Chromosome-wide distribution and characterization of intersubspecific meiotic noncrossovers in mice

Vaclav Gergelits^{1,2,*}, Emil Parvanov¹, Petr Simecek¹ and Jiri Forejt^{1,*}

¹Laboratory of Mouse Molecular Genetics, Division BIOCEV, Institute of Molecular Genetics,
Czech Academy of Sciences, Vestec, Czech Republic CZ-25250;

²Charles University, Faculty of Science, Prague, Czech Republic CZ-12000,

*Corresponding authors

Abstract

During meiosis, the recombination-initiating DNA double-strand breaks (DSBs) are repaired by crossovers or noncrossovers (gene conversions). While crossovers are easily detectable, the noncrossover identification is hampered by the small size of their converted tracts and the necessity of sequence polymorphism to detect them. We report identification and characterization of a mouse chromosome-wide set of noncrossovers by NGS sequencing of 10 mouse chromosome substitution strains. Based on 94 identified noncrossovers we determined the mean length of a conversion tract to be 32 basepairs. The spatial chromosome-wide distribution of noncrossovers and crossovers significantly differed, though both sets overlapped the known hotspots of PRDM9-directed histone methylation and DNA DSBs, thus proving their origin in the standard DSB repair pathway. A significant deficit of noncrossovers descending from asymmetric DSBs pointed to sister chromatids as an alternative template for their repair. The finding has implications for the molecular mechanism of hybrid sterility in mice.

Introduction

Homologous recombination between paternal and maternal chromosomes and their synapsis are unique features of the meiotic phase of gametogenesis in the majority of the sexually reproducing organisms. In humans, mice and most other mammals the localization of recombination breakpoints is under the control of the PRDM9 histone methyltransferase. The PRDM9 zinc finger domain binds at allele-specific sites of genomic DNA and its PR/SET domain ensures trimethylation of histone 3 at lysine 4 and lysine 36 (H3K4me3 and H3K36me3) surrounding both sides of the binding motif¹⁻⁴. The SPO11 topoisomerase-like protein targets a subset of trimethylated PRDM9 binding sites and generates programmed DNA double-strand breaks (DSBs)^{5,6}. The DSBs are repaired by homologous recombination, either by crossovers (COs) accompanied by the exchange of chromatid arms between homologous chromosomes or by noncrossovers (NCOs, gene conversions) created by unidirectional transfer of genetic information from the homologous chromatids. Conversion tracts, usually longer than in NCOs, also accompany COs⁷. The COs are under a strong homeostatic control ensuring at least one CO per chromosome arm (obligate crossover rule)^{8,9}, a condition necessary for the proper segregation of homologous chromosomes into gametes. In contrast, NCOs are indispensable for synapsis of homologous chromosomes as illustrated by asynapsis in spermatocytes with reduced or absent DSBs due to attenuated or abolished expression of the *Spo11* gene¹⁰⁻¹³. The significance of NCOs for meiotic synapsis of homologous chromosomes is further corroborated by the synapsis failure in meiocytes of the mouse subspecific hybrids with a reduced number of repairable (symmetric) DSBs¹⁴⁻¹⁶. The NCO events are hard to detect because the conversion tracts are rather short, estimated at 86 ± 49 bp in two mouse hotspots¹⁷ and 55-290 bp in humans¹⁸. Moreover, the NCOs are detectable only when the conversion tracts carry SNPs or short indels distinguishing the

paternal from the maternal sequence. Thus, the efficiency of identification of NCOs depends on the density and distribution of these allelic variants.

To identify the NCOs and their chromosome-wide distribution we took advantage of the fact that NCOs are enriched in consomic chromosomes of chromosome substitution (consomic) strains. To construct a consomic strain only the individuals with nonrecombinant donor chromosomes (without COs) are selected for transfer to the next generation during 10 consecutive backcrosses to the recipient inbred strain^{19,20} (Figure 1). Hence the transferred chromosome is 10-fold enriched for NCOs but devoid of COs. In the final intercross the consomic chromosome is fixed in the homozygous state allowing direct identification of NCOs by sequencing. Here we searched for NCOs in 10 strains of a panel of C57BL/6J-Chr #^{PWD}/Ph/ForeJ (abbreviated B6.PWD-Chr #) mouse consomic strains²⁰. In these mice the donor chromosome descends from the PWD/Ph (PWD) inbred strain derived from the *Mus musculus musculus* (*M. m. musculus*) subspecies²¹, while the genome of the recipient C57BL/6J (B6) strain is predominantly of *Mus musculus domesticus* (*M. m. domesticus*) origin²². Because both subspecies diverged roughly 350-500 thousand years ago^{23,24} the high level of their genome disparity, one SNP per 130 bp, is comparable to the divergence between the human and chimpanzee²⁵. The accumulation of NCOs during the backcrosses and the higher chance to identify them by converted SNPs are the factors that considerably amplified the efficiency of NCO mapping. Moreover, better characterization of NCOs can contribute to genetic dissection of reproductive isolation between the mouse subspecies. The difficulties to repair DNA DSBs by homologous recombination and subsequent deficiency of NCOs was proposed to interfere with the synapsis of homologous chromosomes and cause early meiotic arrest in the (PWD × B6)F1 model of hybrid male sterility^{14-16,26,27}.

Results

Chromosome-wide identification of noncrossovers

To identify NCOs we sought the B6-PWD SNPs in the PWD sequence of individual consomic chromosomes (Figure 2a). Nine B6.PWD-Chr # consomic strains carrying PWD autosomes # 1, 2, 3, 4, 5, 6, 17, 18, 19 and B6.PWD-Chr X.1s subconsomic strain with the proximal part of Chr X of PWD origin²⁸ were sequenced with the minimal coverage of 24×. The B6 SNPs were detected by simultaneous alignment of the whole genome sequence (WGS) Illumina reads of a consomic sequence to the B6 reference genome, the pseudo-reference of PWK/PhJ²⁹, and were compared to WGS of the PWD/PhJ and PWK/PhJ strains. The latter two are closely related *M. m. musculus*-derived inbred strains. The false-positive variants were eliminated by a multi-step filtering procedure based on the comparison of PWD-B6 SNPs in consomic strains and their B6 and PWD progenitors (see Materials and Methods). Altogether we detected 91 NCOs in nine consomic autosomes and three NCOs in the proximal 69.6 Mb of Chr X (Figure 2b, Table 1). Sixty-nine percent of NCO gene conversion tracts encompassed one converted SNP; 20% and 11% carried two and more SNPs (or possibly short indels), respectively.

The ranges of possible length of each NCO conversion tract (Table 1) were based on the converted and surrounding non-converted SNPs. The conversion tracts varied in length. For instance, we found 13 tracts necessarily shorter and four tracts necessarily longer than 100 bp (Figure 3). Assuming exponential distribution, we estimated the mean length of the detected NCO conversion tracts at 32 bp. In addition, a single complex conversion event (1%) on Chr 2 contained two converted SNPs 231 bp apart from each other with two non-converted SNPs in between. We confirmed by Sanger sequencing that this complex event

more likely resulted from a single meiosis rather than from aggregation of two events in different backcross generations. Thus, the incidence of complex NCO gene conversions was much lower than 46% or 65% reported in the human study³⁰.

A representative subset (48/94) of identified autosomal NCOs was further validated by Sanger sequencing and all of them were confirmed to be true, proving the 100% specificity of the filtering methods.

NCOs colocalize with the sites of meiotic DSBs

The meiotic recombination pathway is initiated by activation of PRDM9 binding sites, detectable by chromatin immunoprecipitation of histone 3 trimethylated at lysine 4 and lysine 36 (H3K4me3 and H3K36me3 ChIP-seq) and massively parallel DNA sequencing. About 4700 such sites are activated per leptotene spermatocyte³¹, but only 250-300 acquire DNA DSBs after being targeted by the SPO11 protein. Hotspots of DNA DSBs can be localized and quantified by sequencing SPO11-covalently bound oligos³² or by ChIP-seq of DMC1 protein bound DNA recombination intermediates³³. At kilobase resolution, 96% of the identified NCOs colocalized with the sites of meiotic DMC1 hotspots based on the available data^{15,34} on males of one or more of the strains/crosses – B6, PWD, (B6 × PWD)F1 and (PWD × B6)F1 (Table 1). The overlap showed high specificity of our detection approach, since the random overlap of NCOs with the superset of these DSBs sites was only 7%, thus confirming that the detected NCOs were real. As expected, the same 96% of NCO conversion tracts colocalized with the PRDM9-induced H3K4me3 histone modifications (Table 1-table supplement 1). The NCO counts per chromosome correlated well (Spearman's $\rho = 0.90377$, $P = 0.0008274$) with the SPO11 oligo signal³⁵ used as a proxy for the DSB counts (Table 1-table supplement 2), reflecting their dependence as shown in detail below. Information for the H3K36me3 and

SPO11 oligo hotspots is not available for (PWD × B6) and (B6 × PWD)F1 hybrids; nevertheless in the subset of NCO sites overlapping the DMC1 hotspots of the B6 strain (72%, 68/94) we assessed the overlap with the PRDM9-mediated H3K36me3 hotspots and the overlap with SPO11 sites of programmed DSBs in B6 meiocytes³⁵. NCOs overlapped with H3K36me3 in 79% of cases (all DMC1 sites overlapped H3K36me3 sites in 56%) and NCOs overlapped with SPO11 in 96% cases (all DMC1 sites overlapped SPO11 sites in 72% cases). The higher overlap of NCOs with H3K36me3 and SPO11 sites could be explained by the fact that NCOs are typically localized in the hotter DMC1 sites more likely reflecting higher affinity for SPO11 and H3K36me3. The high overlap of NCOs with the markers of meiotic DSB induction and repair showed that the vast majority of NCOs resulted from the standard meiotic DSB repair pathway. The four DMC1-negative NCO sites, non-overlapping any DSB site in DMC1 datasets of (B6 × PWD)F1, (PWD × B6)F1, B6 and PWD were confirmed by Sanger sequencing to be converted.

Identification of PRDM9 binding motifs within NCO sites

We identified the PRDM9 binding motif within an interval 500 bp upstream or downstream of the converted SNP in most of the NCOs conversion tracts. When the NCOs were compared with the DMC1 hotspot data, 78% matched with the PRDM9^{B6} hotspots, 5% with PRDM9^{PWD}, 11% possibly originated from either PRDM9^{B6} or PRDM9^{PWD} but could not be determined, 2% were default (present in the B6.*Prdm9*^{-/-} dataset) and 4% of NCOs did not match with any DMC1 hotspot (Table 1-table supplement 3). The inferred PRDM9 alleles responsible for the identified NCOs (Figure 4) were fully consistent with the *Prdm9* genotypes of the particular backcross generation of the emerging consomic strains. While the F1 parents were *Prdm9*^{B6/PWD} heterozygous, the *Prdm9*^{B6/B6} homozygosity was achieved no later than at BC3

in all available lines with the obvious exception of the B6.PWD-Chr 17 line (*Prdm9* is localized on Chr 17, Figure 4-table supplement 1). Remarkably, we detected two NCOs initiated by PRDM9^{B6} on Chr 17 (PWD), while only one NCO initiated by PRDM9^{PWD} (Figure 4). Despite the limited sample size, the ratio was in agreement with the previous finding that 71.8% of DMC1 signal occurred in asymmetric DSB hotspots in (PWD × B6)F1 or (B6 × PWD)F1 hybrids¹⁵. The number of BC generations with the *Prdm9*^{B6/PWD} or *Prdm9*^{B6/B6} genotype corresponded well to the proportion of NCOs initiated by the respective alleles ($r = 0.80$, $P < 0.0001$).

Chromosome-wide colocalization of NCO and CO events

To compare the localization and properties of NCOs and COs we genotyped 102 (B6 × PWD)F1 male-derived meioses using MegaMUGA genotyping arrays. The identified 1486 autosomal COs were mapped with ~80 kb median resolution sufficient for monitoring the CO spatial distribution at the chromosomal level. Out of the total, 627 (42%) CO sites were found repeatedly (2 to 13 times) within a single interval defined by MegaMUGA, showing the presence of an active CO hotspot.

The COs were highly abundant in the subtelomeric regions and at the 2nd decile of the chromosomal lengths (Figure 5a) in agreement with the non-uniform spatial distribution previously described in male meiosis of mice^{36,37} as well as of humans and other species³⁸⁻⁴⁰.

The spatial chromosome-wide distribution of NCOs was significantly different from the CO distribution ($P \leq 0.0109$, see Materials and Methods). In contrast to COs, a significant deficit of NCO events was observed in the 10th decile of the chromosomal length ($P \leq 0.0333$).

Moreover, NCOs were more abundant around the 4th decile of the chromosomal length (Figure 5b). The NCOs were compared to two unrelated datasets of COs, both showing a

difference. First, NCOs were compared to the COs in (B6 × PWD)F1 male meioses detected in this study ($P = 0.0029$, permutation test) to control for the genetic background. Second, NCOs were compared to a dataset of male and female COs from the G2:F1 population of the Collaborative Cross project³⁷ ($P = 0.0109$, permutation test, see Materials and Methods) to account for the previously described sex differences in the CO distribution. Consistently in both comparisons, the frequency of NCO events in the 10th decile of the relative chromosomal length was lower than that of CO events ($P = 0.013$, $P = 0.0333$, respectively, permutation test adjusted for multiple testing). Since none of the published chromosome-wide distributions related to recombination activity including SPO11-oligos in B6³⁵, DMC1 heat and H3K4me3 enrichment in B6 and (B6 × PWD)F1^{15,34} had comparable abundance of signal at the subtelomeric region, the resolution of DNA DSBs by COs is more probable than by NCOs at the subtelomeric regions compared to the other parts of the genome. To compare the CO and NCO outcomes and their PRDM9 initiation with a kilobase resolution we focused on a subset of (B6 × PWD)F1 COs whose position determined by the MegaMUGA chip overlapped a single DMC1 (B6 × PWD)F1 hotspot^{15,34} (hereafter DMC1-unique CO sites). We assumed that the respective COs were initiated in the centers of such DMC1 sites. Limiting the DMC1-unique CO sites to those of the inferred PRDM9^{B6} origin, the COs were found more frequently in the NCO sites than in the sites without detected NCOs ($P = 0.0401$, Mann-Whitney test), suggesting colocalization of CO and NCO hotspots.

Higher H3K4me3 and DMC1 hotspot activity in NCO and CO sites

To search for the differences between CO and NCO resolution pathways, we compared histone 3 lysine 4 trimethylation and DMC1 protein abundance at the CO, NCO and DSB sites. The CO sites were filtered to the above-defined subset of DMC1-unique CO sites. NCO

sites were filtered to those overlapping DSB sites (DMC1) in the (B6 × PWD)F1 database also used as a resource of DSB sites. We assumed the localization of CO, NCO and DSB sites in the DMC1 site center and we filtered for the sites initiated by PRDM9^{B6} to take into account the differences between the models for CO, DSB and NCO detection. The (B6 × PWD)F1 DMC1 heat was higher in both NCOs (median = 176) and COs (203) than in general DSB (40.4) sites (Figure 5-figure supplement 1a). Similarly, the H3K4me3 enrichment was higher in both NCOs (median = 1.33) and COs (1.70) than in general DSBs (0.46) (Figure 5-figure supplement 1b) but to a lesser extent. The higher DMC1 and H3K4me3 signal in both COs and NCOs is consistent with the higher frequency of the CO and NCO events and thus higher chance of their detection.

The COs, DSBs and NCOs overlapped to a comparable extent with genes (47%, 54%, 54%, respectively). In particular, the overlaps with the gene promoters (3%, 3%, 3%) and exons (3%, 5%, 6%) were comparable as well.

Asymmetry of PRDM9 binding and deficiency of nonparental NCOs support the DSB repair mechanism of hybrid sterility

During species evolution the PRDM9 hotspots are gradually eroded due to the biased gene conversion that favors the PRDM9 binding sites with lower affinity^{3,41-43}. In hybrids between mouse subspecies *M. m. musculus* represented by the PWD mouse strain and *M. m. domesticus* (strain B6) the presence of (partially) erased parental PRDM9 hotspots leads to their heterozygosity. The resulting PRDM9 binding site “asymmetry”¹⁵ is supposed to cause hybrid sterility of (PWD × B6)F1 males^{14,26,27}. It was proposed that the DSBs in asymmetric hotspots are more difficult to repair or are fixed by the noncanonical sister-chromatid

repair¹⁵. The consequent lack of symmetric sites is supposed to result in synaptic defects of homologous chromosomes and male sterility^{16,28,44,45}.

If, as proposed, the repair of DSBs by non-sister homologous recombination is inhibited by the asymmetry of their initiation, then a lower than expected frequency of NCOs overlapping the asymmetric DMC1 hotspots could be expected. To examine the idea we classified all NCOs according to the presence of their conversion tract sites within the B6, PWD, (B6 × PWD)F1 or (PWD × B6)F1 DMC1 hotspot datasets¹⁵ and their *Prdm9* binding motif (Table 1-table supplement 3) Only 17.4% of NCO conversion tract sites (15 out of 86) were classified as “nonparental” (named “novel” in ³⁴), being absent in PWD and B6 DMC1 hotspot datasets while present in F1 hybrids (Table 1-table supplement 3, see Materials and Methods for exact definition). This is in contrast with 32% ($P = 0.0099$, $OR = 0.481$, $95\% CI = (0.266, 0.817)$, logistic regression, see Materials and Methods) of nonparental DMC1 hotspots observed in (B6 × PWD)F1 and (PWD × B6)F1 hybrids used as a proxy for the asymmetry. An even higher proportion of nonparental DMC1 sites could be expected in the consomic model because the NCOs were found almost exclusively in the sites where PRDM9^{B6} bound to the PWD chromosome. By default, such sites are more likely to be asymmetric. Thus, the observed deficiency of NCOs that have arisen from asymmetric hotspots matched up with the expected difficulties of repair from the homologous chromatid template. In theory, this might be explained by an increased tendency for CO repair in the nonparental sites.

However, this alternative is highly unlikely since we found almost identical proportion of nonparental CO (19.0%) and NCO (17.4%) sites. To conclude, the observed deficiency of both, NCOs and COs incurred from asymmetric hotspots in accordance with the expected difficulties of their repair from the homologous chromatid template.

Next, we asked whether the autosomes more sensitive to asynapsis display larger proportions of nonparental NCOs/COs. Interestingly, no such correlation was found ($P = 0.5092$, Spearman's $\rho = 0.1614$, Figure 6) suggesting that the DSBs on short, highly asynaptic chromosomes were not particularly more difficult to repair. Rather, the shorter (more asynaptic) chromosomes contain less DSBs in total, making the selection for the properly repaired DSBs more difficult. This finding supports our model of a threshold number of repaired DSBs necessary for the proper pairing^{16,45}.

NCO-associated GC bias

GC-biased gene conversion refers to a situation where GC in a conversion tract is transmitted to gametes more often than AT in a conversion tract from GC/AT heterozygotes^{46,47}. In our set of NCO sites the bases A or T (so-called “weak” bases, w) were converted to bases C or G (“strong” bases, s) 3.4 times more often than *vice versa*, showing a GC bias of 77% ($P = 2.078 \times 10^{-8}$, two-tailed binomial test, $\#w \rightarrow s / (\#w \rightarrow s + \#s \rightarrow w) = 81/(81+24)$). The observed GC bias is higher than 68% or 69% reported in humans³⁰. Remarkably, we observed a similar GC bias both, in single-SNP NCOs (76%, $P = 2.057 \times 10^{-6}$, $42/(42+13)$) and multiple-SNP NCOs (“co-conversions”) (78%, $P = 9.021 \times 10^{-5}$, $39/(39+11)$) (Figure 7a). Our finding is in contrast with the observation in a parallel study⁴⁸ where the *Prdm9*^{Hum}-initiated NCOs were analyzed in detail to see the effect of GC bias in *Prdm9*-uneroded genome. The difference may indicate that the mechanisms causing the GC bias operate differently in a *Prdm9*-eroded and *Prdm9*-uneroded genome.

We observed the same pattern of GC bias in the surrounding non-converted sites, both upstream and downstream from the NCO converted sites, most probably as a consequence of historical meiotic gene conversions. As expected, the rate of GC bias decreased with the

distance from the *Prdm9* binding sites with no GC bias in the 500 bp distance (Figure 7b). This might be explained by combination of more localized NCO gene conversions and more spread CO gene conversion events occurring in the same site in the history of the locus. In addition, we observed a stronger GC bias for parental (63% for *Prdm9*^{B6}, Figure 7c) than for nonparental sites (59%) suggesting that the process of gene conversion leading to GC bias lasted longer in evolution in parental sites and has affected more sites in the surroundings of the PRDM9 motif. Measured alternatively by the level of asymmetry, we also observed a stronger GC bias for the symmetric sites than for the asymmetric ones as expected. Simultaneously, we found that there was an excess of PWD-B6 SNPs in the centers (± 18 bp) of nonparental DMC1 sites (47.6%) than in the centers of parental DMC1 sites (30.0%). These findings are consistent with the view that the binding motifs of nonparental DMC1 sites were disrupted at early generations in the evolution, the binding sites were not used so frequently or even at all since that moment and the gene conversions accompanying the repair of meiotic DSB at these sites created a smaller GC bias. Consistently, the parental DMC1 sites are still active in the parental strains and have created a higher GC bias. Finally, the identified NCO gene conversions disrupted the PRDM9 motif in 20.6% of cases (Figure 7d) showing perfect agreement with the motif erosion localized in a single A3 hotspot¹⁷.

Discussion

In mice, NCOs are the prevalent outcome of the repair of programmed meiotic DSBs with the NCOs/COs overall ratio 9:1, based on cytological data and sperm typing^{7,49,50}. Despite their prevalence and assumed functional significance for meiotic synapsis of homologous chromosomes^{10,11,50} little is known about their genome-wide distribution and properties. In this study, we used whole-genome sequencing to detect and characterize a unique dataset of 95 NCO events in 10 mouse chromosomes using the B6.PWD-Chr # panel of mouse consomic strains. To compare the chromosome-wide distribution and colocalization of NCOs and COs we identified nearly 1500 CO events resulting from (B6 × PWD)F1 meioses. When writing the manuscript another genome-wide study of mouse NCOs was reported⁴⁸. Previous studies of individual hotspots in mice and humans determined the average length of CO-associated conversion tracts to be ~500 bp^{17,18,50}. The NCO gene conversion tracts were much shorter, 50–300 bp at the mouse *Psmb9* hotspot and two hotspots on chromosome 1^{51,52} and 55–290 bp at the MHC hotspot in human chromosome 6¹⁸. More recently, a human genome-wide study based on Illumina SNP arrays indicated the multiple-SNP NCO gene conversion tract length in the range of 100–1000 bp⁵³, while most of NCO events involved only one SNP due to low SNP density. Thus our estimate, 32 bp, based on the screen of 10 mouse chromosomes is much shorter than reported previously for individual hotspots but close to another recent mouse study showing 30 bp and 41 bp for two different *Prdm9* alleles⁴⁸. Another difference from the human studies is an almost complete absence of complex NCO gene conversions in our dataset. Only one out of 94 NCO events was a 231 bp long complex conversion interrupted by one nonconverted SNP in contrast to 46% or 65% of complex events found in human studies^{30,53}.

Until recently, the studies of NCOs in mice were based on the analysis of individual recombination hotspots^{7,17,51,52,54}. Our finding of 96% colocalization of NCO conversion tracts with DMC1 hotspots confirmed both the origin of NCOs from meiotic DSBs and our efficiency to filter out the false positive events. Moreover, the almost exclusive occurrence of NCO events within DMC1 hotspots testifies against a significant share of female-specific NCOs in our dataset. The DMC1 data used for comparison came exclusively from the males, while the identified NCOs were of male and female origin. The first study measuring the DMC1 signal in female meiosis⁵⁵ showed only negligible sex difference in the genomic sites of DMC1 signal. What differed was the activity of DMC1 hotspots corresponding to the variations in H3K4 trimethylation and affected in males by DNA methylation. Thus the findings by Brick and coworkers⁵⁵ of negligible sex differences in the localization of DSBs can be extended to NCOs.

Based on cytological and sperm typing studies the NCOs are supposed to represent ~90% of all repaired DSBs; however the ratio varies depending on the particular hotspot^{49,52}. The mechanism responsible for the decision between CO and NCO resolution is largely unknown. Recently, a study based on whole-genome sequencing of single sperm indicated that PRDM9 protein binding to the uncut template homolog decreases the time necessary for homolog recognition and single-strand DNA invasion into the duplex DNA and increases the chance to process the recombination intermediate into CO⁵⁶. The same study confirmed the previous findings of nonrandom distribution of COs along chromosomes in male meiosis of mice and humans, showing CO excess in subtelomeric regions and their paucity at the centromeric ends^{36-38,49,52}. Our findings of the excess of COs in the subtelomeric regions and in the second decile of the chromosomal length in (B6 x PWD) male meiosis and significant deficit of NCOs in the 10th decile of relative chromosome length are supported by these studies and show

that DNA DSBs at the subtelomeric regions are more prone to be resolved as COs relative to NCOs.

NCOs are associated with biased gene conversion (BGC), which results in uneven recovery of each allele in the next generation and can be generated by recombination initiation bias when the PRDM9 binding motif of the recipient haplotype has higher probability to be targeted by SPO11 than the donor (template) haplotype. The mechanism of GC-biased gene conversion (gBGC) can be explained by biased repair of mismatched AT/GC alleles in heteroduplex DNA favoring strong G/C over weak A/T base pairs^{46,48} or subtle preference of the A/T allele at initiation of recombination^{48,57}. Meiotic drive in favor of the GC allele (~68%) was documented in the human genome^{30,53,58,59} and more recently in mice (60%-68%)⁴⁸. In our study a significant GC bias (76%) was detected genome-wide in the NCO sites, but also in the surrounding non-converted sites, suggesting the DSB repair activity in the past. The GC bias was gradually decreasing outwards from the DMC1 center, with no bias in the ± 500 bp distance, similarly as reported earlier⁴⁷. This fading effect most likely resulted from the historical recombination accompanied by NCO and CO gene conversions, which had migrated from the DSB initiation site. The humanized *Prdm9*^{Hum} binding sites have no history of erosion in classical (e.g., C57BL/6J) or wild-derived laboratory (e.g., PWD/Ph, CAST/EiJ) mouse genomes and the surrounding regions are not GC-biased. Li and coworkers⁴⁸ observed the differences between NCO gene conversion tracts initiated by PRDM9^{Hum} (41 bp) and PRDM9^{Cast} (mean = 30 bp) in the genomes of B6 and CAST crosses. Here, we showed the mean conversion tract length of 32 bp, identical as for PRDM9^{Cast}, for the NCO sites almost exclusively initiated by PRDM9^{B6}. We suggest a possible relationship between the degree of PRDM9 binding site erosion and the length of a gene conversion tract that is not based on inherent properties of the *Prdm9* alleles. Rather, the mean length of conversion

tracts could reflect the preceding activity of PRDM9 in the binding site and surrounding regions and consequent GC bias. Such GC-biased regions are significantly more resistant to be converted again, leading to shorter gene conversion lengths. Consequently, this could imply a gradual decrease in time of building the GC bias as the surrounding bases of the PRDM9 binding site become increasingly G/C rich, which limits the tract length. Asymmetric binding of PRDM9 to allelic sites was identified in several combinations of intersubspecific F1 hybrids^{15,34,42}. It was proposed that excessive asymmetry of DNA DSB hotspots is the key feature of early meiotic arrest in the (PWD × B6)F1 model of hybrid sterility¹⁵. Although many aspects of the role of *Prdm9* in hybrid sterility remain to be clarified⁶⁰, new data bring additional support for the DSB asymmetry hypothesis^{16,45,56}. The repair of asymmetric DSBs can be delayed or aborted, resulting in chromosome breakage and their deficiency in hybrids. Indeed, only 12% of our NCOs appeared as nonparental, generated from a subset of asymmetric DSBs. Despite being significantly below the expected 32%, it shows that some asymmetric sites are either only partially eroded and can be distinguished in a fraction of cells, or the sister chromatid taboo is breached and sister chromatids are used as a template, or a combination of both.

Materials and Methods

Animals and ethics statement

The animals were maintained at the Institute of Molecular Genetics in Prague, Czech Republic. The project was approved by the Animal Care and Use Committee of the Institute of Molecular Genetics ASCR, protocol No 005/2017. The principles of laboratory animal care, Czech Act No 246/1992 Sb., compatible with EU Council Directive 86/609/EEC and Appendix

of the Council of Europe Convention ETS, were observed. The PWD/Ph inbred strain originated from a single pair of wild mice of the *M. m. musculus* subspecies trapped in 1972 in Central Bohemia, Czech Republic²¹. The C57BL/6J (B6) inbred strain was purchased from The Jackson Laboratory. The panel of 27 chromosome substitution strains C57BL/6J-Chr #PWD (B6.PWD-Chr #) maintained by the Institute of Molecular Genetics AS CR, Czech Republic, was prepared in our laboratory²⁰. All mice were maintained in a 12 hr light/12 hr dark cycle in a specific pathogen-free barrier facility. The mice had unlimited access to a standard rodent diet (ST-1, 3.4% fat, VELAZ) and water. All males were sacrificed at age 40-60 d. The tails of progenitors of chromosomal substitution strains have been archived in a -80 °C freezer.

NGS sequencing and genotyping

Partially purified DNA from mouse spleens was isolated by Puregene Core Kit A (QIAGEN 158267) according to the manufacturer's instructions. Next generation sequencing was performed in Illumina HiSeq X to obtain paired-end reads of lengths 2x150bp. The Kapa library preparation protocol was used. The obtained reads were checked for quality by FASTQC and deduplicated. The paired-end reads were aligned to both mm10 reference genome and to PWK pseudo-reference genome and used for identification of NCO sites as described below.

Detection of noncrossover sites

The NCO sites were identified by subsequent filtering steps using both the conservative and anti-conservative method. In the conservative method, the Illumina paired end 2x150 bp reads were aligned to both B6 reference and PWD pseudo-reference fasta files. The

candidate reads were required to align perfectly to the B6 reference and not perfectly but at most with six variants to the PWD pseudo-reference. The detected variants in the reads were then compared to four datasets of Illumina reads of B6, PWK, PWD^{15,61} and the B6.PWD-Chr # strains (this study) and the sites heterozygous either in B6 or PWD parental strains were excluded. Similarly, the gene conversion candidate sites in the transferred PWD chromosomes of B6.PWD-Chr # strains were required to be homozygous. Further, two regions of B6 origin of lengths ~3Mb were found on the background of transferred, otherwise PWD chromosomes in the B6.PWD-Chr 3 and B6.PWD-Chr 4 strains, respectively: chr3:22181981-25498721 and chr4:153634374-156356368. Most likely, they resulted from hidden double crossovers during the preparation of chromosomal substitution strains. As such they were excluded from further analyses. Using the NGS data, all sites were manually curated using the IGV genome browser. Other steps of validation were performed subsequently and independently, namely Sanger sequencing, overlap with the known sites of recombination detected in previous studies^{15,34} as described below.

Noncrossover site amplification and Sanger sequencing

The DNA from mouse tails was amplified at selected NCO sites (see primer list). Samples were amplified by a BIOER XP Cyclor with program: 95 °C for 2 min followed by 40 cycles at 95 °C for 10 s, 55 °C for 20 s and 72 °C for 30 s, and single step at 72 °C for 3 min. Samples showing a single correct size band were treated by ExoSAP-IT enzyme according to manufacturer's instructions (Thermofisher Scientific 78201) and sequenced.

Prdm9 exon 12 amplification

DNA was amplified with primers Exon12-L1 – TGAGATCTGAGGAAAGTAAGAG and Exon12-R – TGCTGTTGGCTTTCTCATTC with PCR program: 94 °C for 5 min and 40 cycles at 94 °C for 30 s, 61 °C for 1 min, 68 °C for 2 min, and a single step at 72 °C for 7 min. The PCR products were resolved in 2% agarose gel allowing visual discrimination of B6 from PWD alleles of *Prdm9* by their different migration.

Dynamics of NCO gene conversion occurrence

For inferring the dynamics of gene conversion during 10 generations of backcrosses, the converted SNPs in NCO sites were additionally sequenced by Sanger sequencing in a step-wise procedure. First, the male parents from BC-10 and BC-6 generations, then the samples from the remaining generations between BC-1, BC-6 and BC-10 were either genotyped, when the closest known generation before was PWD/B6 and the closest generation after was B6/B6, or inferred when the closest known generation before and after were both of the same genotype. The parents from BC-2 and BC-3 used for preparation of B6.PWD-Chr 2 were not archived.

Overlap of NCOs and COs with DMC1 and H3K4me3 ChIP-seq sites

To determine the overlap of NCO and CO sites with DMC1 hotspots, we used the DMC1 datasets of (B6 × PWD)F1, (PWD × B6)F1, B6 and PWD males from the previous studies^{15,34}. The datasets differed in the number of detected hotspots between the biological replicates^{15, 34}; however, the extra DMC1 hotspots had the lowest heat. The heat of DMC1 sites overlapping between the biological replicates was highly correlated ($\rho = 0.92$, $P < 10^{-16}$, for (B6 × PWD)F1). Importantly, the extent of the overlap of autosomal NCO sites with the

respective DMC1 datasets differed by only 2% (higher in Davies and coworkers¹⁵). Furthermore, only the latter datasets¹⁵ provide the information about the inferred PRDM9 initiation of each DSB and the level of asymmetry. For the Chr X, which is missing in the dataset of Davies and coworkers¹⁵, we used the information on the DMC1 overlap from datasets of Smagulova and coworkers³⁴. Similarly, the vast majority (98%) of NCOs overlapping either (B6 × PWD)F1 or (PWD × B6)F1 DMC1 sites overlapped each other. The DMC1 sites were considered uniformly of 2001 bp length, consistently with the reported median DMC1 site length³⁴. Because the autosomal COs in our study were detected in the (B6 × PWD)F1 hybrids, we assumed their localization in DMC1 hotspots of the (B6 × PWD)F1 DMC1 dataset. To investigate the presence of H3K4me3 signal in NCO and CO sites of (B6 × PWD)F1, (PWD × B6)F1 and B6 males we used the data from Davies and coworkers¹⁵. For the comparison of the amount of DMC1 and H3K4me3 signal in NCO and CO sites, we used only the data from the (B6 × PWD)F1 dataset¹⁵. The H3K4me3 sites were considered uniformly of the 2001 bp length.

Comparison of NCO and CO spatial chromosome-wide distribution

The NCOs were compared to the COs in (B6 × PWD)F1 male meioses generated in this study to control for the genetic background. Next, the NCOs were compared to both male and female COs inferred from the G2:F1 population of the Collaborative Cross (CC) project⁶². In both cases, the comparison was performed by a permutation test. The respective CO events were binned into 10 groups according to their relative chromosomal coordinate to give the estimates of the relative proportions of CO events in each bin, which were used to specify the probabilistic distribution. Provided the null hypothesis NCOs do not differ from COs, a dataset of 91 NCO events was simulated from the respective CO distributions. The

simulations were performed 10,000 times and the real NCO distribution was compared to the expected NCO distribution. In addition, to compare with the male and female COs from the CC project, the CO distribution based on 50% of male COs and 50% of female COs and CO distribution based on 25% of male COs and 75% of females COs were calculated as two extreme possibilities. This was done to account for the uncertainty of sex for which the NCOs were observed. The more conservative p-value was reported.

Proportion of nonparental NCO, CO and DSB events as a proxy for the level of asymmetry

We compared the abundance of nonparental NCO, CO and general DSB sites to assess the hypothesis that the asymmetric DSB sites are harder to repair and are possibly repaired by sister chromatids. As a result, they would not be observed as either NCO or CO outcome. Here we defined an “event” as either an NCO event, CO event, or meiotic DSB event. Second, we defined the NCO/CO/DSB events as “nonparental” if found in the DMC1 dataset of (B6 × PWD)F1 and/or (PWD × B6)F1 hybrids and not in the DMC1 dataset of B6 or PWD parents and B6.*Prdm9*^{-/-}. Third, the “parental” events were defined as found in the DMC1 dataset of (B6 × PWD)F1 and/or (PWD × B6)F1 hybrids and also in one of the B6 or PWD DMC1 parental datasets but not in the B6.*Prdm9*^{-/-} dataset. The events not found in any of (B6 × PWD)F1 or (PWD × B6)F1 DMC1 datasets and the events found in the B6.*Prdm9*^{-/-} DMC1 dataset were excluded from the analysis. We calculated the proportion of nonparental events as a ratio (# of nonparental events) / (# of nonparental and parental events) for the NCO events. For the comparison to CO and DSB events we calculated the ratio similarly, but considered two additional features of the DMC1 data. First, DMC1 heat was significantly correlated with the level of asymmetry ($\rho = -0.26$, $P < 0.0001$; $\rho = -0.22$, $P < 0.0001$), and with the proportion of nonparental sites ($\rho = 0.11$, $P < 0.0001$; $\rho = 0.08$, $P <$

0.0001), in (B6 × PWD)F1 hybrids for both PRDM9^{B6} and PRDM9^{PWD} initiated sites respectively. In addition, the median of DMC1 heat was much lower in the overall DSB hotspots than in the DSB hotspots with the detected NCO/CO occurrence (Figure 5-figure supplement 1a). Thus, the DMC1 heat statistically confounded the relationship between the parental/nonparental origin of the events and the class of the event (NCO/CO/DSB) and was controlled for (Table 1-figure supplement 1) in the used logistic regression models. Second, a vast majority (~90%) of the detected NCO events were initiated by PRDM9^{B6} and all the detected NCO events occurred on the PWD chromatid. This was an inherent consequence of using chromosomal substitution strains as a model. This means that the NCO events occurred by design in more asymmetric sites than the DSB sites detected in (B6 × PWD)F1 and (PWD × B6)F1 hybrids by DMC1 ChIP-seq. Thus, the referred difference of proportions of nonparental NCO and nonparental DSB events should be interpreted as a conservative (minimal) estimate implying that the deficit of nonparental NCO events is likely to be even larger.

Statistics

All statistical tests were performed and figures were plotted using R 3.4.4, the tidyverse package (dplyr, ggplot)⁶³ in particular. The mean lengths of NCO gene conversions were assumed to follow an exponential distribution according to Li and coworkers⁴⁸ to provide a direct comparability. Composite likelihood of the rate parameter λ was estimated using the probabilities that the neighboring SNPs were converted. Other, non-standard statistical analyses are described in the Materials and Methods part above. Uniformly, the permutations tests used to calculate the expected overlap between the genomic features

516 were done using $N = 10,000$ permutations. The standard statistical tests used are mentioned
517 in the Results part.

518

519 [Data availability](#)

520 The NGS data are deposited under SRA BioProject number PRJNA564925.

References

- 1 Parvanov, E. D., Petkov, P. M. & Paigen, K. Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835, doi:science.1181495 [pii] 10.1126/science.1181495 (2010).
- 2 Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836-840, doi:science.1183439 [pii] 10.1126/science.1183439 (2010).
- 3 Myers, S. *et al.* Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**, 876-879, doi:science.1182363 [pii] 10.1126/science.1182363 (2010).
- 4 Powers, N. R. *et al.* The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS genetics* **12**, e1006146, doi:10.1371/journal.pgen.1006146 (2016).
- 5 Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375-384 (1997).
- 6 Keeney, S. *et al.* A mouse homolog of the *Saccharomyces cerevisiae* meiotic recombination DNA transesterase Spo11p. *Genomics* **61**, 170-182, doi:10.1006/geno.1999.5956 (1999).
- 7 Cole, F., Keeney, S. & Jasin, M. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. *Molecular cell* **39**, 700-710, doi:10.1016/j.molcel.2010.08.017 (2010).
- 8 Cole, F. *et al.* Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nat Cell Biol* **14**, 424-430, doi:10.1038/ncb2451 (2012).

- 545 9 Gray, S. & Cohen, P. E. Control of Meiotic Crossovers: From Double-Strand Break
546 Formation to Designation. *Annu Rev Genet* **50**, 175-210, doi:10.1146/annurev-genet-
547 120215-035111 (2016).
- 548 10 Romanienko, P. J. & Camerini-Otero, R. D. The mouse Spo11 gene is required for
549 meiotic chromosome synapsis. *Molecular cell* **6**, 975-987, doi:S1097-2765(00)00097-6
550 [pii] (2000).
- 551 11 Carofiglio, F. *et al.* Repair of exogenous DNA double-strand breaks promotes
552 chromosome synapsis in SPO11-mutant mouse meiocytes, and is altered in the
553 absence of HORMAD1. *DNA repair* **63**, 25-38, doi:10.1016/j.dnarep.2018.01.007
554 (2018).
- 555 12 Kauppi, L. *et al.* Numerical constraints and feedback control of double-strand breaks
556 in mouse meiosis. *Genes & development* **27**, 873-886, doi:gad.213652.113 [pii]
557 10.1101/gad.213652.113 (2013).
- 558 13 Baudat, F., Manova, K., Yuen, J. P., Jasin, M. & Keeney, S. Chromosome synapsis
559 defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Molecular*
560 *cell* **6**, 989-998, doi:S1097-2765(00)00098-8 [pii] (2000).
- 561 14 Bhattacharyya, T. *et al.* Mechanistic basis of infertility of mouse intersubspecific
562 hybrids. *Proceedings of the National Academy of Sciences of the United States of*
563 *America* **110**, E468-477, doi:1219126110 [pii]
564 10.1073/pnas.1219126110 (2013).
- 565 15 Davies, B. *et al.* Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in
566 mice. *Nature* **530**, 171-176, doi:10.1038/nature16931 (2016).

- 16 Gregorova, S. *et al.* Modulation of Prdm9-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. *eLife* **7**, pii: e34282. doi: 34210.37554/eLife.34282., doi:10.7554/eLife.34282 (2018).
- 17 Cole, F. *et al.* Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. *Nature genetics* **46**, 1072-1080, doi:10.1038/ng.3068 (2014).
- 18 Jeffreys, A. J. & May, C. A. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature genetics* **36**, 151-156, doi:10.1038/ng1287 (2004).
- 19 Nadeau, J. H., Forejt, J., Takada, T. & Shiroishi, T. Chromosome substitution strains: gene discovery, functional analysis, and systems studies. *Mammalian genome : official journal of the International Mammalian Genome Society* **23**, 693-705, doi:10.1007/s00335-012-9426-y (2012).
- 20 Gregorova, S. *et al.* Mouse consomic strains: exploiting genetic divergence between *Mus m. musculus* and *Mus m. domesticus* subspecies. *Genome research* **18**, 509-515, doi:gr.7160508 [pii] 10.1101/gr.7160508 (2008).
- 21 Gregorova, S. & Forejt, J. PWD/Ph and PWK/Ph inbred mouse strains of *Mus m. musculus* subspecies--a valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biol (Praha)* **46**, 31-41 (2000).
- 22 Yang, H. *et al.* Subspecific origin and haplotype diversity in the laboratory mouse. *Nature genetics* **43**, 648-655, doi:ng.847 [pii] 10.1038/ng.847 (2011).

590 23 Geraldles, A. *et al.* Inferring the history of speciation in house mice from autosomal,
591 X-linked, Y-linked and mitochondrial genes. *Molecular ecology* **17**, 5349-5363,
592 doi:10.1111/j.1365-294X.2008.04005.x (2008).

593 24 Din, W. *et al.* Origin and radiation of the house mouse: Clues from nuclear genes. *J*
594 *Evolution Biol* **9**, 519-539 (1996).

595 25 Ideraabdullah, F. Y. *et al.* Genetic and haplotype diversity among wild-derived mouse
596 inbred strains. *Genome research* **14**, 1880-1887 (2004).

597 26 Forejt, J. Hybrid sterility in the mouse. *Trends Genet* **12**, 412-417, doi:0168-
598 9525(96)10040-8 [pii] (1996).

599 27 Forejt, J., Pialek, J. & Trachtulec, Z. in *Evolution of the House Mouse* (eds M.
600 Macholan, S. J.E. Baird, P. Muclinger, & J. Pialek) 482-503 (Cambridge University
601 Press, 2012).

602 28 Bhattacharyya, T. *et al.* X chromosome control of meiotic chromosome synapsis in
603 mouse inter-subspecific hybrids. *PLoS genetics* **10**, e1004088,
604 doi:10.1371/journal.pgen.1004088 (2014).

605 29 Lilue, J. *et al.* Sixteen diverse laboratory mouse reference genomes define strain-
606 specific haplotypes and novel functional loci. *Nat Genet* **50**, 1574-1583,
607 doi:10.1038/s41588-018-0223-8 (2018).

608 30 Halldorsson, B. V. *et al.* The rate of meiotic gene conversion varies by sex and age.
609 *Nature genetics* **48**, 1377 (2016).

610 31 Baker, C. L., Walker, M., Kajita, S., Petkov, P. M. & Paigen, K. PRDM9 binding
611 organizes hotspot nucleosomes and limits Holliday junction migration. *Genome*
612 *research* **24**, 724-732, doi:10.1101/gr.170167.113 (2014).

613 32 Lam, I., Mohibullah, N. & Keeney, S. Sequencing Spo11 Oligonucleotides for Mapping
614 Meiotic DNA Double-Strand Breaks in Yeast. *Methods Mol Biol* **1471**, 51-98,
615 doi:10.1007/978-1-4939-6340-9_3 (2017).

616 33 Smagulova, F. *et al.* Genome-wide analysis reveals novel molecular features of mouse
617 recombination hotspots. *Nature* **472**, 375-378, doi:nature09869 [pii]
618 10.1038/nature09869 (2011).

619 34 Smagulova, F., Brick, K., Pu, Y. M., Camerini-Otero, R. D. & Petukhova, G. V. The
620 evolutionary turnover of recombination hot spots contributes to speciation in mice.
621 *Genes & development* **30**, 266-280, doi:10.1101/gad.270009.115 (2016).

622 35 Lange, J. *et al.* The Landscape of Mouse Meiotic Double-Strand Break Formation,
623 Processing, and Repair. *Cell* **167**, 695-708 e616, doi:10.1016/j.cell.2016.09.035
624 (2016).

625 36 Cox, A. *et al.* A new standard genetic map for the laboratory mouse. *Genetics* **182**,
626 1335-1344, doi:10.1534/genetics.109.105486 (2009).

627 37 Liu, E. Y. *et al.* High-resolution sex-specific linkage maps of the mouse reveal
628 polarized distribution of crossovers in male germline. *Genetics* **197**, 91-106,
629 doi:10.1534/genetics.114.161653 (2014).

630 38 Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L.
631 Comprehensive human genetic maps: individual and sex-specific variation in
632 recombination. *Am J Hum Genet* **63**, 861-869, doi:10.1086/302011 (1998).

633 39 Littrell, J. *et al.* A High-Resolution Genetic Map for the Laboratory Rat. *G3* **8**, 2241-
634 2248, doi:10.1534/g3.118.200187 (2018).

635 40 Shen, B., Jiang, J., Seroussi, E., Liu, G. E. & Ma, L. Characterization of recombination
636 features and the genetic basis in multiple cattle breeds. *BMC genomics* **19**, 304,
637 doi:10.1186/s12864-018-4705-y (2018).

638 41 Boulton, A., Myers, R. S. & Redfield, R. J. The hotspot conversion paradox and the
639 evolution of meiotic recombination. *Proceedings of the National Academy of Sciences*
640 *of the United States of America* **94**, 8058-8063 (1997).

641 42 Baker, C. L. *et al.* PRDM9 drives evolutionary erosion of hotspots in *Mus musculus*
642 through haplotype-specific initiation of meiotic recombination. *PLoS genetics* **11**,
643 e1004916, doi:10.1371/journal.pgen.1004916 (2015).

644 43 Tiemann-Boege, I., Schwarz, T., Striedner, Y. & Heissl, A. The consequences of
645 sequence erosion in the evolution of recombination hotspots. *Philos Trans R Soc Lond*
646 *B Biol Sci* **372**, doi:10.1098/rstb.2016.0462 (2017).

647 44 Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. A mouse speciation
648 gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373-375,
649 doi:1163601 [pii]
650 10.1126/science.1163601 (2009).

651 45 Wang, L., Valiskova, B. & Forejt, J. Cisplatin-induced DNA double-strand breaks
652 promote meiotic chromosome synapsis in PRDM9-controlled mouse hybrid sterility.
653 *eLife* **7**, doi:10.7554/eLife.42511 (2018).

654 46 Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian
655 genomic landscapes. *Annu Rev Genomics Hum Genet* **10**, 285-311,
656 doi:10.1146/annurev-genom-082908-150001 (2009).

657 47 Clement, Y. & Arndt, P. F. Meiotic recombination strongly influences GC-content
658 evolution in short regions in the mouse genome. *Mol Biol Evol* **30**, 2612-2618,
659 doi:10.1093/molbev/mst154 (2013).

660 48 Li, R. *et al.* A high-resolution map of non-crossover events reveals impacts of genetic
661 diversity on mammalian meiotic recombination. *Nat Commun* **10**, 3900,
662 doi:10.1038/s41467-019-11675-y (2019).

663 49 Baudat, F. & de Massy, B. Regulating double-stranded DNA break repair towards
664 crossover or non-crossover during mammalian meiosis. *Chromosome Res* **15**, 565-
665 577, doi:10.1007/s10577-007-1140-3 (2007).

666 50 Guillon, H., Baudat, F., Grey, C., Liskay, R. M. & de Massy, B. Crossover and
667 noncrossover pathways in mouse meiosis. *Molecular cell* **20**, 563-573,
668 doi:10.1016/j.molcel.2005.09.021 (2005).

669 51 Guillon, H. & de Massy, B. An initiation site for meiotic crossing-over and gene
670 conversion in the mouse. *Nature genetics* **32**, 296-299, doi:10.1038/ng990 (2002).

671 52 de Boer, E., Jasin, M. & Keeney, S. Local and sex-specific biases in crossover vs.
672 noncrossover outcomes at meiotic recombination hot spots in mice. *Genes &*
673 *development* **29**, 1721-1733, doi:10.1101/gad.265561.115 (2015).

674 53 Williams, A. L. *et al.* Non-crossover gene conversions show strong GC bias and
675 unexpected clustering in humans. *eLife* **4**, doi:10.7554/eLife.04637 (2015).

676 54 Ng, S. H., Parvanov, E., Petkov, P. M. & Paigen, K. A quantitative assay for crossover
677 and noncrossover molecular events at individual recombination hotspots in both
678 male and female gametes. *Genomics* **92**, 204-209, doi:10.1016/j.ygeno.2008.06.008
679 (2008).

680 55 Brick, K. *et al.* Extensive sex differences at the initiation of genetic recombination.
681 *Nature* **561**, 338 (2018).

682 56 Hinch, A. G. *et al.* Factors influencing meiotic recombination revealed by whole-
683 genome sequencing of single sperm. *Science* **363**, doi:10.1126/science.aau8861
684 (2019).

685 57 Webb, A. J., Berg, I. L. & Jeffreys, A. Sperm cross-over activity in regions of the human
686 genome showing extreme breakdown of marker association. *Proceedings of the*
687 *National Academy of Sciences of the United States of America* **105**, 10471-10476,
688 doi:10.1073/pnas.0804933105 (2008).

689 58 Odenthal-Hesse, L., Berg, I. L., Veselis, A., Jeffreys, A. J. & May, C. A. Transmission
690 distortion affecting human noncrossover but not crossover recombination: a hidden
691 source of meiotic drive. *PLoS genetics* **10**, e1004106,
692 doi:10.1371/journal.pgen.1004106 (2014).

693 59 Arbeithuber, B., Betancourt, A. J., Ebner, T. & Tiemann-Boege, I. Crossovers are
694 associated with mutation and biased gene conversion at recombination hotspots.
695 *Proceedings of the National Academy of Sciences of the United States of America* **112**,
696 2109-2114, doi:10.1073/pnas.1416622112 (2015).

697 60 Paigen, K. & Petkov, P. M. PRDM9 and Its Role in Genetic Recombination. *Trends*
698 *Genet* **34**, 291-300, doi:10.1016/j.tig.2017.12.017 (2018).

699 61 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene
700 regulation. *Nature* **477**, 289-294, doi:nature10413 [pii]
701 10.1038/nature10413 (2011).

702 62 Churchill, G. A. *et al.* The Collaborative Cross, a community resource for the genetic
703 analysis of complex traits. *Nature genetics* **36**, 1133-1137, doi:ng1104-1133 [pii]

10.1038/ng1104-1133 (2004).

Wickham, H. The tidyverse. *R package ver. 1.1. 1* (2017).

Davies, B. *et al.* Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* **530**, 171-176, doi:10.1038/nature16931 (2016).

Legends to figures

Figure 1. Scheme of construction of chromosomal substitution strains B6.PWD-Chr #. The donor chromosomes (PWD) were transferred to the recipient inbred strain (B6) by ten consecutive backcrosses. In the final intercross the consomic chromosome was fixed in homozygous state allowing direct identification of NCOs by sequencing. The example of B6.PWD-Chr 2 is shown here.

Figure 2. Detection of NCO events. **a** The scheme of NCO detection by comparison of chromosomal substitution strains B6.PWD-Chr# and parental B6 and PWD strains. The NCO site could be covered by 0, 1, or 2 or more SNPs. NCO sites overlapped by 0 SNPs cannot be detected. **b** Localization of NCO events identified on the transferred PWD chromosomes in ten chromosomal substitution strains. The NCO gene conversion with single and multiple converted SNPs are distinguished by color. For the detection of NCOs on chromosome X the subconsomic B6.PWD-ChrX.2 strain carrying the PWD sequence in the proximal 69.6 Mbs was used.

Figure 3. Maximal and minimal possible length of each NCO gene conversion tract. The maximal length is given by the distance of the two nearest non-converted PWD-B6 SNPs. The minimal length is given by the distance of the most outer converted PWD-B6 SNPs; in 71%

cases, the conversion tract contains only one SNP giving a minimal length of 1 bp. *The gene conversion with minimal length = 2659 bp and maximal length = 3966 bp is not displayed.

Figure 4. The inferred PRDM9 allele initiating the autosomal DSBs repaired by NCOs. The initiating PRDM9 allele was inferred based on the overlap with parental B6 and PWD strains and the detected PRDM9 binding motif in datasets of Davies and coworkers¹⁵ and Smagulova and coworkers³⁴.

Figure 5. Histograms of relative chromosomal positions of **a** CO and **b** NCO events. The spatial distributions are significantly different, with a significant deficit of NCO events in comparison to CO events at the 10th decile of relative length.

Figure 6. The deficit of nonparental NCOs or COs does not affect F1 asynapsis rates. The NCOs and COs were pooled into one group as their level of nonparental events did not differ. For each autosome the deficit of nonparental NCOs or COs was calculated as the difference between the proportion of nonparental DSBs and the proportion of nonparental NCOs or COs. The asynapsis rate of each autosome was determined for (PWD × B6)F1 males previously¹⁶. For each autosome the DSBs were paired with NCOs and COs with respect to the (statistically confounding) (B6 × PWD)F1 DMC1 heat determined previously¹⁵.

Figure 7 Genome-wide effects of NCO/CO gene conversions. **a** The observed NCO gene conversion converted the A/T “weak” base into the G/C “strong” base 3.4 times more frequently than *vice versa*, the “strong” base into the “weak” base. **b** GC bias already present due to historical recombination in (B6 × PWD)F1 hybrids in DMC1 sites¹⁵ and ± 1000 bp. The genome is GC-biased (and *Prdm9*-eroded) in the autosomes of both PWD and B6 strains.

Each dot represents the average GC bias in the sliding window of 100 bp. **c** The CG bias (shown here for PRDM9^{B6} and PRDM9^{PWD}) is higher for parental than for nonparental DMC1 sites. The GC bias was calculated analogously as in (b). **d** Histogram of distances of minimal possible NCO tracts from the inferred PRDM9^{B6} binding motif. NCO gene conversions disrupted the binding motif in 20.6% of cases.

Legends to supplementary figures

Figure 5-figure supplement 1. Histograms of the DMC1 heat and H3K4me3 enrichment in (B6 × PWD)F1 hybrids. **a** DMC1 data from ⁶⁴. Red vertical lines labeled q₂₅, q₅₀, q₇₅ represent 25th, 50th and 75th quantiles of the respective distributions. **b** H3K4me3 data from ⁶⁴. Red vertical lines labeled q₂₅, q₅₀, q₇₅ represent 25th, 50th and 75th quantiles of the respective distributions. 1 CO (0.14%) and 68 DSBs (0.39%) with H3K4me3 enrichment greater than 15 are not visualized.

Table 1-figure supplement 1. Deficiency of nonparental NCO and CO sites. The nonparental NCO and CO sites are deficient in comparison to total nonparental DSB sites. The DMC1 heat is binned into deciles according to DMC1 heat for NCO dataset. This was done to visually account for the statistically confounding DMC1 heat. The DMC1 heat data is from ⁶⁴.

a

Chromosome

B6

PWD

B6.PWD-Chr #

B6 conversion

b

Chr #

a COs in B6xPWD, male meiosis

b NCOs in consomics, male & female meiosis

Table 1. Localization and origin of NCO events.

NCO localization					DMC1 coloc. ¹				NCO localization					DMC1 coloc. ¹			
Chr	Start	End	Min ²	Max ²	B6 (PWD×B6)F1	PWD	B6.Prdm9 ^{-/-}	Nonparental ³	Chr	Start	End	Min ²	Max ²	B6 (PWD×B6)F1	PWD	B6.Prdm9 ^{-/-}	Nonparental ³
1	13845860	13845929	70	277	T	T	F	F	4	92481179	92481179	1	596	F	F	F	NA
1	46907980	46907980	1	419	T	T	F	F	4	125451332	125451332	1	661	T	T	F	F
1	59359320	59359321	2	692	T	T	F	F	4	142990293	142990308	16	364	T	T	F	F
1	61872329	61872329	1	69	T	T	F	F	5	6242093	6242323	231	407	T	T	F	F
1	63879266	63879271	6	117	F	T	F	F	5	20789067	20789067	1	115	T	T	F	F
1	67960742	67963400	2659	3994	T	T	F	F	5	27852657	27852657	1	338	T	T	F	F
1	128517316	128517318	3	985	T	T	F	F	5	51691842	51691842	1	92	T	T	F	F
1	137135844	137135844	1	141	T	T	F	F	5	55484917	55484946	30	278	T	T	F	F
1	165896307	165896311	5	680	T	T	F	F	5	90775650	90775650	1	61	T	T	F	F
1	183578162	183578162	1	542	T	T	F	F	5	96571457	96571457	1	94	T	T	F	F
2	3736957	3736957	1	57	T	T	F	F	5	102062273	102062273	1	190	T	T	F	F
2	5040747	5040747	1	133	T	T	F	F	5	104084656	104084656	1	340	T	T	F	F
2	10336587	10336592	6	264	T	T	F	F	5	108572496	108572496	1	441	T	T	F	F
2	11688991	11689010	20	549	F	T	F	F	5	117528325	117528325	1	380	T	T	F	F
2	13697711	13697711	1	431	F	T	T	F	5	119304743	119304775	33	110	T	T	F	F
2	28188305	28188305	1	361	F	T	F	F	5	125695123	125695137	15	211	T	T	F	F
2	32226478	32226493	16	147	F	T	T	F	5	141144023	141144023	1	246	T	T	T	F
2	58517611	58517618	8	270	T	T	F	F	5	148185214	148185214	1	54	F	F	F	NA
2	61489637	61489637	1	643	T	T	F	F	6	18095110	18095110	1	159	T	T	F	F
2	72241857	72241857	1	462	T	T	F	F	6	36814629	36814629	1	299	T	T	F	F
2	74937052	74937064	13	384	T	T	F	F	6	39355424	39355448	25	176	F	T	T	F
2	78357403	78357403	1	50	T	T	F	F	6	101509677	101509677	1	148	T	T	F	F
2	79679676	79679676	1	330	T	T	F	F	6	103949614	103949614	1	122	T	T	T	F
2	85528071	85528071	1	131	T	T	F	F	6	115864297	115864297	1	315	F	T	T	F
2	96760589	96760589	1	169	T	T	F	F	6	128995778	128995882	105	337	T	T	F	F
2	116516318	116516318	1	126	F	T	F	F	6	134249703	134249745	43	388	T	T	F	T
2	135984537	135984537	1	46	F	T	F	F	17	14350774	14350820	47	587	T	T	F	F
2	147805390	147805390	1	383	T	T	F	F	17	25991437	25991437	1	162	F	T	F	T
2	150590002	150590002	1	687	T	T	F	F	17	31984106	31984106	1	113	F	T	T	F
2	166617293	166617524	232	370	F	T	F	F	17	51576043	51576050	8	491	F	T	F	T
3	16828944	16828944	1	267	T	T	F	F	17	67796579	67796615	37	97	F	F	F	NA
3	22128566	22128566	1	378	T	F	F	NA	17	79667734	79667734	1	258	T	T	T	T
3	36830205	36830205	1	492	T	T	F	F	18	4596959	4596972	14	394	T	T	F	F
3	51188409	51188409	1	182	T	T	F	F	18	28878184	28878184	1	305	T	F	F	NA
3	65603278	65603278	1	662	T	T	F	F	18	32083435	32083435	1	46	T	T	F	F
3	65655136	65655136	1	466	F	T	F	F	18	38278002	38278002	1	251	T	T	F	F
3	69739785	69739785	1	232	T	T	F	F	18	40885813	40885839	27	230	T	T	F	F
3	85792585	85792585	1	138	T	T	F	F	18	41946723	41946739	17	162	T	T	F	F
3	95869211	95869211	1	36	T	T	F	F	18	78715451	78715451	1	72	F	T	F	T
3	107617094	107617095	2	185	T	T	F	F	19	4199635	4199635	1	183	T	T	F	F
3	130702831	130702831	1	177	T	T	F	F	19	16998839	16998847	9	130	F	T	F	T
3	136064382	136064385	4	622	F	T	F	F	19	28429944	28429944	1	112	T	T	F	F
4	6396524	6396541	18	399	F	T	F	F	19	28519574	28519585	12	97	T	T	F	F
4	10936243	10936243	1	448	F	T	F	F	19	44938647	44938648	2	444	T	T	F	F
4	49528333	49528367	35	225	T	T	F	F	X	42581841	42581841	1	239	T	T	F	F
4	59369670	59369670	1	286	T	T	F	F	X	52427650	52427650	1	220	F	T	F	T
4	83240097	83240097	1	260	F	T	F	F	X	63297750	63297750	1	193	F	F	F	NA

¹ DMC1 colocalization; data from Davies *et al.* 2016 and Smagulova *et al.* 2016, overlap within ±1000 bp from DMC1 center. T, True; F, False.

² Minimal and Maximal NCO conversion tract lengths.

³ The nonparental event is defined as found in (B6×PWD)F1 but not in parental B6 nor PWD DMC1 datasets.

⁴ Default hotspot; i.e. present in B6.Prdm9^{-/-} DMC1 dataset.

* Complex NCO event, see the main text.

Table 1—table supplement 1.

chromosome	start	end	min_length	max_length	n_variants	Sanger_validated	DMC1_related	DMC1_B6	DMC1_BP	DMC1_P8	DMC1_PWD	DMC1_Prdm9_KO	PRDM9_origin	SPO11_B6	H3K36me3_B6	variants_coords	variants_B6	variants_PWD	left_nonconv_coord	right_nonconv_coord	left_nonconv_B6	left_nonconv_PWD	right_nonconv_B6	right_nonconv_PWD
chr1	13845860	13845929	70	277	3	T	T	T	T	T	T	T	F	T	T	71; 13845929	TTT; C; G	TTTT; T; A	13845847	13846125	C	T	C	T
chr1	46907980	46907980	1	419	1	T	T	T	T	T	T	T	F	T	T	46907980	G	A	46907905	46908325	CTT	---	C	T
chr1	59359320	59359321	2	692	1	T	T	T	T	T	T	T	F	T	T	59359320	TG	GT	59358707	59359400	G	A	C	T
chr1	61872329	61872329	1	69	1	n.d.	n.d.	T	T	T	T	T	F	T	T	61872329		A	61872313	61872383	G	T	C	T
chr1	63879266	63879271	6	117	2	n.d.	n.d.	F	T	T	T	T	F	T	T	63879271	A; G; C	A; G	63879207	63879325	G	A	A	G
chr1	67960742	67963400	2659	3966	8	T	T	T	T	T	T	T	F	T	T	20; 67963400	A; G; C; T	A; G; C; T	67960131	67964098	A	-----	3CCGAG	
chr1	128517316	128517318	3	985	2	n.d.	n.d.	T	T	T	T	T	F	T	T	5; 128517318	C; G	T; T	128517205	128518191	G	A	T	A
chr1	137135844	137135844	1	141	1	n.d.	n.d.	T	T	T	T	T	F	T	T	137135844	G	C	137135739	137135881	A	G	C	T
chr1	165896307	165896311	5	680	2	T	T	T	T	T	T	T	F	T	T	7; 165896311	C; G	T; A	165895698	165896379	C	T	A	C
chr1	183578162	183578162	1	542	1	T	T	T	T	T	T	T	F	T	T	183578162	G	T	183577957	183578500	C	G	T	C
chr2	3736957	3736957	1	57	1	n.d.	n.d.	T	T	T	T	T	F	T	T	3736957	G	T	3736938	3736996	T	C	G	A
chr2	5040747	5040747	1	133	1	n.d.	n.d.	T	T	T	T	T	F	T	T	5040747	G	A	5040687	5040821	AAA	---	C	T
chr2	10336587	10336592	6	264	2	T	T	T	T	T	T	T	F	T	T	87; 10336592	T; C	G; T	10336480	10336745	T	A	A	T
chr2	11688991	11689010	20	549	3	T	T	T	T	T	T	T	F	T	T	07; 11689010	C; G; G	A; A; A	11688523	11689073	A	-	C	T
chr2	13697711	13697711	1	431	1	T	T	T	T	T	T	T	F	T	T	13697711	A	G	13697334	13697766	AC	--	T	G
chr2	28188305	28188305	1	361	1	T	T	T	T	T	T	T	F	T	T	28188305	G	A	28188155	28188517	G	G	A	G
chr2	32226478	32226493	16	147	2	n.d.	n.d.	T	T	T	T	T	F	T	T	78; 32226493	C; A	A; C	32226396	32226544	C	G	A	G
chr2	58517611	58517618	8	270	2	T	T	T	T	T	T	T	F	T	T	11; 58517618	G; C	A; T	58517431	58517702	G	A	IGTAGC	-----
chr2	61489637	61489637	1	643	1	T	T	T	T	T	T	T	F	T	T	61489637	C	G	61489162	61489806	-----	3TGTTG	T	-
chr2	72241857	72241857	1	462	1	T	T	T	T	T	T	T	F	T	T	72241857	G	A	72241474	72241937	G	A	T	A
chr2	74937052	74937064	13	384	2	T	T	T	T	T	T	T	F	T	T	52; 74937064	G; C	A; T	74936895	74937280	C	T	C	A
chr2	78357403	78357403	1	50	1	n.d.	n.d.	T	T	T	T	T	F	T	T	78357403	C	G	78357374	78357425	G	A	G	A
chr2	79679676	79679676	1	330	1	T	T	T	T	T	T	T	F	T	T	79679676	G	A	79679609	79679940	A	G	A	T
chr2	85528071	85528071	1	131	1	n.d.	n.d.	T	T	T	T	T	F	T	T	85528071	G	A	85528033	85528165	C	T	C	T
chr2	96760589	96760589	1	169	1	T	T	T	T	T	T	T	F	T	T	96760589	C	T	96760507	96760767	G	A	C	G
chr2	116516318	116516318	1	126	1	n.d.	n.d.	T	T	T	T	T	F	T	T	116516318	T	G	116516266	116516393	A	G	A	G
chr2	135984537	135984537	1	46	1	n.d.	n.d.	T	T	T	T	T	F	T	T	135984537	T	C	135984527	135984574	G	A	T	G
chr2	147805390	147805390	1	383	1	T	T	T	T	T	T	T	F	T	T	147805390	A	T	147805306	147805690	G	A	C	T
chr2	150590002	150590002	1	687	1	T	T	T	T	T	T	T	F	T	T	150590002	C	T	150589924	150590612	-	C	C	A
chr2	166617293	166617524	232	370	1	T	T	T	T	T	T	T	F	T	T	3; 166617524	ncC*; ncC*; G; ncT*; ncT*; A		166617226	166617597	G	A	G	A
chr3	16828944	16828944	1	267	1	T	T	T	T	T	T	T	F	T	T	16828944	C	T	16828894	16829162	A	G	G	T
chr3	22128566	22128566	1	378	1	T	T	T	T	T	T	T	F	T	T	22128566	G	A	22128341	22128720	T	C	T	C
chr3	36830205	36830205	1	492	1	T	T	T	T	T	T	T	F	T	T	36830205	G	A	36829989	36830482	-	T	G	T
chr3	51188409	51188409	1	182	1	T	T	T	T	T	T	T	F	T	T	51188409	T	C	51188366	51188549	T	A	AAAA	-----
chr3	65603278	65603278	1	662	1	T	T	T	T	T	T	T	F	T	T	65603278	C	T	65602922	65603585	G	A	-	-
chr3	65655136	65655136	1	466	1	T	T	T	T	T	T	T	F	T	T	65655136	C	G	65654711	65655178	C	T	G	T
chr3	69739785	69739785	1	232	1	n.d.	n.d.	T	T	T	T	T	F	T	T	69739785	G	T	69739758	69739991	A	G	A	T
chr3	85792585	85792585	1	138	1	n.d.	n.d.	T	T	T	T	T	F	T	T	85792585	G	A	85792540	85792679	G	T	C	A
chr3	95869211	95869211	1	36	1	n.d.	n.d.	T	T	T	T	T	F	T	T	95869211	G	A	95869199	95869236	C	T	G	A
chr3	107617094	107617095	2	185	1	T	T	T	T	T	T	T	F	T	T	107617094	CG	TA	107616928	107617114	G	A	A	T
chr3	130702831	130702831	1	177	1	T	T	T	T	T	T	T	F	T	T	130702831	C	T	130702792	130702970	A	C	A	G
chr3	136064382	136064385	4	622	1	T	T	T	T	T	T	T	F	T	T	3; 136064385	G; C	A; T	136063966	136064589	G	A	GGA	ATT
chr4	6396524	6396524	18	399	2	T	T	T	T	T	T	T	F	T	T	524; 6396524	T; C	T	6396368	6396768	C	T	GGA	T
chr4	10936243	10936243	1	448	1	T	T	T	T	T	T	T	F	T	T	10936243	C	T	10935986	10936435	--	GT	G	A
chr4	49528333	49528367	35	225	3	T	T	T	T	T	T	T	F	T	T	35; 49528367	C; C; C	T; G; T	49528278	49528504	G	A	C	T
chr4	59369670	59369670	1	286	1	T	T	T	T	T	T	T	F	T	T	59369670	C	T	59369506	59369793	A	G	T	C
chr4	83240097	83240097	1	260	1	n.d.	n.d.	T	T	T	T	T	F	T	T	83240097	C	T	83240079	83240340	T	G	A	G
chr4	92481179	92481179	1	596	1	T	T	T	T	T	T	T	F	T	T	92481179	T	C	92480956	92481553	A	G	GTGTG	-----
chr4	125451332	125451332	1	661	1	T	T	T	T	T	T	T	F	T	T	125451332	C	T	125451295	125451957	A	T	C	T
chr4	142990293	142990308	16	364	2	n.d.	n.d.	T	T	T	T	T	F	T	T	3; 142990308	--; C	CT; T	142990122	142990487	A	C	G	A
chr5	6242093	6242323	231	407	2	T	T	T	T	T	T	T	F	T	T	093; 6242323	C; C	T; G	6242023	6242431	AT	GG	A	C
chr5	20789067	20789067	1	115	1	n.d.	n.d.	T	T	T	T	T	F	T	T	20789067	C	T	20789051	20789167	C	T	-	G
chr5	27852657	27852657	1	338	1	T	T	T	T	T	T	T	F	T	T	27852657	G	T	27852582	27852921	-----	GAGGG	G	A
chr5	51691842	51691842	1	92	1	n.d.	n.d.	T	T	T	T	T	F	T	T	51691842	A	G	51691831	51691924	G	A	-	-
chr5	55484917	55484946	30	278	2	T	T	T	T	T	T	T	F	T	T	17; 55484946	C; C	T; G	55484708	55484987	C	A	CCCCCT	-----
chr5	90775650	90775650	1	61	1	n.d.	n.d.	T	T	T	T	T	F	T	T	90775650	G	A	90775609	90775671	A	G	C	A
chr5	96571457	96571457	1	94	1	n.d.	n.d.	T	T	T	T	T	F	T	T	96571457	G	T	96571409	96571504	C	T	G	A
chr5	102062273	102062273	1	190	1	T	T	T	T	T	T	T	F	T	T	102062273	G	A	102062209	102062400	C	T	A	T
chr5	104084656	104084656	1	340	1	T	T	T	T	T	T	T	F	T	T	104084656	T	A	104084370	104084711	G	A	-----	CTAGG
chr5	108572496	108572496	1	441	1	T	T	T	T	T	T	T	F	T	T	108572496	G	A	108572066	108572508	C	T	-	CC
chr5	117528325	117528325	1	380	1	T	T	T	T	T	T	T	F	T	T	117528325	G	A	117528280	117528661	G	A	C	T
chr5	119304743	119304775	33	110	2	n.d.	n.d.	T	T	T	T	T	F	T	T	3; 119304775	A; C	G; T	119304692	119304803	C	T	G	-
chr5	125695123	125695137	15	211	2	T	T	T	T	T	T	T	F	T	T	3; 125695137	T; C	C; A	125695019	125695231	A	-	T	C
chr5	141144023	141144023	1	246	1	n.d.	n.d.	T	T	T	T	T	F	T	T	141144023	T	C	141143846	141144093	G	A	G	A
chr5	148185214	148185214	1	54	1	T	T	T	T	T	T	T	F	T	T	148185214	T	C	148185183	148185238	C	-	T	G
chr6	18095110	18095110	1	159	1	n.d.																		

Table 1—table supplement 2.

Chromosome	Length_mm10	SPO11_oligos	NCO_count
chr1	191909192	69606	10
chr2	178326651	73115	20
chr3	156398855	53878	12
chr4	152055611	60332	8
chr5	147919674	62217	15
chr6	146336543	52717	8
chr17	91707462	39641	6
chr18	87452634	34544	7
chr19	58205856	25629	5

SPO11 oligos data are taken from Lange et al., 2016

Table 1–table supplement 3. Classification of NCOs based on the inferred initiating PRDM9 allele and predicted hotspot symmetry.

Predicted NCO class	Parentality	DMC1 dataset *				Hotspot origin *	No of NCOs
		PWD	B6	B6×PWD	PWD×B6		
Weak PRDM9 motif, symmetric	parental	+	+	+	+	undetermined	2
PWD sym	parental	+	-	+	+	PWD	5
PWD asym	nonparental	-	-	+	+	PWD	0
B6 sym **	parental	-	+	+	+	B6	66
B6 sym	parental, incomparable	-	+	-	-	B6	2
B6 asym †	nonparental	-	-	+	+	B6	10
Weak PRDM9 motif, asymmetric †	nonparental	-	-	+	+	undetermined	8
Anomalous, default ‡	anomalous	+	+	+	+	B6. <i>Prdm9</i> ^{-/-}	2
Anomalous, DMC1 negative	anomalous	-	-	-	-	none	4

* Davies et al. 2016, Smagulova et al. 2016

** includes 1 NCO from (PWD-, B6+, B6×PWD- and PWD×B6+)

† includes 1 NCO from (PWD-, B6-, B6×PWD- and PWD×B6+)

†† includes 1 NCO from (PWD-, B6-, B6×PWD- and PWD×B6+)

‡ includes 1 NCO from (PWD-, B6+, B6×PWD+, PWD×B6+)

Figure 4—table supplement 1. Genotype of *Prdm9* in 10 generations of backcrosses used for preparation of chromosomal substitution strains.

Line	BC-1	BC-2	BC-3	BC-4	BC-5	BC-6	BC-7	BC-8	BC-9	BC-10
B6.PWD-Chr1	BP	BB	BB	BB	BB	BB	BB	BB	BB	BB
B6.PWD-Chr2	BP	BP	NA	NA	BB	BB	BB	BB	BB	BB
B6.PWD-Chr3	BP	BB	BB	BB	BB	BB	BB	BB	BB	BB
B6.PWD-Chr4	BP	BP	NA	BB	BB	BB	BB	BB	BB	BB
B6.PWD-Chr5	BP	BP	BB	BB	BB	BB	BB	BB	BB	BB
B6.PWD-Chr6	BP	BB	BB	BB	BB	BB	BB	BB	BB	BB
B6.PWD-Chr17	BP	BP	BP	BP	BP	BP	BP	BP	BP	BP
B6.PWD-Chr18	BP	BB	BB	BB	BB	BB	BB	BB	BB	BB
B6.PWD-Chr19	BP	BP	NA	NA	BB	BB	BB	BB	BB	BB