD, Petukhova GV. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. Genome Res. 2012; 22:957–965. [PubMed: 22367190]
38. Lu L-Y, Yu X. Double-strand break repair on sex chromosomes: challenges during male meiotic prophase. Cell Cycle. 2015; 14:516–525. [PubMed: 25565522]
39. Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. Genome Res. 2014; 24:724–732. [PubMed: 24604780]
40. White MA, Ikeda A, Payseur BA. A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice. Mammalian Genome. 2012; 23:454–466. [PubMed: 22763584]
41. Kauppi L, et al. Distinct properties of the XY pseudoautosomal region crucial for male meiosis. Science. 2011; 331:916–920. [PubMed: 21330546]
42. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. Genetic recombination is directed away from functional genomic elements in mice. Nature. 2012; 485:642–645. [PubMed: 22660327]

43. Hinch AG, Altemose N, Noor N, Donnelly P, Myers SR. Recombination in the human Pseudoautosomal region PAR1. PLoS Genet. 2014; 10:e1004503. [PubMed: 25033397]
44. Kobayashi W, et al. Chromatin architecture may dictate the target site for DMC1, but not for RAD51, during homologous pairing. Scientific Reports. 2016; 6:24228. [PubMed: 27052786]
45. Imai Y, et al. The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. Chromosoma. 2017; 126:681–695. [PubMed: 28527011]
46. Scherthan H. Telomere attachment and clustering during meiosis. Cell Mol Life Sci. 2007; 64:117–124. [PubMed: 17219025]
47. Lee C-S, et al. Chromosome position determines the success of double-strand break repair. Proceedings of the National Academy of Sciences. 2016; 113:E146–54.
48. Lee C-Y, et al. Mechanism and regulation of rapid telomere prophase movements in mouse meiotic chromosomes. Cell Rep. 2015; 11:551–563. [PubMed: 25892231]
49. Wang J, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. 2012; 22:1798–1812. [PubMed: 22955990]
50. Reynolds A, Qiao H, Yang Y, Chen JK, Jackson N. RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat Genet. 2013
51. Gruhn JR, et al. Correlations between Synaptic Initiation and Meiotic Recombination: A Study of Humans and Mice. Am J Hum Genet. 2016; 98:102–115. [PubMed: 26749305]
52. Börner GV, Kleckner N, Hunter N. Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell. 2004; 117:29–45. [PubMed: 15066280]
53. Yokoo R, et al. COSA-1 reveals robust homeostasis and separable licensing and reinforcement steps governing meiotic crossovers. Cell. 2012; 149:75–87. [PubMed: 22464324]
54. Baudat F, de Massy B. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res. 2007; 15:565–577. [PubMed: 17674146]
55. Woglar A, Villeneuve AM. Dynamic Architecture of DNA Repair Complexes and the Synaptonemal Complex at Sites of Meiotic Recombination. Cell. 2018; 173:1678–1691.e16. [PubMed: 29754818]
56. Mimitou EP, Yamada S, Keeney S. A global view of meiotic double-strand break end resection. Science. 2017; 355:40–45. [PubMed: 28059759]
57. Zhang L, et al. Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci USA. 1992; 89:5847–5851. [PubMed: 1631067]
58. Rimmer A, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014; 46:912–918. [PubMed: 25017105]
59. Billings T, et al. DNA binding specificities of the long zinc finger recombination protein PRDM9. Genome Biol. 2013; 14:R35. [PubMed: 23618393]
60. Blom, Gunnar. Transformations of the binomial, negative binomial, Poisson AND $\chi 2$ distributions. Biometrika. 1954; 41:302–316.
61. Lange J, et al. ATM controls meiotic double-strand-break formation. Nature. 2011; 479:237–240. [PubMed: 22002603]
62. Brown MS, Bishop DK. DNA Strand Exchange and RecA Homologs in Meiosis. Cold Spring Harb Perspect Biol. 2014; 7:a016659. [PubMed: 25475089]
63. Cloud V, Chan Y-L, Grubb J, Budke B, Bishop DK. Rad51 Is an Accessory Factor for Dmc1-Mediated Joint Molecule Formation During Meiosis. Science. 2012; 337:1222–1225. [PubMed: 22955832]
64. Da Ines O, Abe K, Goubely C, Gallego ME, White CI. Differing Requirements for RAD51 and DMC1 in Meiotic Pairing of Centromeres and Chromosome Arms in Arabidopsis thaliana. PLoS Genet. 2012; 8:e1002636. [PubMed: 22532804]
65. Singh G, Da Ines O, Gallego ME, White CI. Analysis of the impact of the absence of RAD51 strand exchange activity in Arabidopsis meiosis. PLoS ONE. 2017; 12:e0183006. [PubMed: 28797117]
66. Smagulova F, et al. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature. 2011; 472:375–378. [PubMed: 21460839]

67. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25:402–408. [PubMed: 11846609]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

One Sentence Summary

Molecular assays of initiation combined with a high-resolution map of crossovers identifies key determinants of meiotic recombination outcomes.

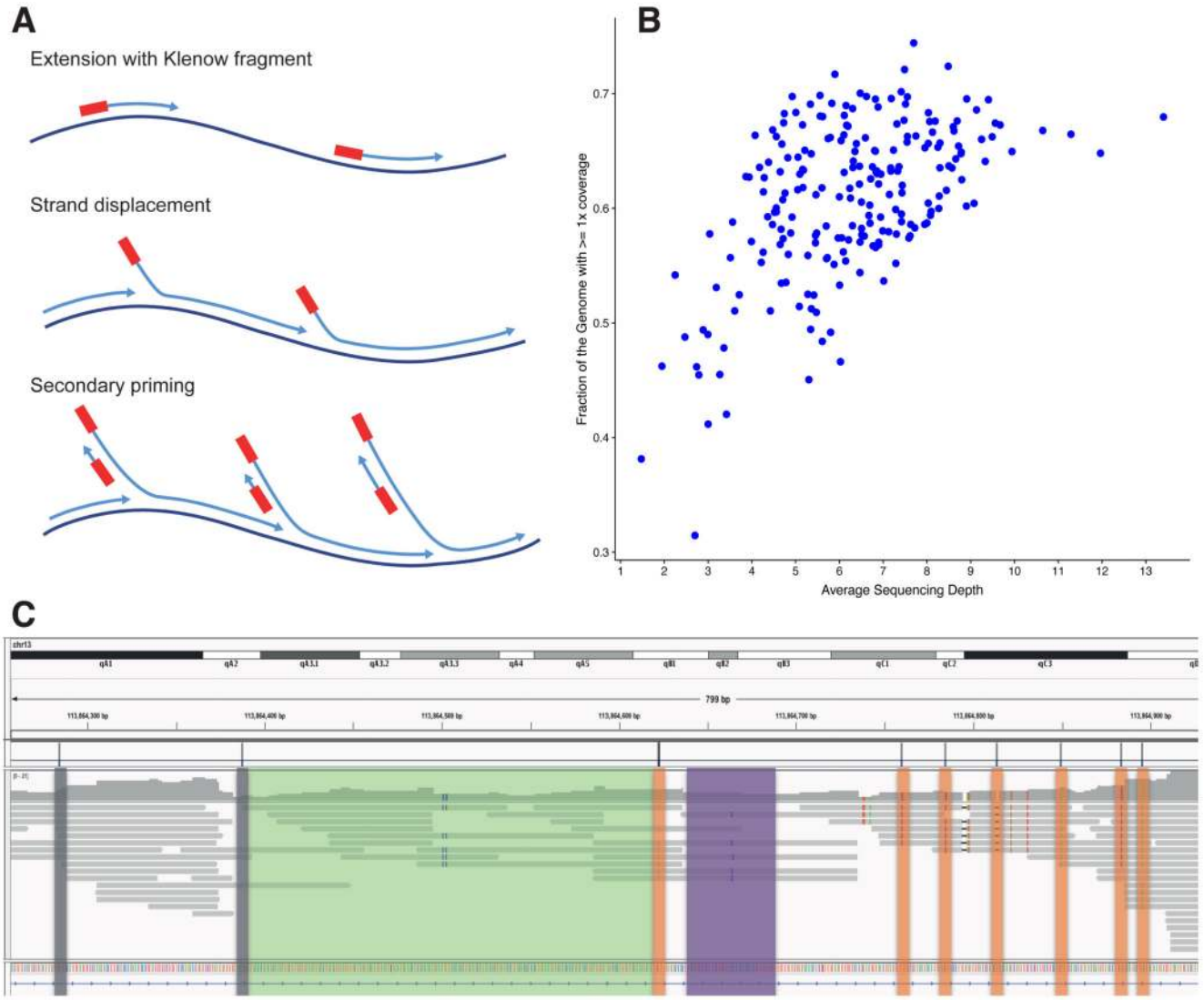

Figure 1. Experimental design for inferring crossovers from single sperm cells.

A. An illustration of the method for whole-genome amplification (WGA) of isolated single sperm cells (25). Random RNA oligonucleotides act as primers for WGA mediated by Klenow fragment, which displaces adjacent synthesized fragments to form overlapping single-stranded DNA copies. These, in turn, serve as templates for primer annealing and chain extension. The resulting amplicons are converted into double-stranded DNA for sequencing.

B: Sequencing depth and genome coverage achieved for each of 217 sperm.

C: An Integrative Genomics Viewer (IGV) illustration of how a crossover was called by our method. The horizontal light gray lines show the reads that mapped in a region of chromosome 13 for a particular sperm. Vertical dark gray bars highlight variants found only in B6, while orange bars highlight variants found only in CAST. The crossover breakpoint lies within a region of uncertainty (green). This crossover overlapped a PRDM9^{HUM}

Europe PMC Funders Author Manuscripts

hotspot, identified by DMC1 ChIP-seq (25), whose center was inferred to be at 113,864,493. A good match to the PRDM9^{HUM} binding motif occurs in the purple region.

Europe PMC Funders Author Manuscripts

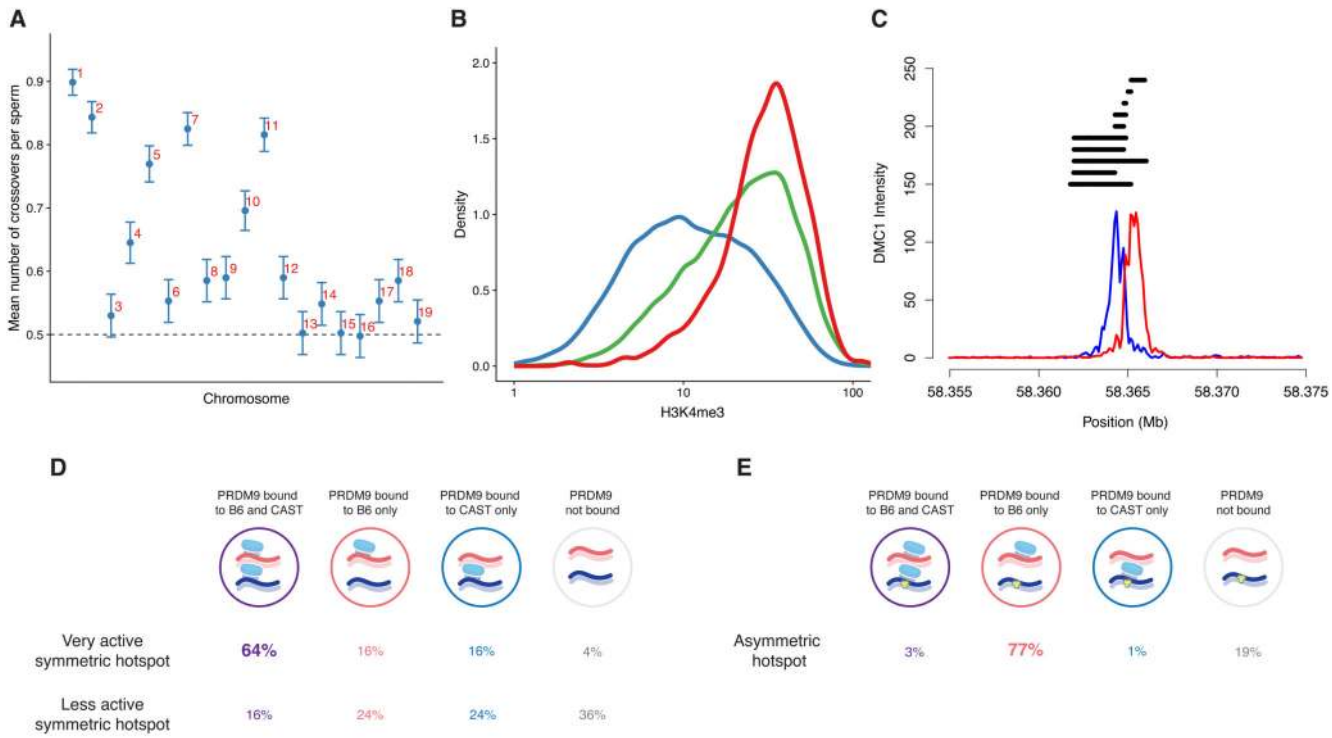

Figure 2. Properties of crossovers and recombination hotspots.

A: Average number of crossovers called per chromosome per sperm (bars show 1 standard error), showing at least, and in many cases almost exactly, one crossover per chromosome per meiosis (equivalently 0.5 crossovers per haploid sperm).

B: Distribution of H3K4me3 intensity in all autosomal recombination hotspots identified by DMC1 ChIP-seq (blue), after removing hotspots showing evidence of PRDM9-independent H3K4me3 (e.g. transcription-start sites, (25)). If crossovers occurred in proportion to the hotspot heat, the distribution of H3K4me3 in hotspots with crossovers should be the corresponding size-biased distribution (green). The observed distribution of H3K4me3 in hotspots with crossovers (red) is skewed further towards hotspots with greater H3K4me3 ($p = 10^{-90}$).

C: The most active autosomal hotspot for crossover is on the centromere-distal end of chromosome 19. DMC1 binds the 3' ssDNA overhangs on either side of the DSB, which leads to a shift between DMC1 coverage on the forward (blue) and reverse (red) strands (200 bp smoothing). Regions containing the crossover breakpoint in each sperm are in black. Crossovers at the same locus in distinct sperm can have different resolution, depending on the actual sequencing coverage achieved in each case.

D: PRDM9 binding at a hotspot is a stochastic event in a cell. In a population of cells, some proportion of cells will have one, both, or neither homologue bound by PRDM9 (sky blue). Here we show the proportion of times each of these possibilities occurs at two illustrative symmetric hotspots. In the very active hotspot (top row), PRDM9 binds the B6 (red) and CAST (blue) homologues with probability 80% each. As a result, PRDM9 is bound to *both* homologues in the majority of cells (64%). In the less active hotspot, the probability of

Europe PMC Funders Author Manuscripts

PRDM9 binding each homologue is 40%. The proportion of cells in which PRDM9 is bound to both homologues is lower (16%).

E. As in D, a comparison of the proportion of cells in which PRDM9 (sky blue) binds one or both homologues, B6 (red) and CAST (dark blue) but at an illustrative asymmetric hotspot. The probability of PRDM9 binding the B6 homologue is ~80% but only ~4% for the CAST homologue. This is due to a SNP (yellow) in the PRDM9 motif on the CAST homologue, which partially disrupts binding. Only a small minority of cells have PRDM9 bound to both homologues.

Europe PMC Funders Author Manuscripts

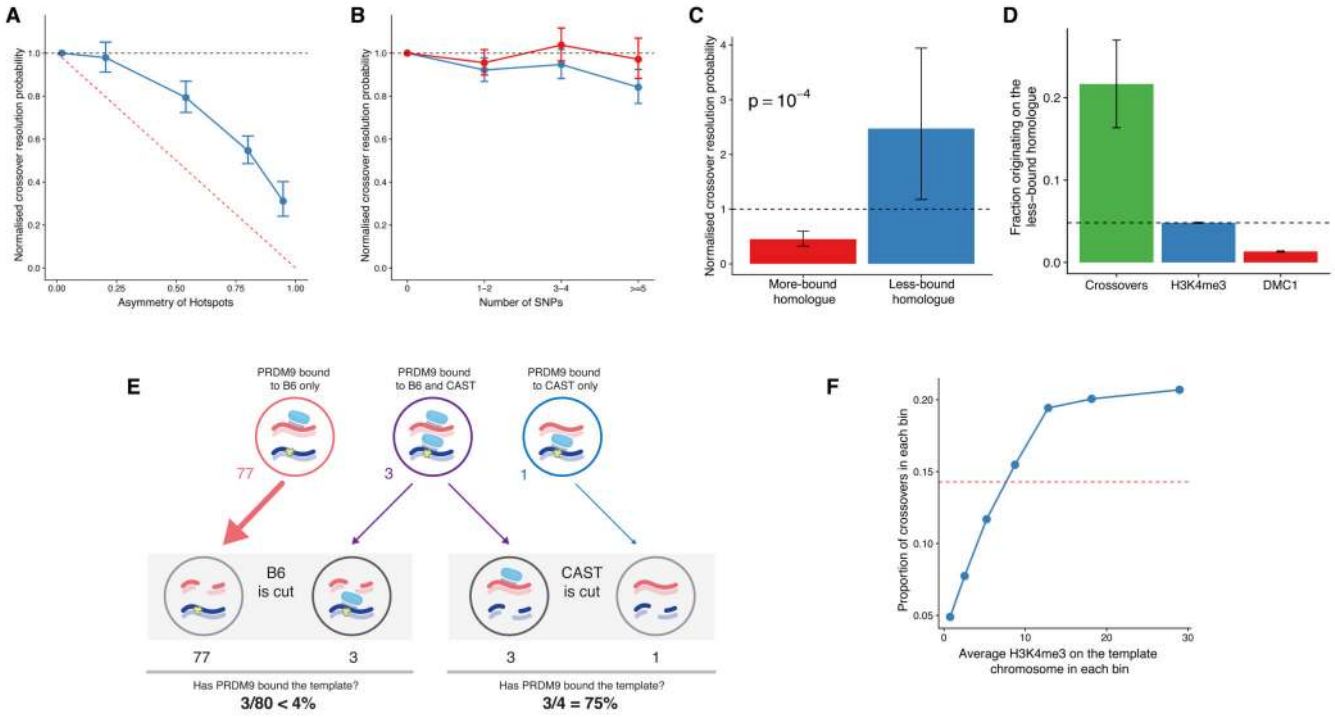

Figure 3. PRDM9 binding on the non-initiating, template homologue affects crossover resolution.

A: Hotspots were binned into five groups depending on the level of asymmetry in PRDM9 binding of the homologues (25). The crossover resolution probability, which accounts for differences in H3K4me3, in each bin (normalized relative to the bin with the most symmetric hotspots, y-axis, 1 standard error bars), is plotted against the mean asymmetry of hotspots in that bin (x-axis). Predicted effects on crossover resolution if PRDM9 binding on the template homologue was irrelevant (black) and if it was essential (red) are shown for comparison.

B: Hotspots were grouped into four bins depending on the number of SNP differences between B6 and CAST chromosomes in the central 200 bases of the hotspot. The crossover resolution probability in each bin (blue) was inferred relative to the bin containing hotspots with zero SNPs. Red points show the same quantity after correcting for asymmetry in PRDM9 binding. Bars show 1 standard error.

C: Crossover resolution probability is significantly higher for DSBs initiated on the "less-bound" homologue than the "more-bound" homologue in asymmetric hotspots. Crossover resolution probabilities for initiation on the more-bound (red, n=47) and less-bound homologues (blue, n=13), after accounting for differences in H3K4me3 on them. Probabilities were normalized against the average for symmetric hotspots (dashed black line), bars show 95% confidence intervals (25).

D: Fraction of crossovers (green), H3K4me3 (blue) and DMC1 (red) originating on the less-bound chromosome in asymmetric hotspots, with dashed line marking the proportion expected from H3K4me3. While the fraction of crossovers initiating on the less-bound chromosome is significantly greater than expected from H3K4me3 ($p=5x10^{-6}$), the fraction of DMC1 is significantly lower than expected from H3K4me3 ($p<10^{-16}$). Bars show 1 standard error.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

E: Illustration that the probability of PRDM9 having bound the *template* depends on which homologue is initially cut for the same asymmetric hotspot as in Fig. 2E. A DSB is more likely to occur on the more-bound homologue B6 (red). When it does, fewer than 4% of cells (3/80) will have the template CAST homologue (blue) bound. In the less likely event that the CAST homologue is cut, the B6 homologue will have been bound in 75% of cells (3/4). (Note that cells with PRDM9 bound on both homologues are twice as likely to be cut at this hotspot than cells with only one homologue bound.)

F: Crossover resolution is influenced by PRDM9 binding on the template homologue. All potential sites for recombination initiation, i.e., the B6 and CAST homologous sites in each hotspot, were sorted according to the H3K4me3 on their respective template homologues. The initiating sites were then binned into 7 bins, such that the total H3K4me3 intensity on the initiating sites in each bin is the same. The proportion of crossovers that initiated in each bin (out of 685 crossovers where the initiating homologue could be inferred) is shown against the average H3K4me3 on the corresponding template homologues (x-axis). Dotted red line shows the expected relationship if H3K4me3 on the template were unrelated to crossover outcome.

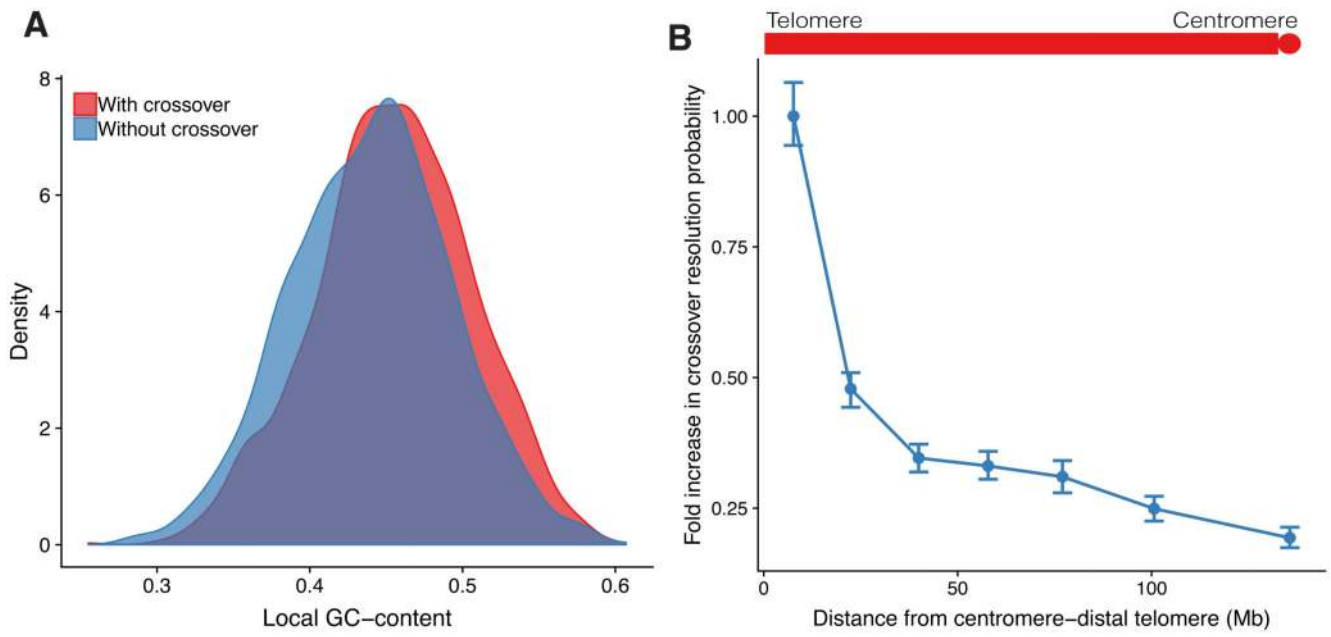

Figure 4. Crossover resolution is affected by local GC-content and Telomere proximity.

A: Each autosomal hotspot with a crossover was paired with another hotspot lacking a crossover for the same PRDM9 variant, on the same chromosome and with very similar H3K4me3 on both homologues (25). The distribution of local GC-content (500 bp around the hotspot center) is compared between the two matched sets (n=1355, $p=1.2\times10^{-14}$, paired t-test).

B: Hotspots were divided into 7 bins depending on their distance from the distal telomere of their respective chromosome. Crossover resolution probability (relative to the leftmost bin) is shown for each bin (1 standard error bars). Chromosomes with more than one crossover in an individual sperm were removed to avoid confounding with crossover interference (see Figs. S20-S21 for additional views).

Europe PMC Funders Author Manuscripts

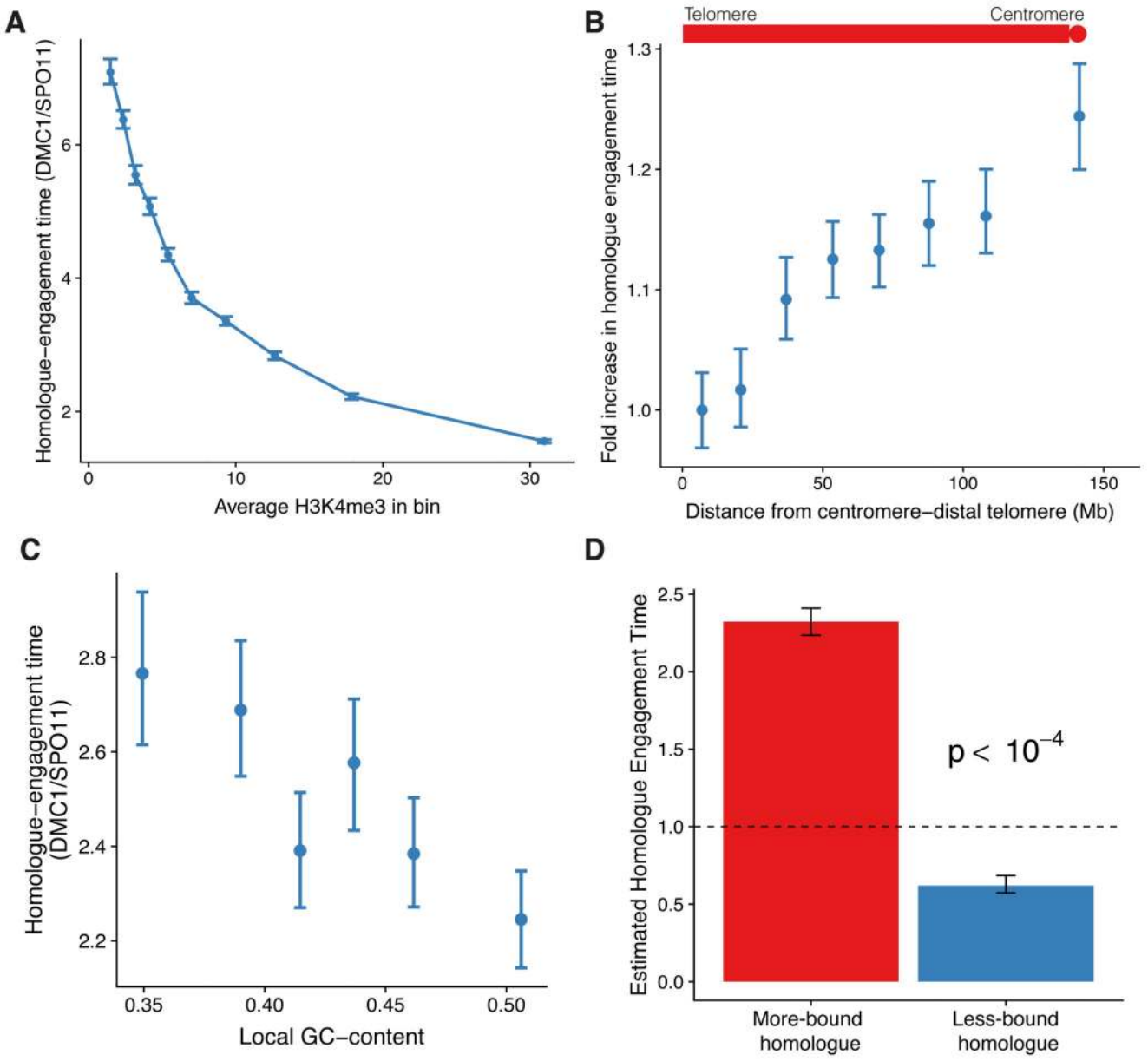

Figure 5. Factors affecting homologue-engagement time in the repair of DSBs.

A: Hotspots in the B6 mouse were ordered by their H3K4me3 intensity and divided into 10 bins. Average homologue-engagement time, the ratio of total DMC1 with total SPO11 per bin (y-axis), is shown relative to the average H3K4me3 per hotspot in each bin (x-axis).

B: Hotspots in the B6 mouse were divided into 8 bins depending on their distance from the distal telomere of their respective chromosome. Average homologue-engagement time (ratio of total DMC1 with total SPO11 in each bin) is shown (1 standard error bars).

C: Hotspots in the B6 mouse were divided into 6 bins depending on their local GC-content (± 500 bp around the hotspot center). Average homologue-engagement time per bin is shown (bars show 95% confidence intervals).

D: Comparison of estimated homologue-engagement time for the more-bound homologue (red) and less-bound homologue (blue) in asymmetric hotspots (corresponding to Fig. 3C, 95% confidence intervals, (25)). Estimated homologue-engagement time (ratio of DMC1 with H3K4me3) is normalized against the average for symmetric hotspots (dashed black line).

Europe PMC Funders Author Manuscripts

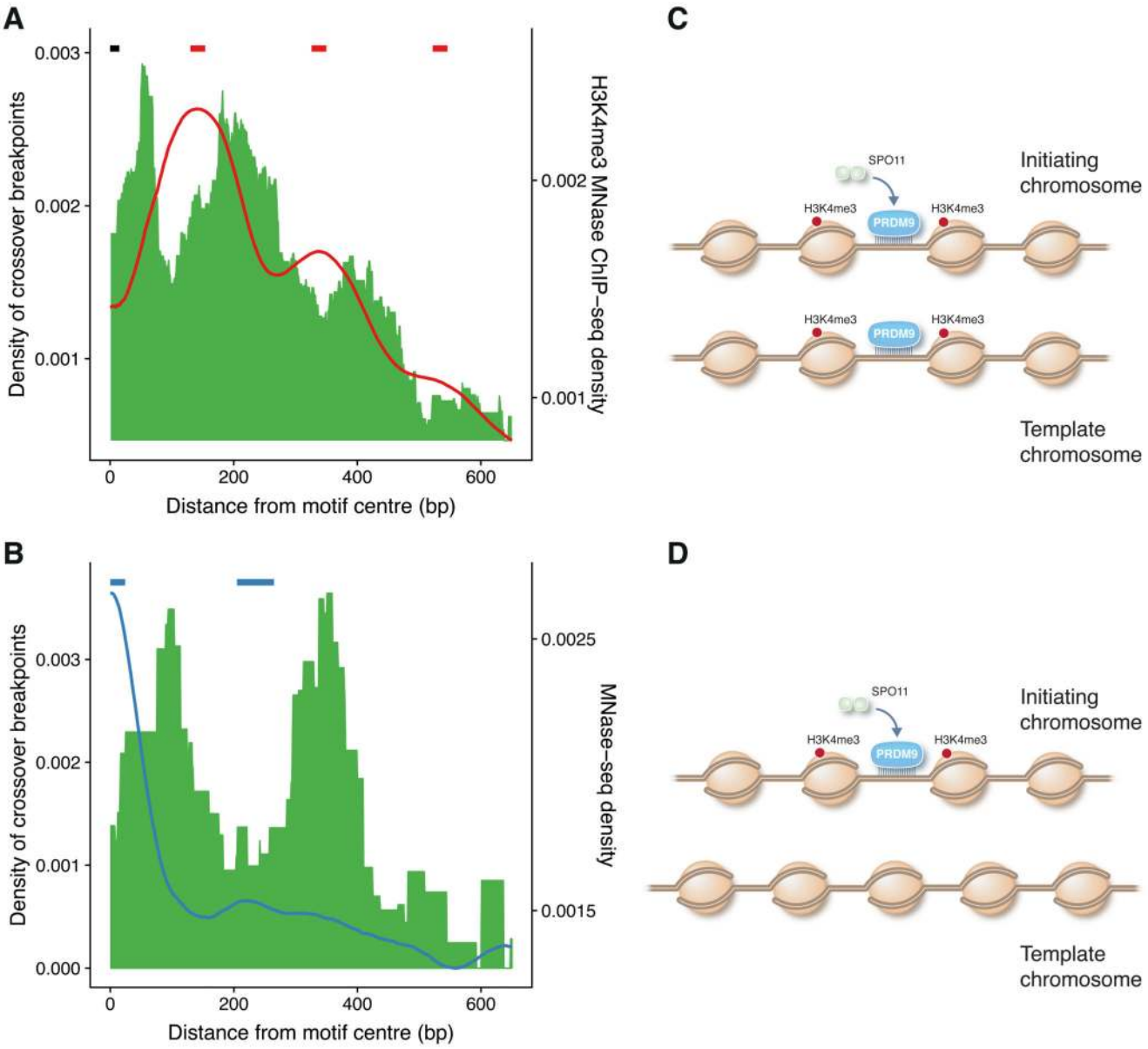

Figure 6. Positioning of crossover breakpoints is influenced by nucleosome positioning on the template chromosome.

A: Distribution of crossover breakpoints from the motif center (green) for crossovers that overlap symmetric PRDM9^{CAST} hotspots with a well-identified motif site, and have breakpoint resolution ≤50bp (n=132). To deal with the uncertainty in crossover breakpoint location in each sperm we assign equal weight to all possible breakpoint positions in that sperm (25). H3K4me3 ChIP-seq with MNase averaged over PRDM9^{CAST} hotspots in red (20bp smoothing). Red bars at top show average inferred positions of nucleosomes, black bar shows the PRDM9^{CAST} binding site.

B: As (A) but for crossovers that overlap asymmetric PRDM9^{CAST} hotspots (n=33). Average MNase-seq for the less-bound chromosome of asymmetric hotspots (blue, 50bp smoothing), with blue bars at top showing average inferred nucleosome positions. This is an estimate of

the nucleosome positioning at hotspot sites when PRDM9 is not bound (25). The peak in MNase-seq at the hotspot center is consistent with the presence of a nucleosome in $PRDM9^{CAST}$ hotspots in the absence of PRDM9 binding (Fig. S37).

C: Illustration of nucleosome positions when the template homologue is bound by $PRDM9^{CAST}$. DNA (dark brown) around histones (light brown), with red dots indicating H3K4me3 mark. Nucleosome positions on the DSB-initiating and template homologues are the same. This is more likely in symmetric hotspots (A).

D: Illustration of nucleosome positions when the template homologue is not bound by $PRDM9^{CAST}$. Colours as in (C). Typical nucleosome positioning at sites bound by PRDM9 is shifted relative to unbound sites, resulting in a difference between the DSB-initiating and template chromosomes. This is more likely in asymmetric hotspots (B). The shift in crossover breakpoints between (A) and (B) is consistent with the shift in nucleosome positions on the template homologue, as illustrated in (C) and (D).

Europe PMC Funders Author Manuscripts

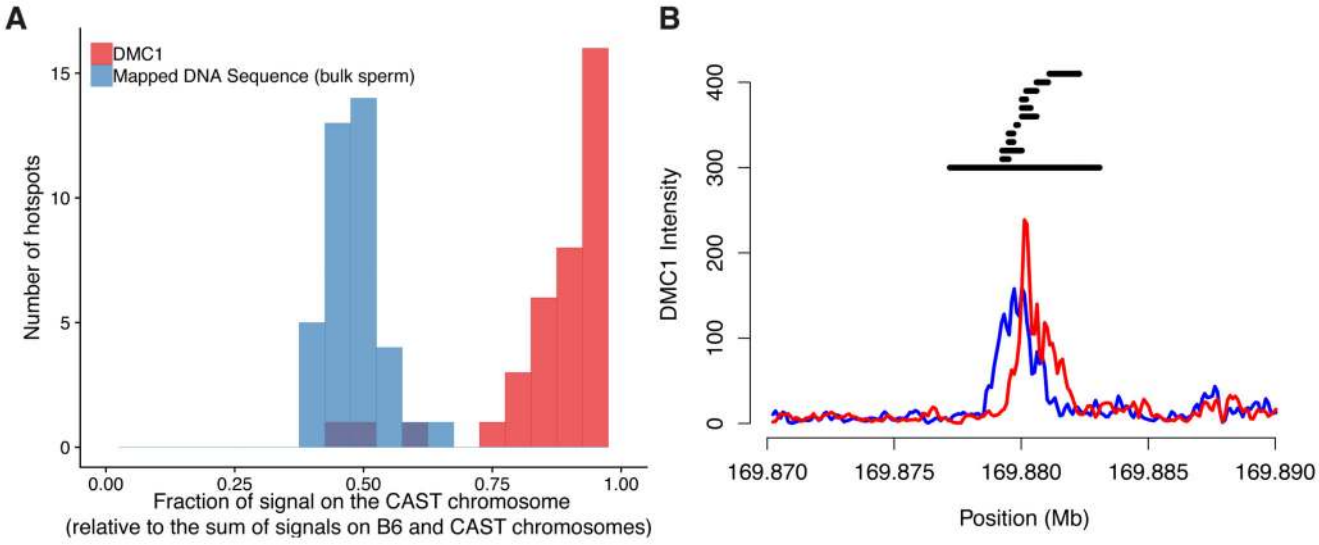

Figure 7. Differences in recombination in the pseudoautosomal region.

A: Histogram of the fraction of DMC1 reads on the CAST chromosome across hotspots (red, n=38). For the same regions, the corresponding histogram for reads from sequencing of bulk sperm is shown (blue) as a control to assess potential mapping artefacts. While DMC1 is significantly biased towards the CAST haplotype ($p<10^{-4}$, (25)), there is no significant bias in bulk sequencing (median=0.51, p=0.85, (25)).

B: The most active hotspot for crossovers in the entire genome is in the het-PAR and is PRDM9-independent. DMC1 coverage (200bp smoothing) is shown for the forward (blue) and reverse (red) strands. Crossover breakpoints are in black. See Fig. S43 for a further het-PAR hotspot.

Europe PMC Funders Author Manuscripts