A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination

Ran Li1,2,3,5, Emmanuelle Bitoun1,2,5, Nicolas Altemose1,2,4,5, Robert W. Davies1,2, Benjamin Davies1 & Simon R. Myers1,2

During meiotic recombination, homologue-templated repair of programmed DNA double-strand breaks (DSBs) produces relatively few crossovers and many difficult-to-detect non-crossovers. By intercrossing two diverged mouse subspecies over five generations and deep-sequencing 119 offspring, we detect thousands of crossover and non-crossover events genome-wide with unprecedented power and spatial resolution. We find that both crossovers and non-crossovers are strongly depleted at DSB hotspots where the DSB-positioning protein PRDM9 fails to bind to the unbroken homologous chromosome, revealing that PRDM9 also functions to promote homologue-templated repair. Our results show that complex non-crossovers are much rarer in mice than humans, consistent with complex events arising from accumulated non-programmed DNA damage. Unexpectedly, we also find that GC-biased gene conversion is restricted to non-crossover tracts containing only one mismatch. These results demonstrate that local genetic diversity profoundly alters meiotic repair pathway decisions via at least two distinct mechanisms, impacting genome evolution and Prdm9-related hybrid infertility.
During meiosis, genetic information is exchanged between homologous chromosomes via the process of recombination. In mammals and other taxa, recombination is essential for the proper pairing of homologous chromosomes (synapsis) and their segregation into gametes and, together with mutation, generates all genetic variation. In many species, most recombination events cluster into small 1–2 kb regions of the genome, called recombination hotspots. In mice, humans and many other vertebrate species, these hotspots are positioned mainly by PRDM9, a zinc-finger protein that binds specific DNA sequence motifs and deposits at least two histone modifications, H3K4me3 and H3K36me3, on the surrounding nucleosomes. DNA double-strand breaks (DSBs) subsequently form near a small subset of PRDM9 binding sites in each cell, and DSB processing results in single-stranded DNA decorated with the strand exchange proteins RAD51 and DMC1. Meiotic recombination hotspots can be mapped genome-wide by performing versions of Chromatin Immunoprecipitation with high-throughput sequencing (ChIP-seq) in bulk gonadal tissue samples. Because H3K4me3 is deposited at essentially all PRDM9 binding sites, meiosis-specific H3K4me3 ChIP-seq enrichment can be used to approximate the frequency of PRDM9 binding at each site. DMC1 marks DSB sites prior to repair processing, so DMC1 enrichment at individual hotspots increases both with the rate at which DSBs occur and with the average time until these DSBs are repaired.

Each DSB can ultimately repair by homologous recombination in several ways. First, the single-stranded ends surrounding each DSB must participate in homology search, to identify a suitable repair template within the vast search space of the genome. The homologue is favoured as a repair template, as opposed to the sister chromatid. One exception is the X chromosome, which has no homologue in males and is thought to rely on sister-templated repair late in male meiotic prophase. A minority of DSBs form crossovers (COs), involving reciprocal exchanges between homologues, while many more DSBs become non-crossovers (NCOs), in which a section of genetic material is copied (converted) from the homologue, without the donating chromosome being altered. Although the role of recombination in shaping genetic variation is well understood, our understanding of possible effects in the reverse direction—of local genetic variation on CO and NCO event outcomes—remains very incomplete in mammals. Previous work has revealed impacts of genetic variation on the formation of DSBs, by altering PRDM9 binding properties. We previously published evidence that the degree to which PRDM9 binding is symmetric—that is, whether PRDM9 binds both homologues equally well at each site—predicts increased synopsis and fertility in certain hybrid male mice. Moreover, individual asymmetric hotspots show elevated DMC1 ChIP-seq enrichment relative to H3K4me3, most consistent with the possibility that these DSBs take longer to repair. One study in mice reported that hotspots with high polymorphism rates, particularly those with asymmetry, show fewer COs than expected from their DMC1 enrichment in males. However, their data could not discriminate between the following causal possibilities: (i) genetic diversity per se was responsible (similar to findings in yeast); (ii) DMC1 elevation through repair delay entirely drove the signal; or (iii) these sites were preferentially repaired instead by NCO recombination. Despite this important progress, it therefore remains unclear how, or even if, local genetic variation might impact eventual recombination outcomes following DSB formation, or how this might differ for COs and NCOs and between sexes.

Furthermore, many fundamental questions remain about the process of NCO recombination itself due to the difficulty of detecting NCO events, as NCOs are very short and the ability to detect them relies on the conversion of nearby SNPs, which is less likely to occur in individuals with low heterozygosity. Previous studies in humans and mice have revealed that in males, most (~70%) NCO events occur within PRDM9-positioned recombination hotspots and are predominantly short (<1 kb) and simple: they comprise contiguous tracts of converted SNPs. In contrast, complex NCO events, which contain both converted and non-converted SNPs and often extend over 1 kb, are seen at a greater rate in females than males, and they show an association with maternal age. Finally, human NCO events show a strong overall bias towards G/C bases (68%) as opposed to A/T bases, occurring via an unresolved molecular mechanism. This phenomenon is thought to have driven regional differences in the GC-content of many species genome-wide. Possible causes of this GC-bias include either subtle event initiation biases, or heteroduplex DNA repair pathways. It has been unclear to what extent these findings for humans might generalise to other species, or to what extent findings from individual hotspots might generalise to hotspots genome-wide. Importantly, lack of power has prevented resolution thus far of basic questions about meiotic recombination, including any precise estimate of the extent these findings might generalise to hotspots genome-wide. Importantly, lack of power has prevented resolution thus far of basic questions about meiotic recombination, including any precise estimate of the length of underlying NCO tracts, the total number of homologous recombination events per meiosis, or where NCO events position relative to PRDM9 binding sites and DSBs, although studies at individual hotspots in mice have suggested a fairly broad distribution.

To investigate links between genetic variation and repair outcomes, we mapped both CO and NCO events in mice, including mice humanised at Prdm9, in both sexes. Critically, we also gathered complementary H3K4me3 and DMC1 ChIP-seq data, allowing us to analyse PRDM9 binding and DSB formation. Together, these data provide an unprecedented opportunity to investigate each step of meiotic recombination genome-wide and with high resolution, from PRDM9 binding, to DSB formation, to CO and NCO repair.

**Results**

**Building a high-resolution CO and NCO map.** We identified both CO and NCO events in hybrids of two mouse strains: C57BL/6J, humanised at Prdm9 (hereafter B6Hum) and predominantly of Mus musculus domesticus origin) and CAST/EJ (hereafter CAST, predominantly of Mus musculus castaneus origin). Their high sequence divergence (0.7%; Methods) improves power to detect NCO events in offspring. B6Hum is identical to C57BL/6J except that the portion of the B6 Prdm9 exon 10 encoding the DNA-binding zinc-finger array has been replaced with the orthologous sequence from the human PRMD9 B allele. To produce a new allele we label Prdm9Hum, distinct from the Prdm9Cast allele possessed by CAST. The different Prdm9 alleles allow us to distinguish the properties of Prdm9Cast and Prdm9Hum controlled recombination hotspots, with the humanised allele being of interest because it has not co-evolved with either mouse subspecies’ genome. We sequenced 11 F2 offspring of (B6xCAST)F1 mice, and after breeding for five generations in total to accumulate recombination events controlled by Prdm9Hum, we sequenced 72 (B6xCAST)F5-Padm9Hum mice and their 36 F4 parents (Methods; Fig. 1a; Supplementary Fig. 1; Supplementary Data 1). We also gathered ChIP-seq data for both DMC1 and H3K4me3 in testes from male (B6xCAST)F1-Prdm9Hum/Cast mice, allowing us to compare these to NCO/CO event outcomes. Importantly, the high heterozygosity in these mice enabled us to assign ChIP-seq reads to each homologue separately, providing a measure of homologue-
specific PRDM9 binding and DSB formation. 23,748 DMC1 peaks correspond to DSB hotspots, and 63,050 PRDM9-dependent H3K4me3 peaks mark PRDM9 binding sites. Essentially all DMC1 peaks show evidence of H3K4me3 enrichment (98% at \( p < 0.05 \) by likelihood ratio testing; Supplementary Data 2). For most peaks, we are able to determine which \( \text{Prdm9} \) allele controls them (Supplementary Note 1). Together these data allow us to compare signatures of PRDM9 binding, DSB formation, and NCO/CO events in both sexes.

To find both CO and NCO events, we developed and applied a Hidden Markov Model (HMM)-based algorithm to infer background states (B6/B6, B6/CAST and CAST/CAST) across the genome in each mouse to test potential gene conversions against (Fig. 1b and Supplementary Note 2). Each SNP is assigned a particular background state. CO events correspond to background changes between successive SNPs. SNPs with genotypes not matching their local background represent either rare NCO events, or common sequencing errors. Following careful filtering to exclude such errors (Supplementary Note 3 and Supplementary Table 1) we identified 183 NCOs and 295 CO events on autosomes from the 11 F2 animals (Fig.1c) and 1,392 NCOs and 2,205 COs in the 72 F5 mice (Fig. 1d and Supplementary Data 3 and 4). This represents ~3-fold more NCO events identified by direct sequencing, which avoids ascertainment biases, than in the largest previous mammalian study, which was performed in humans - allowing inter-species comparisons. Sequencing-based validation of F2 events (Methods) estimated that 91% of the identified NCO events are real, while using simulations we estimate our power to identify those NCO events containing at least 1 SNP as 73% or above (Methods

---

**Fig. 1** Study design and properties of CO and NCO events. **a** Study design. Arrows indicate locations of de novo CO and NCO events. 11 F2 mice, 36 F4 mice, and 72 F5 mice were sequenced. **b** Detection of NCOs by comparing observed genotypes and background. **c, d** Distribution of identified CO breakpoints and NCOs across autosomes from F2 (c) and F5 (d) animals. **d** Includes both de novo and distinct inherited events from the F5 mice. Asterisk on Chr17 indicates the position of the \( \text{Prdm9} \) gene, which was selected to be homozygous and may show lower power to detect nearby F5 events. e Binning all events by their distance to the centromere or telomere (x-axis); shaded regions represent 90% bootstrap CIs. "Null" represents the results of scrambling NCO positions uniformly across each chromosome.
and Supplementary Fig. 2a, b). In the F5 mice, we were able to identify both de novo and parentally inherited NCO and CO events; and we were able to assign a subset to the maternal or paternal meiosis (Methods). Among the 821 F5 de novo COs, we identified 321 paternal events and 382 maternal events (along with 118 events that were unable to be assigned). NCO events can be assigned to a genetic background by determining whether they result from a DSB on the B6 or CAST chromosome. In the 510 de novo NCOs from F5 mice, we identified 121 paternal events and 130 maternal events (with 259 unassigned). We note that our CO breakpoints are resolved to the sequence level, with a median CO breakpoint resolution of 396 bp.

**Overall event properties.** NCO and CO events, as well as DMC1 and H3K4me3, show enrichment nearer to telomeres, especially male COs (Fig. 1e and Supplementary Fig. 2c, d). This is broadly similar to patterns observed in other mice41–43, humans44, and other organisms such as dogs (which have lost PRDM9)38, although COs show somewhat stronger telomeric enrichment than NCOs. NCOs are also enriched near centromeres, especially on smaller chromosomes (Fig. 1e and Supplementary Fig. 2e). Notably, the telomere effect appears less pronounced among NCO events with recombination hotspots identified by ChIP-seq (Fig. 2f). By examining the joint distribution of COs and NCOs in F2 mice, we confirm that NCOs can occur on both sister chromatids within each pair, regardless of whether each sister also has a CO (Supplementary Table 2). We also found that broad-scale CO and NCO rates have a strong positive association with GC content; after controlling for GC content, we found that COs are less likely to occur in regions belonging to the gene-rich chromatid compartment A45 (Supplementary Table 3 and Supplementary Note 4).

Surprisingly and in contrast to events in human females, 99.5% of observed NCO events were simple and comprised contiguous tracts of converted SNPs, with no non-converted SNPs amongst them. Similarly, 99.4% of de novo COs were simple background switches. This implies that complex events are extremely unusual in mice. Moreover, we observed a very high overlap of both CO and NCO events with recombination hotspots identified by ChIP-seq, stronger than observed in humans21,26,27, with little recombination in the remainder of the genome. In F2 mice (whose parents have the same F1 genetic background as the ChIP-seq samples), 96% of CO events and 92% of NCO events overlap either DMC1 or H3K4me3 ChIP-seq peaks (adjusted for false-positive NCO events and chance overlap; 84% unadjusted). This reduces only modestly in F5 mice, where only Prdm9<sup>Cast</sup>-controlled recombinat hotspots are active (Supplementary Table 4), so hotspots identified in the heterozygous F1 mouse are still informative for meioses occurring in F4 mice. Our findings also confirm that female recombination mainly occurs in hotspots also active in male mice, in which our ChIP-seq data were gathered, although there may be sex-specific differences in hotspot heat15 (Supplementary Fig. 2d).

NCO and CO events both occur in individual recombination hotspots with probabilities approximately proportional to their estimated heat using either DMC1 or H3K4me3 enrichment (Fig. 2a–c, Supplementary Fig. 3f): over 50% of all hotspot-associated F2 NCO or CO events occur in only the 4,000 hottest hotspots (ranked by DMC1 enrichment), around one sixth of all hotspots. Strong dominance of Prdm9<sup>Cast</sup>-controlled over Prdm9<sup>Hum</sup>-controlled hotspots is observed for both event types, and in our ChIP-seq data (Fig. 2d, e and Supplementary Fig. 3a, b). Because binding sites for the humanised allele have not experienced evolutionary hotspot erosion29, this phenomenon cannot explain the dominance of the Prdm9<sup>Cast</sup> allele. Others have recently provided evidence that higher expression of Prdm9<sup>Cast</sup> than Prdm9<sup>Hum</sup> is also unlikely as an explanation40, although we cannot rule out that the dominance of the Prdm9<sup>Cast</sup> allele over Prdm9<sup>Hum</sup> is due to the presence of greater levels of (or more stable) PRDM9<sup>Cast</sup> protein.

Alternatively, or in conjunction, stronger binding affinity of the CAST protein for its motif may underlie the observed dominance. Indeed, both a smaller number and weaker average intensity of PRDM9 (and mirroring H3K4me3) ChIP-seq peaks were recently reported in testes from B6 compared with CAST mice46. Similarly, we observe weaker H3K4me3 enrichment at Prdm9<sup>Hum</sup>-controlled hotspots compared to Prdm9<sup>Cast</sup>-controlled hotspots (Supplementary Fig. 3e). If occurring also in mice, the propensity of PRDM9<sup>Hum</sup> to bind to promoters, which have lower DSB and recombination rates in humans and comprise ~10% of its binding sites16, might also help make the Prdm9<sup>Cast</sup> allele appear even more dominant.

After accounting for sampling variation (Methods), we estimated correlation between recombination rates at different scales (Fig. 2f, g and Supplementary Fig. 3c,d). This released sex differences in recombination rates (combining COs and NCOs to gain power), with 100% correlation excluded, and decreasing correlations at broader scales (Fig. 2f). We observe strong (>70%) correlation between sex-averaged NCO and CO rates, although we also find very strong evidence that these events differ in their positioning along the chromosome, especially at broad scales (Fig. 1e and Supplementary Fig. 2f), and the NCO rate is much higher than the CO rate at all scales.

**Length, number, and positioning of NCO tracts.** We leveraged the high SNP density (~1 SNP per 170 bp) and large number of events in our system to estimate the underlying NCO event tract lengths (accounting for the fact that if a NCO event does not contain a SNP, it is not observed; Methods), separately for hotspots controlled by Prdm9<sup>Cast</sup> and Prdm9<sup>Hum</sup>. The data show relatively good fits to an exponential distribution (Fig. 2h), but with significant differences in estimated mean NCO tract length (p = 0.0018 by bootstrapping): 30 bp for Prdm9<sup>Cast</sup> (95% confidence interval (CI) 25–35 bp), and 41 bp for Prdm9<sup>Hum</sup> (35–48 bp CI). This is unexpected and implies that Prdm9 alleles can differ in basic properties of how recombination events resolve. We confirmed that this difference cannot be explained by a difference in SNP density surrounding hotspots controlled by each allele (Supplementary Note 5). These tract length estimates are at the lowest end of the broad existing estimates for humans and mice25–27,38.

Previous studies using microscopy have reported 200–400 visible DMC1 foci marking individual DSB sites per meiosis in mice38,47,48. However, some of these DSBs, e.g. those occurring on the X-chromosome in males, might repair using the sister chromatid, yielding no converted SNPs. Using our sequencing data, we directly estimated the total number of DSBs that repair from the homologue per meiosis. Given our tract length estimates, we inferred an average total of 300.5 DSBs (95% CI 258.5–370.5) per meiosis repairing using the homologue, 90% of these being NCOs48 (Supplementary Note 6). Estimates are very similar in both F1 and F4 parents. This suggests that the vast majority of the estimated 200–400 DSBs per meiosis undergo homologue-templated repair, rather than sister-templated repair46,51.

Both NCO and CO event centres distribute symmetrically around PRDM9 binding motifs that we identified within hotspots (Methods, Fig. 3a–d and Supplementary Fig. 4a–d). NCO events cluster very near to motifs (potentially overlapping them in 70% of cases; Fig. 3a, d), slightly less strongly than clustering of
Effects of SNP density on GC-biased gene conversion. Our NCO events show strong evidence of AT-to-GC bias, though initially weaker than seen in humans, for both Prdm9Cast-controlled (64%) and Prdm9Hum-controlled (60%) hotspots \( (p < 2 \times 10^{-5}, \text{two-sided binomial test}) \). We next focussed on NCO events within Prdm9Hum-controlled hotspots for further investigation, because the genomic GC-content has not evolved alongside this allele. We tested for a difference in NCO tract lengths, >100 bp from the nearest SNP, showing correlations between (sex-averaged) NCO and CO rates at different scales. As f, but showing correlations between (sex-averaged) NCO and CO rates at different scales. h Decay in probability that nearby SNPs are co-converted, with inter-SNP distance, conditional on a SNP being converted.

mapped DSB breakpoints, but with a far tighter range than the DMC1 and H3K4me3 ChIP-seq enrichment values (Fig. 3e, f). CO events spread more broadly (Fig. 3b, c), consistent with previous studies. Thus, NCO gene conversion appears restricted to sites very close to initiating DSBs themselves, and more distantly positioned NCOs appear to occur only rarely, in contrast to results from a single mouse hotspot.

Fig. 2 DMC1, H3K4me3, and Prdm9 allele predict CO and NCO properties. a DMC1 and H3K4me3 ChIP-seq peaks in a 50-kb region on Chr10 (starting at mm10 position 122871029), with single NCO and CO events overlapping these peaks. b, c the proportion of F2 CO + NCO events at hotspots binned by increasing DMC1 (b) and H3K4me3 (c) ChIP-seq enrichment. d, e Fraction of COs (d) and NCOs (e) controlled by the Prdm9Cast or Prdm9Hum alleles from F2 mice overlapping DMC1 hotspots in the Prdm9 knockout mouse, or non-identifiable (Unknown). f Correlation of underlying total recombination rates between females and males (pooling NCO and CO events), for rates binned at different scales (x-axis); dotted lines show 95% CIs for true correlations. g As f, but showing correlations between (sex-averaged) NCO and CO rates at different scales. h Decay in probability that nearby SNPs are co-converted, with inter-SNP distance, conditional on a SNP being converted.
This implies that local genetic diversity itself influences GC-biased gene conversion at NCOs, and therefore there must be at least two distinct processes operating to repair heteroduplex stretches formed at DSBs, one which is strongly GC-biased, and another which dominates when multiple mismatches exist, and shows no GC-bias.

To further characterise GC-bias, we estimated conversion rates of different types of SNPs in the donor and recipient chromosomes at single-SNP NCO sites (Fig. 4c and Supplementary Note 7). We normalised these relative to their conversion rates in multi-SNP events (Supplementary Fig. 5b), or to flanking SNP composition (Fig. 4c), both of which show...
no GC-bias and gave near-identical results. The simplest model which can explain the data is if there are two distinct conversion rates: a lower rate if the recipient chromosome (i.e. the homologue in which the DSB occurs) carries a G or a C, and higher rate if the recipient carries an A or a T. For example, G/C transversions appear to convert at the lower rate. This could be explained by a model where a GC-biased repair process can resolve heteroduplex DNA in favour of the recipient chromosome if it carries a G or C base—effectively blocking conversion of that base.

Notably, we do not observe a consistent GC-bias for CO events, which are accompanied by long conversion tracts of ~500 bp in size. However, we did observe a very small number of complex CO events (0.6% of all de novo COs observed), which contain non-converted markers surrounded by converted markers resulting from the same meiosis. In humans, complex COs have also been observed and have been shown to have GC-bias. We hypothesised that complex COs and NCOs might result from blocking of conversion of particular markers where the recipient chromosome carries a G or C base, thus generating a GC-bias in the remaining markers which are converted. We observed a total of 12 such non-converted markers within 8 complex CO events, and 7 within 7 complex NCO events. Remarkably, for 18 of these 19 cases the recipient chromosome carries a G or C base (p = 7.6 × 10⁻⁵ by two-sided binomial test). These results suggest that while a GC-neutral mechanism operates for most multi-SNP NCO tracts, producing simple events, occasionally some local, almost 100%
GC-biased mechanism may operate at individual SNPs to generate rare complex NCO and CO events.

CO and NCO depletion when PRDM9 fails to bind the homologue. For NCO events, we can identify on which homologue the underlying DSB occurred. In the F2 mice, we observed a bias: 60% of the observed NCOs were initiated on the B6 background (p < 10⁻³, two-sided binomial test). This is due to the behaviour of the Prdm9<sup>Cast</sup> allele, which accounts for 80% of observed NCO events (Fig. 2e), and which shows a strong preference for binding to the B6 background (Supplementary Fig. 6a; 66% of NCOs, p < 10⁻³, two-sided binomial test), explained by evolutionary hotspot erosion of CAST-controlled hotspots on the CAST genetic background<sup>20,22</sup>. Because PRDM9 binding, DSB evolution, and hotspot erosion or other evolutionary forces<sup>23</sup>, we instead observed a strong depletion of both NCO and CO events in the most asymmetric hotspots (Fig. 2b, c). However, when we grouped Prdm9<sup>Hum</sup> allele binds and initiates recombination events equally on both homologues. Among the most asymmetric hotspots, we observed SNPs or indel polymorphisms within 96% of identified motifs overall (Supplementary Fig. 6), implying their asymmetry is almost always driven by sequence changes disrupting PRDM9 binding on one homologue. Consequently, asymmetric hotspots remain asymmetric in both sexes and between F2 and F5 animals. As in other hybrid mice<sup>23</sup>, we observed that the ratio of DMC1 to H3K4me3 enrichment increases roughly two-fold at asymmetric hotspots compared to symmetric hotspots (Supplementary Fig. 6g), a fact most easily explained by delayed repair of DSBs for which the homologue is not bound by PRDM9<sup>23</sup>. To test whether such asymmetry might also change the outcomes of DSB repair, we measured the numbers of NCO and CO events actually occurring in asymmetric vs. symmetric hotspots, relative to their expectations according to DMC1 and H3K4me3 enrichment (Methods and Supplementary Note 8).

Recombination hotspots can be separated into asymmetric cases where DSBs occur mainly on one homologous chromosome, and symmetric cases where DSBs occur equally on both homologues. Among the most asymmetric hotspots, we observed SNPs or indel polymorphisms within 96% of identified motifs overall (Supplementary Fig. 6), implying their asymmetry is almost always driven by sequence changes disrupting PRDM9 binding on one homologue. Consequently, asymmetric hotspots remain asymmetric in both sexes and between F2 and F5 animals. As in other hybrid mice<sup>23</sup>, we observed that the ratio of DMC1 to H3K4me3 enrichment increases roughly two-fold at asymmetric hotspots compared to symmetric hotspots (Supplementary Fig. 6g), a fact most easily explained by delayed repair of DSBs for which the homologue is not bound by PRDM9<sup>23</sup>. To test whether such asymmetry might also change the outcomes of DSB repair, we measured the numbers of NCO and CO events actually occurring in asymmetric vs. symmetric hotspots, relative to their expectations according to DMC1 and H3K4me3 enrichment (Methods and Supplementary Note 8).

With all else being equal we expect any two groups of hotspots that are matched to have the same total DMC1 or H3K4me3 enrichment to also have similar numbers of CO and NCO events (Fig. 2b, c). However, when we grouped Prdm9<sup>Hum</sup>-controlled hotspots according to their (a)symmetry we instead observed a strong depletion of both NCO and CO events in the most asymmetric hotspots (p = 10⁻²⁷ and p = 10⁻²³, chi-square test, respectively, after controlling for factors in Supplementary Table 6). Furthermore, for DSBs occurring on the less-bound chromosome of asymmetric hotspots (measured by homologue-specific DMC1 enrichment), we found that NCO events occur at the expected rate for symmetric hotspots (Supplementary Note 8). Similarly, we observe that a greater fraction of both CO and NCO events initiate on the less-bound homologue compared to the expectation from H3K4me3 or DMC1 asymmetry (Fig. 3c, d). This implies that when DSBs occur at asymmetric hotspots on the more frequently bound chromosome, the resulting lack of observed NCO or CO recombination events must be specifically driven by a lack of PRDM9 binding to its homologue, rather than any impact of diversity per se, even within PRDM9 binding motifs.

**Discussion**

Our results indicate a sex-averaged NCO rate in mice carrying humanised Prdm9 of around 10⁻⁶ per base, which is strikingly below human estimates of around 4.1 × 10⁻⁶ and 7.7 × 10⁻⁶ in males and females, respectively<sup>27</sup>. We estimate that there are 300.5 total COs and NCOs per meiosis in mice, consistent with previous microscopy studies showing there are 200–400 DSBs per meiotic cell<sup>38,47,48</sup>. This implies that most DSBs in mice undergo homologue-templated repair, and it suggests that the true number of NCOs per meiosis in mice is unlikely to be much larger than our estimate. The greater overall NCO rate in humans may be due in part to the greater frequency of long, complex NCOs in humans, which can each convert a large number of sites<sup>26,27</sup>. These long, complex events can extend up to 100 kb and comprise up to 11.3% of human NCOs detected by sequencing; they were also found to be enriched in females and to increase in frequency with maternal age<sup>27</sup>. We found that similar events are nearly absent in mice (0.5% of all NCO events detected in this study, and never as long as those seen in humans). We suggest that this difference may reflect the timespan of dictyate arrest, which occurs before the completion of female recombination, and lasts decades in humans vs. months in mice. These findings support the hypothesis that complex NCO events in humans might arise from the repair of non-programmed DNA damage occurring over time, consistent with the fact that they are different in other ways: they mainly occur outside PRDM9-controlled hotspots, they are often longer than 1 kb, and they show GC-bias regardless of their size<sup>27</sup>.

Using DMC1<sup>40</sup> and H3K4me3 ChIP-seq data, one can infer that DSBs form at a small subset<sup>12</sup> of PRDM9 binding sites in each cell on average proportionally to the rate of PRDM9 binding to each site<sup>23</sup>. However, we find that the processing, repair and eventual recombination outcomes at each PRDM9 binding site all depend strongly on the sequence of the homologue at that site. In this study, we revealed for the first time that both CO and NCO events are depleted, in both sexes, at asymmetric recombination hotspots (relative to expectations from both H3K4me3 and DMC1 enrichment; Fig. 5 and Supplementary Fig. 7). That is, when the unbroken homologue is not strongly bound and/or marked by PRDM9 (primarily due to motif-disrupting SNPs on the homologue; Supplementary Fig. 6f), the DSB seems less able to repair by homologue-templated recombination. This effect cannot be explained by an increase in genetic diversity alone, as has been suggested<sup>22</sup>; we found that only nearby SNPs that
abolish PRDM9 binding symmetry have any association with the CO or NCO rate (Supplementary Table 6). One potential consequence of the reduction of homologue-templated repair at asymmetric hotspots is the mitigation of hotspot erosion caused by the over-transmission of alleles that disrupt PRDM9 binding20,23.

A recent crossover-mapping study, which sequenced single sperm from the same mouse cross examined here, similarly found CO depletion and DMC1 excess at asymmetric hotspots40. Here we have gone further by demonstrating that both COs and NCOs are depleted at asymmetric sites, showing that many DSBs at asymmetric hotspots are not just delayed in their repair, but they often completely fail to ever repair from their homologue. Moreover, this occurs in female as well as male meiosis. This supports the hypothesis that homology search is the key process disrupted at asymmetric hotspots, rather than downstream events like CO versus NCO repair decisions. One hypothesis is that PRDM9 binding and/or its associated chromatin marks on the unbroken homologue assist the homology search machinery with the challenging task of finding the correct homologous template for repair. This could explain the wider asynapsis and infertility seen in male mice where asymmetric hotspots predominate23,53.

**Fig. 5** COs and NCOs are depleted in asymmetric hotspots in both sexes. a Human-controlled DMC1 hotspots were separated into three bins (asymmetric, intermediate, and symmetric) according to symmetry, so that each bin contains the same number of predicted events according to DMC1 enrichment (Supplementary Note 8). Grey bars show the DMC1-predicted expected fraction of events in each bin. The four coloured bars (vertical lines: 95% CIs) show the observed fraction of (sampled) F5 de novo events: COs, NCOs, and paternal or maternal recombination events. b As a, except predicted events were defined using H3K4me3. c, d Expected vs observed initiation biases at symmetric (c) and asymmetric (d) hotspots, as defined by H3K4me3 symmetry. If asymmetry-related recombination deficiency occurred equally on both homologues, then the fraction of DMC1 reads and the fraction of NCO and CO initiations on the less-bound homologue would be predicted to closely match its fraction of H3K4me3, as in the symmetric case (c). Error bars are 95% bootstrap CIs.
although additional factors must act to explain sex differences in hybrid fertility. There still remains the question of how asymmetric hotspots are repaired, if not from the homologue, in hybrid mice whose fertility is not disrupted. We suggest that asymmetric hotspots sometimes behave like DSBs on the X chromosome in males, which repair late from the sister chromatid and, like asymmetric hotspots, show excess DMC1 enrichment18,23. Because sister-templated repair does not convert any SNPs, we are unable to directly detect such events.

Our results demonstrate that local genetic differences between homologues can also affect the process of GC-biased gene conversion (gBGC) at NCO sites. We confirmed gBGC does not depend on DMC1 enrichment or hotspot symmetry (Supplementary Fig. 5 and Supplementary Data 5), and this implies it must operate downstream of DSB formation, during heteroduplex mismatch resolution. We found that within hotspots in both humans and mice, gBGC acts almost exclusively on potential conversion tracts containing only a single SNP (i.e. mismatch), with essentially identical bias in each species (68% of NCOs convert A/T to G/C; Supplementary Data 5).26,27 In contrast, nearly all multi-SNP NCO tracts are simple stretches of converted markers with no apparent GC-bias, implying more than one pathway for heteroduplex repair. It has been hypothesised that gBGC may function to oppose the mutagenic effects of recombination37; our results suggest that this effect is weakened at sites with high local heterozygosity, which are also predicted to have a higher gene conversion rate.

As one possible model to explain these gBGC observations, we propose that there are two strand-aware heteroduplex repair pathways at NCOs (illustrated in Fig. 6). The first proposed pathway can repair all types of heteroduplex mismatches and always selects the bases copied from the donor homologue—we call this the donor-biased pathway. The second proposed pathway operates only at heteroduplexes containing a single mismatch with a G or C base on the recipient strand, and it always selects the recipient strand’s base—we call this the GC-restoring pathway. All mismatches repaired by the donor-biased pathway will be observable as gene conversions. However, because the GC-restoring pathway restores the recipient allele, its outcome cannot be directly observed (equivalently, we say gene conversion is blocked at that site). While we cannot formally rule out other possible models of NCO repair, our proposed strand-aware model can fully explain our observations of selective GC-bias as well as our number of observed NCO events while only requiring ~300 total DSBs per meiosis. In contrast, a strand-unaware model would require ~465 DSBs per meiosis, which is outside the range of previous experimental observations.38,49,50 (Supplementary Fig. 8b and Supplementary Note 7). Furthermore, at 19 observed CO events within highly (>95%) asymmetric hotspots containing a mutation within their PRDM9 motif, we observe transmission of the cold binding site allele to offspring in 95% of cases (Supplementary Note 7), as has been previously observed at an individual hotspot.50 Therefore, for those longer conversion tracts within CO events at least, a strand-aware mechanism of heteroduplex repair appears to bias repair towards the unbroken donor homologue.

Importantly, both proposed pathways can repair single-GC recipient sites: we calculated that the GC-restoring pathway would need to operate at these sites 53% of the time to account for the observed 68% overall GC-bias among observed events (Fig. 6, Supplementary Fig. 8a, Supplementary Note 7). For example, the GC-biased mechanism might operate only at G or C

---

**Fig. 6** Possible model of NCO mismatch repair pathway choice. NCO repair pathways such as Synthesis-Dependent Strand Annealing (SDSA) form heteroduplex DNA between the recipient homologue’s original sequence (blue) and a small tract of DNA copied from the donor homologue (red). Any mismatching bases must be resolved either by restoring to the recipient allele or converting to the donor allele; only the latter can be detected in offspring. Three possible NCO heteroduplex tracts are depicted, differing in the number and type of mismatch sites (these regions are illustrated as unwound for visual aid purposes only). In all cases, one proposed mismatch repair mechanism (orange arrow) converts site(s) on the recipient chromosome in a donor-strand-biased manner to the allele of the donor chromosome (red). However, in the particular case shown in the middle column—where there is a single mismatch site and the recipient chromosome contains a Strong (S = G or C), not Weak (W = A or T), allele—a different proposed repair mechanism (blue arrow) operates ~53% of the time and restores the G/C recipient allele in a recipient-strand-biased manner, resulting in no observable gene conversions at that site (Supplementary Fig. 8 and Supplementary Note 7). The sum of these two effects can explain the overall 68% bias of observed gene conversions towards G/C.
recipient bases, but not both, which would cap the bias at NCO sites to (at most) 67%, very close to the observed fraction. Almost all observed complex NCOs and COs, though very rare, appear to be explained by this GC-restoring pathway operating on individual SNPs within the tract, with the recipient homologue carrying a G or C base in 95% of such non-converted markers we observed (Supplementary Note 7). Similar behaviour was observed at complex COs within a single human hotspot, suggesting a conserved mechanism across species.27 Because the proposed GC-biased pathway operates at single SNPs and shows strong base-specific and strand-specific biases, it may involve proteins like those found in the base excision repair (BER) pathway, as previously suggested.28 In contrast, the proposed donor-biased repair pathway, which can operate on longer stretches including multiple SNPs, is more consistent with proteins in the mismatch repair (MMR) pathway, several of which are known to be essential for meiosis in mice.24,25

Here we have uncovered the first examples of local genetic diversity affecting both the ability to undergo homologue-templated repair (due to polymorphism altering PRDM9 binding potential and symmetry), and the decision to repair via GC-biased gene conversion (depending on the number and spacing of local sequence polymorphisms), both downstream of DSB formation.

Another unexpected influence on NCO events, again occurring downstream of DSB formation, is the Prdm9 allelic state, with Prdm9Hum-controlled NCOs having an average length 11 bp (37%) longer than Prdm9Cast-controlled NCOs. It is unclear whether this reflects PRDM9 binding directly, or some indirect impact, e.g., how PRDM9 binds relative to nucleosome positions or other genetic or epigenetic features. Further work will need to investigate the exact mechanisms by which PRDM9 binding to the unbroken homologue promotes homologue-templated repair as well as the exact repair pathways that can explain why NCO recombination outcomes depend on local heterozygosity.

**Methods**

**Mouse breeding, library preparation and sequencing.** CAST/Env (CAST) mice were sourced from MRC Harwell (UK). The C57BL/6J (B6) line humanised at the Prdm9Hum locus was generated previously.28 Breeding of B6 and B6Hum mice (F0) was carried out in both directions (using females and males of each type) to generate (B6xCAST)F1 hybrid, heterozygous offspring. To study properties of the humanised Prdm9 allele, we genotyped as previously29 and selected 26 F2 mice of each sex homozygous for humanised Prdm9. We crossed these animals and then offspring for three further generations and examined F5 offspring, comprising two of each sex from each of 18 pairs of F4 parents. One B6Hum mouse, one CAST mouse, 11 F2 mice and all 18 F4/F5 families (36 F4 parents and 72 F5 offspring) were subjected to whole genome sequencing. Genomic DNA was extracted from spleen using the DNAeasy Blood and Tissue Kit (Qiagen) and analysed by direct sequencing information corresponding to the converted sites from the donor mouse, such as the allele depth, and we copied the sequencing information for the background from the recipient, such as mate-pair information. We copied the variant call information rather than reads themselves so all the information we employed in our filtering. Because F2 events are controlled by both Prdm9Hum and Prdm9Cast and F5 de novo events are controlled by Prdm9Hum alone, we performed two sets of simulations by using data from 11 F2 samples and 72 F5 samples. Because most recombination events overlap hotspots, we simulated NCOs in hotspot regions. For each event and tract length, we sampled 500 hotspots with probability proportional to their H3K4me3 enrichment. Within each hotspot, we sampled the centre of the NCO tract according to the distribution of NCOs around PRDM9 motifs after correcting for SNP density, and we sampled its tract length from an exponential distribution with a pre-defined mean tract length (which we varied from 10 to 100 bp with step size 10 and from 150 to 300 bp with step size 50). Sampled NCOs containing 8 SNPs were not counted as potentially detectable. Across these 2000 tracts, different animals possessed different ancestral backgrounds. For each tract in each animal, we checked if any of the other animals had a different ancestral background consistent with a gene conversion event in the first animal. If so, we sampled such a donor mouse (other events were ignored). We copied the sequencing information corresponding to the converted sites from the donor mouse, such as the allele depth, and we copied the sequencing information for the background from the recipient, such as mate-pair information. We copied the variant call information rather than reads themselves so all the information we employed in our filtering.
We calculated our power by dividing the total number of simulated tracts deleted after filtering by the total number of simulated tracts overlapping at least one SNP (Supplementary Fig. 2a, b).

H3K4me3 ChIP-seq. We performed ChIP-seq against H3K4me3 in testes from two litterate 23-week-old male (B6xCAST)F1 Prdm9Hum/Cast mice (C57BL/6j-Prdm9Hum/B6 mother, CAST/Ei) father; two biological ChIP replicates—one testis per mouse—plus one input chromatin replicate from one mouse) as previously described32 with several important modifications that increased ChIP stringency (underlined in the full protocol below). In brief, the testis tissue was removed, the seminiferous tubules were dissociated with tweezers and fixed at room temperature in 5 ml 1% formaldehyde in 1× PBS for 5 min followed by glycine quenching (125 mM final conc.). Following 2× centrifugation (2000 x g), and resuspension in 4 °C 1× PBS, pellets were snap-frozen using dry ice and stored at −80 °C. Frozen pellets were thawed and resuspended in 900 µl 1% SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8.0, 2× protease inhibitors), dounced 20 times and sonicated in 300 µl aliquots in a Biouraptor Twin (Diagonode) sonication bath at 4 °C for three 5-min periods of 30 s on, 30 s off at high power, then cell debris was pelleted (14,000 x g, 15 min, 4 °C) and removed and aliquots were pooled. Sonicated lysates were diluted 1:10 in cold IP wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate, filtered with a 0.45 µm filter unit) for antibody incubation. Magnetic beads were washed by adding 50 µl Invitrogen Sheep Anti-Rabbit Dynabeads (Thermo Fisher Scientific) per sample to 950 µl cold PBS/BSA (1× PBS, 5 mg/ml BSA, one tablet Roche Complete protease inhibitor), mixed with 0.45 µm microcon filter. Bead solutions were placed on a magnetic rack and resuspended in 1 ml PBS/BSA four times. The washed beads were resuspended in 100 µl PBS/BSA per sample and added to the diluted chromatin samples for pre-clearing for 2 h at 4 °C with rotation. Beads were removed, and 100 µl of pre-clear ed chromatin was set aside for the input control. Five microcon filtered rabbit polyclonal anti-H3K4me3 antibody (Abcam ab8580) were added to the remaining 200 µl pre-clear ed and incubated overnight at 4 °C with rotation. 50 µl beads were washed and resuspended in 1× PBS/BSA as before, then incubated with the diluted chromatin samples for 2 h at 4 °C with rotation. Beads were then washed three times for 5 min each with 10 ml cold LiCl Wash Buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate, filtered with a 0.45 µm filter unit), then washed once with cold 1X TE (10 mM Tris-HCl pH 7.5, 0.1 mM Na2-EDTA) Bead. Pellets were resuspended in 200 µl room-temperature IP elution buffer (1% SDS, 0.1 M NaHCO3, filtered with a 0.45 µm filter unit) and vortexed to mix. Samples were incubated in a 65 °C water bath for 1 h with 20 µl Proteinase K and 15 min intervals to uncouple beads from protein-DNA complexes. Samples were centrifuged (14,000 rpm, 3 min) and placed on a magnet to pellet beads, and supernatants were isolated and then incubated in a 65 °C water bath overnight to reverse crosslinks. For the input control, 50 µl of pre-clear ed chromatin was processed in parallel. After de-crosslinking, samples were further incubated with 80 µg RNase A at 37 °C for 60 min and then with 80 µg Proteinase K at 55 °C for 90 min. DNA was purified using a Qiagen MinElute reaction cleanup kit and quantified using a Qubit High Sensitivity DNA kit (Thermo Fisher Scientific). This yielded roughly 1 ng of ChIP DNA per testis.

ChIP and total chromatin DNA samples were sequenced in multiplexed paired-end Illumina HiSeq2500 libraries (rapid run), yielding 252 million 51-bp read pairs. We processed the data following the algorithm provided by Kihl et al.33 to map the reads to mm10 and obtain type I reads. We then called DMC1 peaks as previously described33. We defined NCO and CO events as occurring within hotspots if they were less than 1 kb away from either a DMC1 peak or an H3K4me3 peak (covering 4% of the genome).

Testing for correlation rates at chosen scales. We assumed that an unknown vector of (binned) mean values \( W_l \) along the genome, the dense ChIP stringency-related quantity (number of observed recombination or NCO events in various classes), \( N_{\text{arc}} \) follows a Poisson distribution with mean \( W_{\text{arc}} \), in interval \( l \). The \( W_{\text{arc}} \) means vary along the genome and represent the underlying recombination rate parameters; this model is accurate provided (as is likely to be the case for a single meiosis, the number of expected events in each bin is small. Then the variance

\[
\text{Var}(W_{\text{arc}}) = \text{Var}(N_{\text{arc}}) - \text{E}(N_{\text{arc}}) \tag{1}
\]

This enabled estimation of the variation in recombination rate along the genome, using the usual standard estimates of the mean and variance of the number of events, across bins genome-wide.

\[
\text{Var}(W_{\text{arc}}) = \text{Var}(N_{\text{arc}}) - \text{E}(N_{\text{arc}}) \tag{2}
\]

Further, the covariance

\[
\text{Cov}(W_{\text{arc}}, N_{\text{arc}}) = \text{E}(\text{Var}(N_{\text{arc}})) - \text{E}(N_{\text{arc}})^2 \tag{3}
\]

Combining these results enabled estimation of the underlying correlation between \( W_l \) and \( N_{\text{arc}} \) along the genome based on properties only of the observed poisson counts \( N_{\text{arc}} \) and \( N_{\text{arc}} \).

\[
\text{Cov}(W_{\text{arc}}, N_{\text{arc}}) = \text{E}(\text{Var}(N_{\text{arc}})) - \text{E}(N_{\text{arc}})^2 \tag{4}
\]

The quantities in the above equation are all estimated in the usual way using standard estimates of mean and variance from the observed vectors of counts. At any interval size scale, we bootstrap re-sampled (10,000 times) the resulting disjoint intervals of the genome, to compute \( \hat{C} \)s for the estimator.

Motive analysis. We used a Bayesian, ab initio motif finding algorithm to identify motifs within DSB hotspot16,23. For each DSB hotspot that is controlled by Prdm9Cast/−, a 1,000 bp sequence (centred on the hotspot centre) was extracted from the reference sequence (mm10). Ab initio motif identification was performed on the centre 600-bp sequences from the top 1,000 hotspots (ranked by DMC1 enrichment) that contained no overlaps. We first defined a set of positions that were closer than 1 kb away from either a DMC1 peak or an H3K4me3 peak, and then used the Altemose et al. algorithm42 (these are reported in Supplementary Data 2). Motif calling proceeded in two stages: seeding motif identification, and motif refinement. Each seeding motif was obtained by first counting all 10-mers present in all input sequences, and from the top 50 most frequently occurring 10-mers, the one with the greatest overall representation in the central 300 bp of each peak sequence was chosen as the seeding 10-mer. We then defined a refined motif for each peak as described in Davies et al.23. This refined motif was then force-called on the full set of the hotspots (without filtering) by re-running the refinement algorithm, providing a probability of motif occurrence within each hotspot, and also identifying the most likely motif location in each case. This motif was reported for each peak, along with position and strand. We did the same for DSB hotspots controlled by Prdm9Cast and a 48-bp human motif was identified. We identified distinct sequence motifs, and their locations, within 97% of hotspots controlled by Prdm9Cast and 74% of hotspots controlled by Prdm9Hum (Supplementary Fig. 4a,b,23). We used the SNPs generated as described above to determine whether each motif contains a SNP within its span. The distance from a motif to an event was defined as the distance from the centre of the motif to the nearest converted marker (lower bound for NCOs), or zero if a converted marker fell within the motif itself. We associated events < 1 kb from a motif with that motif-containing hotspot.

Estimation of NCO tract lengths. To estimate NCO tract length, we assumed the converted tract follows an exponential distribution with rate parameter \( \lambda \), where \( 1/ \lambda \) is the mean tract length. While exponential tract lengths are not a fully accurate model, we can view this as a summary of tract properties, estimating the probability of co-conversion of pairs of markers as the distance between them increases. We computed a composite likelihood function for our NCOs and estimated \( \lambda \) via maximal likelihood. Specifically, for each converted site, viewing this site as a focal site, we examined the SNPs nearby and recorded for each SNP its distance from the focal SNP, and whether that SNP was also converted. If the SNP was also converted, then it was still in the gene conversion tract, otherwise it was not. Using this approach allowed our approach to be independent of SNP density, because we condition our SNP positions in our analysis. The probability that a SNP nearby a converted site is also converted is

\[
\text{Pr(SNP nearby converted) = Pr(in) = e^{-\lambda d} \tag{5}}
\]
where $d$ is the distance from the nearby SNP to the converted site. We multiply these probabilities across all focal/nearby SNP pairs to obtain the (composite) likelihood of the data:

$$
\Pr(D) = \prod_{i=1}^{n} \Pr(i|\text{SNP})(1 - \Pr(i|\text{SNP}))^{1 - x}
$$

where $x = 1$ if the SNP nearby is also converted and $x = 0$ otherwise. By maximising the likelihood using grid search for $1/\lambda$ from 1 to 1000 with step 0.1, we gained an estimate of tract length. Although distinct NCO events are reasonably modelled as independent, because pairs of SNPs within the same NCO event are not in linkage disequilibrium, this is not a true likelihood (though the resulting estimator is statistically consistent as the number of independent conversion events increases). Therefore, to estimate uncertainty in the resulting estimates, we utilised bootstrapping of NCO events.

To perform bootstraps, we separated autosomal genotypes into 258 non-overlapping 10 Mb blocks (the last block in each chromosome is shorter than 10 Mb). We re-sampled 258 blocks with replacement, where the probability of sampling each block is proportional to the length of that block, and from the resulting bootstrapped set of NCOs, re-estimated tract length via the same procedure. CIs were calculated from a total of 10,000 bootstraps. We implemented this procedure for two sets of NCO events: those overlapping human-controlled and those overlapping CAST-controlled hotspots, respectively. Specifically, we used events that overlap a DM1/H3K4me3 peak to avoid using any false positives. For Prdm9Hum, we used de novo F5 NCOs along with inherited F5 NCOs controlled by Prdm9Hum (815 event total). For Prdm9Cast, we used F2 and inherited F5 NCOs controlled by Prdm9Cast (409 events total).

**Hotspot symmetry estimates.** Sequence differences between the CAST and B6 genomes allowed us to quantify the fraction of ChIP-seq enrichment (either DM1 or H3K4me3), coming from the B6 and CAST chromosomes. This also allowed us to determine whether individual hotspots in these hybrids were symmetric, with DSBs occurring equally on both chromosomes, or asymmetric, with a preference towards either the CAST or B6 chromosome. Using SNPs distinguishing the B6 and CAST genomes, each type I read pair from a hybrid DSB library (DM1 ChIP-seq) was assigned to one of the categories “B6”, “CAST”, “unclassified” or “uninformative” as in ref. 25, replacing PWD (Mus musculus musculus PWD/PH) with CAST. For each CAST DSB hotspot, the B6 cutting ratio was then computed as the fraction of B6 reads mapped within 1 kb of the hotspot centre, over the sum of B6 and CAST reads in that region. We followed a similar approach for H3K4me3 ChIP-seq, further correcting for background as in ref. 25. For both DM1 and H3K4me3, we required ≥10 informative reads to define the B6 cutting ratio.

To order hotspots based on their symmetry, if the fraction of cuts estimated on the B6 and CAST chromosomes were $x$ and $1-x$, respectively, we defined the overall hotspot symmetry as $4 \times (1-x)$, which ranges from 0 for hotspots with events completely on one chromosome to 1 for hotspots with events occurring equally on both chromosomes25. We obtained additional results for events initiating on a known homologue using the log-transformed heat of hotspot symmetry, defined as $hX$, where $h$ is the estimated total heat of the hotspot, for events initiating on the CAST chromosome, and $(1-x)h$ for events initiating on the CAST chromosome (Supplementary Note 8). Separate estimates of hotspot symmetry and homologous heat may be obtained from both H3K4me3 and DM1 ChIP-seq data, for the same collection of hotspots. Because the H3K4me3 homologous heat captures how well the homologous chromosome is bound by PRDM9, it may be of stronger direct interest; however, homologous heat captures how well the homologous chromosome is bound by PRDM9, it may be of stronger direct interest; however, homologous heat may be obtained from both H3K4me3 and DMC1 ChIP-seq (Supplementary Note 8). Separate estimates of hotspot symmetry and

**Estimating fraction of hotspots containing disrupted motifs.** To estimate the proportion of hotspots of different levels of initiation on B6/CAST chromosomes containing SNPs within their PRDM9 binding motifs, we filtered to include only hotspots containing a clear motif (posterior probability > 0.99), and at least 20 informative reads in our DM1 data in order to accurately estimate the proportion of reads from B6, and 5 sequencing reads from each homologue covering the motif region, to provide power to identify variants if present. Supplementary Fig. 6f shows the distribution of hotspots in each binned level of initiation on the B6 chromosome containing a SNP or Indel (using GATK prior to VQSR, or Platypus). Ninety-six percent of identified highly asymmetric hotspots (B6 initiation < 5% or >95% and $p < 10^{-10}$ for binomial test of asymmetry) contained such a polymorphism.

**Testing whether asymmetry or SNP density affects repair.** We fitted a generalised linear regression model to discern whether hotspot asymmetry or local SNP density better predicts variation of CO and NCO rates depending on genetic variation. For each hotspot containing an identified PRDM9 binding motif, we produced a binary response vector indicating whether an overlapping CO event occurred and fit a binomial generalised linear model. As model predictors, we used:

1. The symmetry of the hotspot from DMC1
2. The log-transformed heat of the hotspot estimated by H3K4me3 enrichment (the H3K4me3 enrichment is incremented by a small value 0.001 as there are a few hotspots with zero estimated enrichment)
3. SNP densities around the PRDM9 binding motif at different scales (±100 bp, ±500 bp, ±800 bp)

We then tested each coefficient for significance, conditional on the others. We separated the analysis for Prdm9Cast-controlled COs (all generated in the meiosis from F1 where there are two different Prdm9 alleles) and de novo Prdm9Hum-controlled COs (all generated in the meiosis from F4 where there is only one type of Prdm9 allele) to eliminate any effects of competition between the two alleles. Conditional on H3K4me3 enrichment and hotspot symmetry, SNP density has no significant effect on where COs happen (p-values from all three scales >0.08) while both H3K4me3 enrichment and hotspot symmetry have significant positive effects on CO events conditional on SNP density ($p < 0.05$). For NCO events, we performed a similar analysis, except that we corrected for power to detect NCOs by re-sampling the above hotspots according to the weight generated as described in Supplementary Note 8. We note that some hotspots appeared several times after rejection sampling. Again, all scales showed no significant effect of SNP density conditional on H3K4me3 enrichment and hotspot symmetry ($p > 0.2$). For Prdm9Hum-controlled NCOs, results show that H3K4me3 enrichment and hotspot symmetry have significant positive effects on NCOs conditional on SNP density ($p < 0.003$). Results from Prdm9Cast-controlled NCOs also suggest positive effects on prediction of NCOs, but p-values do not reach significance due to the smaller number of these events (<0.2). We discuss the weaker effect of symmetry for Prdm9Cast-controlled NCOs in Supplementary Note 8.

**Ethical compliance.** All experiments involving research animals received local ethical review approval from the University of Oxford Animal Welfare and Ethical Review Body (Clinical Medicine board) and were carried out in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability** The mouse WGS sequencing data and variant calls are available under the SRA accession SRP189907. The H3K4me3 ChIP-seq reads and peak calls are available under the GEO accession GSE119727. DM1 ChIP-seq data40 were downloaded from GEO accession GSE124941. H3K9me3 ChIP-seq data44 from the broad-scale GLM analysis were downloaded from GEO accession GSE61613. The source data underlying Supplementary Tables 2 and 3 are provided as a Source Data file. Any additional data are available request from the authors.

**Code availability** The computer code developed for the analysis of the datasets in the current study is available here: https://github.com/rosaana leagues/Detect_recombination_events_from_mouse_pedigree

Received: 13 October 2018 Accepted: 17 July 2019
Published online: 29 August 2019

## References

1. Paigen, K. & Petkov, P. Mammalian recombination hot spots: properties, control and evolution. *Nat. Rev. Genet.* **11**, 221–233 (2010).
2. Lichten, M. Meiotic recombination: breaking the genome to save it. *Curr. Biol.* **12**, 253–256 (2001).
3. Baker, Z. et al. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *Elife* **6**, e24133 (2017).
4. Parpanov, E. D., Petkov, P. M. & Paigen, K. Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
5. Baudat, F. et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Genome Res.* **22**, 360–366 (2010).
6. Berg, I. L. et al. PRDM9 variation strongly influences recombination hotspot activity and meiotic instability in humans. *Nat. Genet.* **42**, 859–863 (2010).
7. Myers, S. et al. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**, 876–879 (2010).
22. Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D. & Petukhova, G. V. The rate of meiotic gene conversion varies by sex and unexpected clustering in humans. Genetica 182, 1335–1344 (2009).

23. Shifman, S. et al. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. 4, 2227–2237 (2006).

24. Bruschwig, H. et al. Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetica 191, 757–764 (2012).

25. Axelsson, E., Dehal, P.S. & Prakash, S. Meiotic recombination hotspots coincide with stabilization of the recombination landscape in the dog genome. Genome Res. 29, 51–63 (2012).

26. Patel, L. et al. Dynamic reorganization of the genome shapes the recombination landscape in meiotic prophase. Nat. Struct. Mol. Biol. 26, 164–174 (2019).

27. Grey, C. et al. In vivo binding of PRDM9 reveals interactions with noncanonical genomic sites. Genome Res. 27, 580–590 (2017).

28. Cole, F. et al. Homoeostatic control of recombination is implemented progressively in mouse meiosis. Nat. Cell Biol. 14, 424–430 (2012).

29. Paigen, K. & Petkov, P. Meiotic DSBs and the control of mammalian recombination. Cell 122, 1624–1626 (2005).

30. de Boer, E., Jasim, M. & Keyney, S. Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. Genes Dev. 29, 1721–1733 (2015).

31. Cole, F., Keyney, S. & Jasim, M. Preaching about the converted: how meiotic gene conversion influences genomic diversity. Ann. N.Y. Acad. Sci. 1267, 95–102 (2012).

32. Zickler, D. & Kleckner, N. Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harb. Perspect. Biol. 7, a016626 (2015).

33. Hall드erson, B. V. et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science 363, eaui1043 (2019).

34. Gregorova, S. et al. Modulation of Prdm9-controlled meiotic chromosome anaphase overrides hybrid sterility in mice. Elife 7, e43282 (2018).

35. de Vries, S. S. et al. Mouse Mut5-like protein Mbd3 is required for proper chromosome synapsis in male and female mice. Genes Dev. 33, 532–533 (1999).

36. Knetz, B. et al. Mut5 homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. Genes Dev. 14, 1085–1097 (2000).

37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

38. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–939 (2011).

39. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).

40. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).

41. Hinch, A. G. et al. Factors influencing the generation of recombination hotspots. Nature 479, 700–710 (2011).

42. Patro, R. et al. Recombination initiation maps of individual human genomes. Science 346, 1256442 (2014).

43. Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D. & Petukhova, G. V. The evolutionary turnover of recombination hot spots contributes to speciation in mice. Genes Dev. 30, 266–280 (2016).

44. Davies, B. et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature 530, 171–176 (2016).

45. Chambers, S. R., Hunter, N., Louis, E. J. & Borts, R. H. The mismatch repair system reduces meiotic homologous recombination and stabilizes recombination-dependent chromosome loss. Mol. Cell. Biol. 16, 6110–6119 (1996).

46. Cole, F. et al. The complex landscape of recombination in African Americans. Nat. Genet. 47, 176 (2015).

47. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).

48. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–939 (2011).

49. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).

50. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).

51. Hinch, A. G. et al. The landscape of recombination in African Americans. Nat. Genet. 47, 176 (2015).

52. Cole, F. et al. The complex landscape of recombination in African Americans. Nat. Genet. 47, 176 (2015).

53. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).

54. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–939 (2011).

55. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).

56. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).

57. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–939 (2011).

58. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).

59. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).

60. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21, 936–939 (2011).

61. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).

62. Yang, H. et al. A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6, 663–666 (2009).
**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11675-y.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Peer review information:** Nature Communications thanks Ian Henderson, Molly Schumer and Irene Tiemann-Boege for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.