Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice

Shintaro Yamada[a,b], Seoyoung Kim[a], Sam E. Tischfield[a,c], Julian Lange[a,b], Maria Jasin[d], Scott Keeney[a,b]

[a]Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[b]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[c]Tri-Institutional Training Program in Computational Biology and Medicine, Weill Cornell Medical College, New York, NY, USA

[d]Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Contact
Scott Keeney
s-keeney@ski.mskcc.org
Molecular Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

Keywords
SPO11; meiosis; recombination; DNA double-strand breaks; resection; PRDM9; chromatin; repetitive elements

Abstract

The SPO11-generated DNA double-strand breaks (DSBs) that initiate meiotic recombination occur non-randomly across genomes, but mechanisms shaping their distribution and repair remain incompletely understood. Here, we expand on recent studies of nucleotide-resolution DSB maps in mouse spermatocytes. We find that trimethylation of histone H3 lysine 36 around DSB hotspots is highly correlated, both spatially and quantitatively, with trimethylation of H3 lysine 4, consistent with coordinated formation and action of both PRDM9-dependent histone modifications. In contrast, the DSB-responsive kinase ATM contributes independently of PRDM9 to controlling hotspot activity, and combined action of ATM and PRDM9 can explain nearly two-thirds of the variation in DSB frequency between hotspots. DSBs were modestly underrepresented in most repetitive sequences such as segmental duplications and transposons. Nonetheless, numerous DSBs form within repetitive sequences in each meiosis and some classes of repeats are preferentially targeted. Implications of these findings are discussed for evolution of PRDM9 and its role in hybrid strain sterility in mice. Finally, we document the relationship between mouse strain-specific DNA sequence variants within PRDM9 recognition motifs and attendant differences in recombination outcomes. Our results provide further insights into the complex web of factors that influence meiotic recombination patterns.

Introduction

Cells undergoing meiosis inflict DNA double-strand breaks (DSBs) at many places across the genome to initiate recombination, which physically links homologous chromosomes to promote their accurate segregation. These DSBs occur preferentially (but not exclusively) within highly localized regions called hotspots.[1] The non-random distribution of DSBs governs the evolution and diversity of eukaryotic genomes. Furthermore, failure to properly form and repair meiotic DSBs results in gametes with chromosome structure alterations or aneuploidy, which can lead to developmental disorders.[2,3] An important challenge has been to understand the complex interplay of multiple factors that shape this DSB landscape over size scales ranging from single nucleotides to whole chromosomes.[4-6]

Meiotic DSBs are formed by dimers of the conserved topoisomerase-like protein SPO11 via a transesterase reaction that links a SPO11 molecule to each 5′ end of the broken DNA.[7,8] DNA nicks then release SPO11 covalently bound to short oligonucleotides (SPO11 oligos).[9] 5′→3′ resection generates a single-stranded DNA (ssDNA) tail at each DSB end.[10-13] This ssDNA becomes coated with strand-exchange proteins DMC1 and RAD51 and searches for homologous DNA as a repair template.[14-17] DSB repair is completed as a crossover (reciprocal exchange of chromosome arms that flank the repair site) or a noncrossover.[18]

Among many levels of mammalian DSB landscape organization, hotspot control by the histone methyltransferase PRDM9 has been the most extensively studied. In mice and humans, PRDM9 is a major determinant of hotspot locations via its sequence-specific, multi-zinc-finger DNA-binding domain, the specificity of which evolves rapidly and is highly polymorphic in populations.[19,20] Genome-wide hotspot distributions in these two organisms have been examined by mapping DMC1-bound ssDNA or PRDM9-dependent histone H3 lysine 4 trimethylation (H3K4me3).[21-25] However, constraints on spatial resolution of these maps left questions unanswered about fine-scale DSB patterns, especially at the sub-hotspot level. We recently overcame this issue by sequencing SPO11 oligos purified from mouse testes.[26] SPO11-oligo mapping provided quantitative DSB landscapes at nucleotide resolution, with low background and high dynamic range. The fine-scale maps revealed previously invisible spatial

features of DSBs and relationships between SPO11, PRDM9 and methylated nucleosomes. Here, we combine these SPO11-oligo data with other published genome-wide data to further explore genomic features that influence DSB formation and repair in mice.

Results and Discussion

Relationships of DSB patterns with the trimethylation of H3K36 and H3K4

Studies of DSB and H3K4me3 distributions have shown that DSBs are targeted by PRDM9 to genomic regions that are recognized by the PRDM9 zinc finger DNA-binding domain and subject to local histone methylation by its PR/SET methyltransferase domain.[21,22,24,27-29] Previous work showed that the mean H3K4me3 signal oscillates around DSB hotspots, with immediately adjacent nucleosomes showing stronger H3K4me3 signal than nucleosomes further away (**Fig. 1A**).[22,26] However, PRDM9 also trimethylates histone H3 on lysine 36 *in vitro* and *in vivo*,[24,30-32] so we examined this modification as well using published H3K36me3 chromatin immunoprecipitation (ChIP) data.[32]

When averaged around SPO11-oligo hotspot centers, the H3K36me3 pattern was strikingly similar to that of H3K4me3 (**Fig. 1A**). It was previously reported that the H3K4me3:H3K36me3 ratio around H3K4me3 peaks is higher for nucleosomes immediately adjacent to PRDM9 binding sites than for nucleosomes further away,[32] but this pattern was not apparent when centered on SPO11-oligo hotspots (**Fig. 1A**), suggesting that this is not a robust feature of PRDM9-dependent histone methylation. This pattern was irrespective of whether we applied local normalization.

We previously reported that the apparent symmetry in H3K4me3 disposition around hotspots (see **Fig. 1A**) is an artifact of averaging, and that individual hotspots display a continuum of varying degrees of left–right asymmetry (**Fig. 1B**, left panel).[26] This feature was also seen for H3K36me3, and the profiles for the two modifications were highly correlated (**Fig. 1B–C**). Hotspots can be classified into three groups on the basis of the spatial relationship between local SPO11-oligo patterns and putative 36-base-pair (bp) PRDM9 binding site (**Fig. 1D**).[26] H3K36me3 spatial distribution and ChIP signal strength were comparable among the three hotspot classes, again largely indistinguishable from prior findings for H3K4me3 (**Fig. 1D–E**).[26]

Whereas variation in H3K4me3 ChIP signal could account for 40% of the variation in SPO11-oligo counts at hotspots ($R^2 = 0.40$),[26] H3K36me3 could only account for 26% (**Fig. 1F**). This difference is partially due to higher background signal from transcription-dependent H3K36me3: when we eliminated hotpots in genes previously shown to be transcribed in juvenile testes[33] (where meiotic cells are enriched because of semi-synchronous spermatogenesis), the correlation increased to ~33% (data not shown). By multiple linear regression, combining data for both histone marks for all hotspots gave a significant but quantitatively small improvement over a model with H3K4me3 alone for predicting SPO11-oligo counts ($R^2 = 0.44$; $p < 2.2 \times 10^{-16}$, ANOVA; **Table S1**).

PRDM9 has been proposed to trimethylate H3K4 and H3K36 on the same nucleosome at least some of the time.[32] Taken together, our findings of highly similar H3K4me3 and H3K36me3 patterns (both spatial and quantitative) around SPO11-oligo hotspots agree with this proposal. Conversely, the findings provide little if any support for an alternative hypothesis in which methylation of the two residues contribute to hotspot activity independently.

Combinatorial effects of ATM and PRDM9 in controlling hotspot heat

Hotspots are just one organizational level among many in the DSB landscape. In the budding yeast *Saccharomyces cerevisiae*, DSB distributions are shaped by multiple factors that work hierarchically and combinatorially.[4,6,34-36] For example, most yeast hotspots correspond to the

nucleosome-depleted regions in gene promoters,[4,37,38] but how strong a hotspot will be is shaped not only by factors within the hotspot itself, but also by larger-scale chromosome structures operating over distances from tens of kilobases (kb) up to whole chromosomes.[4,36,38-40] Some of these factors involve feedback circuits that regulate the ability of Spo11 to continue making DSBs depending on whether DSBs have already formed and whether chromosomes are successfully engaging their homologs (reviewed in ref.[41]) The mammalian DSB landscape is thought to be shaped similarly,[24,26,35] but there has been little if any formal exploration of the degree to which different factors interact with one another. To address this question, we compared ATM- and PRDM9-dependent contributions to the DSB landscape.

The ATM (for ataxia telangiectasia mutated) kinase triggers checkpoint signaling and promotes DSB repair.[42] Meiotic DSBs activate ATM, which in turn suppresses further DSB formation.[43] This negative feedback circuit is conserved in *S. cerevisiae*, dependent on the yeast ATM ortholog Tel1.[40,44,45] In both mouse and yeast, ATM/Tel1-dependent DSB control also shapes the DSB landscape.[26,40]

We hypothesized that ATM-mediated DSB control is independent of PRDM9 activity. We showed previously that in the absence of ATM, nearly all hotspots experience more DSBs but weaker hotspots increase more in heat than stronger ones: the ratio of SPO11-oligo counts in ATM-deficient relative to ATM-proficient samples was negatively correlated with hotspot heat.[26] By linear regression this correlation accounted for 16% of the variation in SPO11-oligo counts at hotspots found in the C57BL/6J ("B6") strain (**Fig. 1G**). Differences between hotspots in their response to *Atm* mutation correlated poorly with their H3K4me3 levels (**Fig. 1H**). Furthermore, a multiple linear regression model that combined measures of PRDM9 activity (H3K4me3 and H3K36me3) with measures of the effects of ATM ($Atm^{-/-}$:$Atm^{+/+}$ SPO11-oligo ratio) substantially improved the ability to predict hotspot heat relative to a model incorporating H3 methylation status only ($R^2 = 0.60$; **Fig. 1I** and **Table S1**). Taken together, these findings support the idea that ATM and PRDM9 contribute largely independently to determining the heat of individual hotspots. More generally, these findings illustrate the degree to which different factors can interact to shape the DSB landscape.

Characteristics of DNA sequences around 5′ and 3′ ends of SPO11 oligos

Fine-scale analyses of sequence composition around meiotic DSB sites in budding and fission yeasts revealed non-random base composition but no apparent consensus sequence[4,46-49]. SPO11 generates two-nucleotide (nt) 5′ overhangs,[47,48] predicting an axis of rotational symmetry at the phosphodiester bond between the first and second position for each mapped SPO11 oligo (**Fig. 2A**).[4] We examined possible SPO11 biases in mice by orienting and aligning the DNA sequences surrounding each uniquely mapped SPO11 oligo from 9,060 PRDM9-motif-containing hotspots and then evaluating DNA sequence composition around the predicted dyad axis (**Fig. 2A**). Base composition deviated from random at all positions from six nt upstream to 21 nt downstream of the dyad axis (-6 to +21). The bias from +10 to +21 may reflect preferences related to oligo 3′-end formation, discussed further below. The strong bias in the central 12 nt, a region predicted to contact SPO11,[50] could have reflected preferences of SPO11 itself as inferred in yeasts,[4,49] but could alternatively reflect signatures of the sequence specificity of PRDM9 (whose binding sites lie very close to many SPO11 cleavage sites (**Fig. 1D**)) or biases of the sequencing and/or mapping methods.

We therefore asked if base composition biases near SPO11-oligo mapping positions were rotationally symmetric around the predicted SPO11-dyad axis, as previously demonstrated in *S. cerevisiae*.[4] However, the base composition was not clearly two-fold rotationally symmetric (**Fig. 2B**). For example, A residues were enriched at the -2 position but there was no reciprocal enrichment of T at +2. We further explored this question by assessing the degree to which the dinucleotide frequencies to the left of the dyad axis correlated with dinucleotide frequencies to the right. To do so, we examined the regions from -3 to -16 and +3 to +16 to avoid the artifactual

enrichment or depletion of C-containing dinucleotides around the 5′ end of the mapped SPO11 oligos.[4] For rotationally symmetric patterns, pairwise comparisons of reverse-complementary dinucleotides should show high, positive values of Pearson's r, as previously observed in *S. cerevisiae*.[4] However, we found no such enrichment for positive r values (**Fig. 2C**), thus the sequence bias around mouse SPO11-oligo sites cannot be clearly ascribed to preferences of SPO11 itself.

To test whether presence of PRDM9 binding motifs was obscuring underlying SPO11-associated base composition bias, we focused on two sets of hotspot-associated SPO11 oligos (after first excluding reads with 5′ ends mapping to C residues to minimize a known technical bias[4]): 1) reads whose 5′ ends mapped within a PRDM9 motif, and 2) reads that did not overlap a PRDM9 motif anywhere along their lengths. Base composition around the two sets displayed distinct patterns in addition to the selected-for depletion of C at the -1 position (**Fig. 2B**). For cleavage sites within PRDM9 motifs there was a pronounced 3-nt periodicity in the frequencies of each mononucleotide (blue lines in **Fig. 2B**). This periodicity presumably reflects the fact that each individual PRDM9 zinc finger recognizes a DNA triplet sequence.[19,20,51] However, SPO11-oligo sites that did not overlap a PRDM9 motif still displayed little or no evidence of rotationally symmetric base composition (red lines in **Fig. 2B**). Thus, unlike in yeast, we have been unable to discern in mouse a clear signature that we can ascribe to SPO11 preferences for particular base compositions. These findings do not exclude such preferences contributing to cleavage site choice, but if they exist, such contributions appear to be quantitatively weak relative to other sources of non-randomness in fine-scale SPO11-oligo maps.

In *S. cerevisiae*, the 3′ ends of Spo11 oligos are thought to be formed by a combination of endonuclase and 3′→5′ exonuclease activities of the Mre11-Rad50-Xrs2 complex plus Sae2 (MRE11-RAD50-NBS1 and CTIP in mouse).[9,52-55] Because many SPO11 oligos overlap the positions where PRDM9 binds, we asked whether PRDM9 might influence SPO11-oligo 3′-end formation. On average, oligo 3′ ends were offset by similar lengths from oligo 5′ ends in both directions, reflecting oligos mapping to the top and bottom strands of the DNA (**Fig. 2D**). The offsets of 3′-end positions were as expected for the lengths of mapped SPO11-oligo reads, irrespective of where DSBs occurred relative to PRDM9 binding motifs. This finding implies that PRDM9 has little impact on the nucleolytic processing steps that form the 3′ ends of SPO11 oligos, in turn suggesting that PRDM9 may not be bound to DNA when the MRE11 complex completes its processing function. This inference fits with an earlier proposal that oftentimes PRDM9 has already left its binding sites when SPO11 cleaves the DNA, on the basis of the frequent overlap of SPO11 oligos with PRDM9 binding sites (e.g., **Fig. 1D**).[26] Recombinant PRDM9 forms long-lived complexes with DNA *in vitro*,[56] raising the possibility that it is actively displaced from its binding sites *in vivo* by SPO11 and/or other factors.

DSBs are underrepresented but nonetheless occur frequently within repeated sequences
The mouse genome is replete with repeated sequences.[57,58] These sequences pose a risk because a meiotic DSB formed within one copy of a repeat has the potential to cause a chromosomal rearrangement by repair with a non-allelic copy. Such non-allelic homologous recombination (NAHR) can lead to deleterious deletions, duplications, and other alterations. Indeed, NAHR-mediated chromosomal rearrangements in the germline are the cause of numerous developmental disorders in humans.[3,59-61] In budding yeast, DSB formation tends to be repressed within repetitive DNA genome-wide,[4,62] perhaps as a mechanism to protect against NAHR. We hypothesized that this strategy is also found in mice.

We asked whether SPO11 oligos were less likely to arise from repeated sequences than expected if DSBs were formed in proportion to the relative amount of genomic space occupied by repeats. SPO11 oligos were assigned to two sequence classes: "repeat" and "non-repeat" (**Fig. 3A**). The repeat class, totaling 1,286,319,054 bp, encompassed two sub-groups:

5

interspersed repeats such as transposable elements and low complexity sequences compiled by RepeatMasker (www.repeatmasker.org), which together comprise 45% of the genome;[57] and segmental duplications, defined as genomic segments sharing ≥90% sequence identity over ≥1 kb, which collectively occupy 8.1% of the genome.[58] These two sub-groups are partially overlapping (4.7% of the genome). The non-repeat class encompasses the remainder of the genome and totals 1,366,448,147 bp.

Reads that map to multiple locations are conventionally excluded from Next Generation Sequencing studies because they cannot be assigned unambiguously. However, the exclusion of multi-mapped reads in this analysis would result in an underestimation of the DSB frequency in repeated sequences. Some SPO11 oligos can be mapped uniquely even to repeat-class sequences (**Fig. 3A**), because individual copies of repetitive elements can often be differentiated from one another by DNA sequence variants. Conversely, SPO11-oligos can be assigned to multiple locations that lie within non-repeat sequences as a consequence of the short length of SPO11-oligos and the low complexity of the mouse genome relative to a truly random DNA sequence of the same size (**Fig. 3A–B**).[26] Thus, multi-mapped reads that map to both repeat and non-repeat sequences cannot be unambiguously placed into either class.

We therefore used four methods with varying stringency to allocate multi-mapped reads among the repeat and non-repeat classes (**Fig. 3A–C**). By all four approaches, the frequency of SPO11-oligos in repeat sequences was lower than expected based on the fraction of the genome occupied by such sequences, and this was true for both sub-groups of the repeat class (**Fig. 3A–D**). This apparent DSB suppression was the case for many families of interspersed repeats, including retrotransposons such as the L1 (LINE1) elements L1Md_A, L1Md_Gf and L1Md_T, and the ERVK family LTR (long terminal repeat) elements IAPEz and MMERVK10C, whose meiotic transcription is suppressed by DNA methylation targeted to them by DNMT3L, a catalytically inactive member of the Dnmt3 DNA methyltransferase family (**Fig. 3E** and **Table S2**).[63] Interestingly, however, several repeat element classes had higher SPO11-oligo frequencies than expected, including sequences that RepeatMasker annotated as homologous to the DNA transposon subfamilies MULE-MuDR, TcMar-Mariner, hAT-Charlie, and PiggyBac.

Two main conclusions arise from this analysis. The "glass half empty" viewpoint is that, on average, repetitive DNA sequences incur fewer DSBs than expected by chance. The "glass half full" viewpoint is that this apparent DSB suppression, while it may be strong for individual repeats, is quantitatively modest in global terms: assuming 200–300 DSBs per spermatocyte,[64] as many as 65–100 DSBs are formed in each meiosis within sequences that have at least some repetitive character. The extent to which meiotic NAHR depends on the percent sequence similarity and length between repeats remains unknown. Thus, for many or most of these DSBs, it is likely that sequence differences from non-allelic copies prevent NAHR from occurring. However, it is likely that at least some of these DSBs pose significant risk for NAHR in every meiosis. The mechanisms that would allow a cell to tolerate this risk remain poorly understood.[3,60,65,66]

It appears that active mechanisms may disfavor DSB formation within specific repeat families, e.g., as a result of targeted DNA methylation and heterochromatinization.[63] However, because SPO11-oligo underrepresentation appears to be broadly true of many repeat classes, we propose that this feature of the DSB landscape reflects in part a selective constraint against having too many recombination events involving repeats. More specifically, we hypothesize that PRDM9 proteins that target repetitive DNA too robustly may not be compatible with fertility, thus only those *Prdm9* alleles that do not confer too high a burden of repeat-associated DSBs are likely to be found in populations.

A corollary of this hypothesis is that inappropriately high DSB levels in repeats may contribute to the infertility seen in some mouse hybrid strains, which is caused by incompatibility between the *Prdm9* allele of one strain and the genome of the other.[24,67-69] Prior studies have established that small differences in total DSB numbers can spell the difference between

successful meiosis and infertility caused by spermatogenic failure.[64,70] These findings have been interpreted to mean that there is a threshold in the number of recombination events needed to support successful chromosome pairing, synapsis, and segregation in mice.[64,70-72] DSBs within repeats may not contribute effectively to chromosome pairing and may in fact interfere with pairing by stabilizing illegitimate interactions between non-homologous chromosomes. Thus, given the surprisingly high apparent fraction of DSBs in repeats in the B6 strain (**Fig. 3**), it may take only a modest further increase in this burden to trigger catastrophic meiotic defects.

Nonetheless, it is noteworthy that PRDM9 binding sites are enriched in certain classes of repeated sequence, for example in human L2 LINEs, AluY elements, and the retrovirus-like retrotransposons THE1A and THE1B.[20,73] Thus, paradoxically, it may be that PRDM9 functions include the preferential targeting of certain repeats, perhaps as part of the cell's attempts to limit the spread of certain selfish genetic elements. If so, this function must be balanced against formation of too many DSBs in repeats. We speculate that this balance is an important component of the remarkable evolutionary dynamics of PRDM9, including multiple independent apparent losses of *Prdm9* genes in a number of vertebrate lineages.[74-78]

Histone H3 methylation and the presence of ssDNA
Single-stranded DNA sequencing (SSDS) maps meiotic DSBs by immunopurification and deep-sequencing of DMC1-bound ssDNA from spermatocytes.[21] Because SSDS captures an intermediate that arises after resection, SSDS and SPO11-oligo data can be combined to explore resection.[26] Nucleosomes can impede resection by exonucleases *in vitro*,[79] and chromatin structure strongly influences meiotic resection patterns *in vivo* in *S. cerevisiae*.[13] Given the pronounced asymmetry in PRDM9-dependent histone H3 methylation at many hotspots (**Fig. 1B**), we asked if this feature of the local chromatin structure was correlated with resection patterns. Interestingly, at asymmetrically methylated hotspots, SSDS signal strength was 8% higher on average on the sides of the hotspots with highest H3K4me3 (and H3K36me3) ChIP signal (**Fig. 4A–C**). The distance over which the SSDS signal spread away from hotspots was not apparently different, however, thus the lengths of resection tracts are not correlated with H3 methylation status. [Note that the H3 methylation ChIP signal does not allow conclusions to be drawn about total nucleosome occupancy (methylated plus unmethylated).]

SSDS signal strength reflects both the number of DSBs in the population and the lifespan of DMC1-bound ssDNA.[26] SSDS signal on the left side of hotspots derives from precisely the same DSBs as the SSDS signal on the right, therefore the observed left-versus-right asymmetry in SSDS signal cannot be ascribed to differences in DSB frequency or timing of DSB formation. Instead, the asymmetry may indicate that DMC1 remains bound for a shorter time on average on the low-methylation side of hotspots. For example, this side might tend to engage in strand exchange and DMC1 dissociation earlier than the high-methylation side. Effects on recombination efficiency have been inferred for PRDM9-dependent histone methylation of the intact recombination partner.[24] Alternatively, the asymmetry might reflect systematic differences in the amount of DMC1 bound to the ssDNA on the two sides. We also cannot exclude the possibility that the asymmetry derives from technical biases that affect coverage maps in both SSDS and H3 methylation ChIP experiments.

Non-centromeric ends of autosomes display unusually low SSDS:SPO11-oligo ratios
Previously, we found that hotspots on sex chromosomes show a markedly higher ratio of SSDS to SPO11-oligo counts than hotspots on autosomes.[26] SPO11 oligos have a long lifespan relative to the length of prophase,[43] so (unlike SSDS) SPO11-oligo mapping is thought to be relatively insensitive to variation in timing of DSB formation or lifespan of DSBs. Thus, it was inferred that DSBs have a longer average lifespan on the non-homologous parts of the sex chromosomes than on autosomes, presumably because of delayed repair caused by absence

7

of a homologous chromosome plus a temporary barrier to using the sister chromatid as a recombination partner.[23,26]

Following on this precedent, we extended the comparative analysis of SPO11-oligo and SSDS maps to large domains of autosomes. Mouse chromosomes are acrocentric, i.e., their centromeres lie close to one end of each chromosome. On average, the centromere-proximal 40 megabases (Mb) of autosomes showed similar patterns in both mapping methods, that is, the ratio of SSDS signal to SPO11-oligo counts was indistinguishable from genome average (**Fig. 4D**). (Note that this analysis excludes the large blocks of repetitive satellite DNA that make up the pericentric heterochromatin. These regions are not readily accessible to deep-sequencing methods, but are also known from cytological data to experience few if any meiotic DSBs.[80]) In contrast, the 5 Mb adjacent to centromere-distal telomeres displayed lower mean SSDS coverage than expected from the SPO11-oligo density (**Fig. 4D–E**). This finding suggests that there are systematic differences in the formation, processing, and/or repair of DSBs in these sub-telomeric zones relative to interstitial regions. This pattern could be explained if DSBs tend to form later near telomeres than elsewhere, or if they form with similar timing but with shorter resection distance or less DMC1 loading. Alternatively, sub-telomeric DSBs may tend to be repaired more quickly, thereby reducing the lifespan of DMC1-bound ssDNA. Different repair kinetics could arise from regional differences in the physical proximity of homologs, for example, via tethering to the nuclear envelope.[81]

Reciprocal crossover asymmetry and polymorphisms in the PRDM9 binding motif

Both crossover and noncrossover recombination outcomes can be accompanied by gene conversion of allelic differences around the DSB site (**Fig. 5A**). For crossovers, presence of a gene conversion tract causes the breakpoints on the two recombinant chromosomes to lie at different positions, flanking the conversion tract (**Fig. 5A**). In mice, crossover breakpoints at several recombination hotspots have been characterized by fine-scale analyses of recombinant products in sperm DNA isolated from F_1 hybrids of B6 with other strains.[82-88] In these studies, breakpoints were assayed after PCR amplification of recombinant molecules using allele-specific primers in each of the two orientations (i.e., forward primers specific for B6 polymorphisms with reverse primers specific for the other strain; and vice versa). If DSB formation is about equally likely on the two homologous chromosomes, then the distribution of crossover breakpoints will be similar for both PCR orientations. But if DSB formation is biased in favor of one allele, then the breakpoint distribution detected in one orientation will be shifted relative to breakpoints amplified in the other orientation (**Fig. 5A**).[84,89,90] Such reciprocal crossover asymmetry was observed at the *A3* hotspot in B6×DBA/2J ("DBA") hybrids and was explained by a sequence polymorphism that resulted in differences in the binding of PRDM9 to the B6 and DBA alleles at *A3*.[84,91]

We asked if polymorphisms within PRDM9 binding sites could explain crossover asymmetries observed at other hotspots. Of the 15 published crossover hotspots we examined previously,[26] 13 contained at least one putative PRDM9 binding motif in B6 within 250 bp of the center of a matched SPO11-oligo hotspot. In all 13 cases, a motif was similarly identified in the allelic sequence in the other strain of the tested F_1 hybrid (DBA or A/J). Four of these crossover hotspots were excluded from further analysis: three for which crossovers had been assessed in only one orientation and one that encompassed two prominent SPO11-oligo clusters. For the remaining nine crossover hotspots, we examined putative PRDM9 binding sites — defined as the 36 bp encompassing the hotspot-enriched motif[26] — for polymorphisms between B6 and DBA or A/J (**Fig. 5B–D**).

At *HS59.5*, where the skew in the distribution of crossover breakpoints was 476 bp (**Fig. 5Biii–Biv**), we observed a B6–DBA sequence polymorphism in the putative PRDM9 binding site (**Fig. 5Bi**). In contrast, *HS61.1* displayed no B6–DBA sequence differences in the PRDM9 binding site, and showed essentially no crossover asymmetry (**Fig. 5C**). This pattern extended

to all hotspots examined: crossover asymmetry was minimal for three hotspots where the two strains had identical putative PRDM9 binding sites, but each of the six strains with a polymorphism(s) showed reciprocal crossover asymmetry ranging from 225 to 659 bp (**Fig. 5D** and **Table S3**). We infer that the skew in the distribution of crossover breakpoints reflects differential binding of PRDM9 to, and histone trimethylation of, each pair of alleles. *In vitro* assays would be useful for comparing the binding efficiencies of PRDM9 to each allele. We note that in many cases the only polymorphisms present are outside the most conserved nucleotides of the core motif, consistent with non-core positions contributing to DNA binding affinity *in vivo*, as has been shown *in vitro*.[92]

Materials and methods

Datasets
We used SPO11-oligo, SSDS and H3K4me3 data from GEO accession numbers GSE84689, GSE35498 and GSE52628, respectively.[21,22,26] Pre-processing of these data is described elsewhere.[26] H3K36me3 and RNA-seq data were downloaded from GSE76416 and GSE61613, respectively,[32,33] and were converted to mouse genome assembly GRCm38/mm10 sequence coordinates using the UCSC Genome Browser LiftOver tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver).

Quantification and statistical analyses
Statistical analyses were performed using R versions 3.2.3 to 3.3.1 (http://www.r-project.org). Statistical parameters and tests are reported in the figures and legends. In cases where outliers were removed for plotting purposes, none of the data points were removed from the statistical analysis. For crossover breakpoint analyses, cumulative breakpoint distributions with fitted Gaussian curves were determined using GraphPad Prism version 7.

Base composition analysis
To examine base composition at 5′ ends of SPO11 oligos, we used all uniquely mapped reads from the wild-type B6 dataset. The same conclusions were reached when we used only reads without a 5′-C ambiguity. To compare base composition within and outside of PRDM9 binding sites, we focused on the 201 bp around the peaks of 9,060 SPO11-oligo hotspots with an identified primary PRDM9 motif.[26] To circumvent any contribution from the 12-bp PRDM9 core motif sequence, we eliminated all reads that overlapped part or all of the 9,865 instances of the motifs (some hotspots had more than one motif). Dinucleotide frequency and correlations for two-fold rotational symmetry were calculated as reported previously.[4]

Repetitive sequence analysis
Interspersed repeat sequences such as transposons and simple sequence repeats were as annotated by RepeatMasker. Segmental duplications were as defined by Eichler and colleagues.[58] Tables of RepeatMasker repeats and segmental duplications for mouse genome assembly GRCm38/mm10 were downloaded from the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables). To estimate expected SPO11-oligo counts for **Fig. 3E**, we shuffled repeat locations one million times within the mappable portion of the genome and tallied the SPO11 oligos that fell within the shuffled positions. Because of the substantial difference in overall SPO11-oligo densities between autosomes and sex chromosomes,[26] autosomal repeats were shuffled among autosomes and sex-chromosome repeats were shuffled within their chromosome of origin. We excluded unplaced and unassigned contigs from the analyses. Note that RepeatMasker includes annotations for "rRNA" sequences, but the ribosomal DNA repeat is not assembled in the reference genome (GRCm38/mm10). We

9

excluded these elements from the analysis of specific interspersed repeat element families in **Fig. 3E** but included them for the global analyses in **Fig. 3A–D**.

PRDM9 binding motifs in crossover hotspots

To identify instances of the previously published[26] primary or secondary PRDM9 motif around 15 crossover hotspots, we used MAST[93] to query 501-bp sequences in B6 and A/J or DBA around the centers of matched SPO11-oligo hotspots (**Table S3**). For crossover hotspot *HS18.2*, which does not have a matched SPO11-oligo hotspot, we queried the 501-bp sequence around the midpoint of the region encompassing all crossover breakpoints. This query identified primary and/or secondary PRDM9 motifs in 13 of the 15 crossover hotspots.

To assess crossover asymmetry in hotspots with PRDM9 motifs, we used nine informative loci for which crossover data were available in both orientations and SPO11 oligos were clustered to one zone. The nine crossover hotspots were on chromosome 1 (*A3*, central and distal) and chromosome 19 (*HS18.2*, *HS22*, *HS23.9*, *HS59.5*, *HS61.1* and *HS61.2*). At each of the nine hotspots, we re-mapped previously published crossover data[82,84,86-88] to B6 coordinates (GRCm38/mm10), then plotted cumulative distributions of crossover breakpoints in each orientation with fitted Gaussian curves. The skew at each hotspot was determined by the distance between the two curves at the midpoint for each cumulative plot.

Abbreviations
DSB   double-strand break
ssDNA   single-stranded DNA
ChIP   chromatin immunoprecipitation
NAHR   non-allelic homologous recombination
SSDS   single-stranded DNA sequencing

Disclosure of potential conflicts of interest
No potential conflicts of interest were disclosed.

Acknowledgments

Funding

REFERENCES

[1]   Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. Nat Rev Genet 2013; 14:794-806; PMID:24136506; http://dx.doi.org/10.1038/nrg3573

[2]   Nagaoka SI, Hassold TJ, Hunt PA. Human aneuploidy: mechanisms and new insights into an age-old problem. Nat Rev Genet 2012; 13:493-504; PMID:22705668; http://dx.doi.org/10.1038/nrg3245

[3]   Kim S, Peterson SE, Jasin M, Keeney S. Mechanisms of germ line genome instability. Semin Cell Dev Biol 2016; 54:177-87; PMID:26880205; http://dx.doi.org/10.1016/j.semcdb.2016.02.019

[4]   Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, et al. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 2011; 144:719-31; PMID:21376234; http://dx.doi.org/10.1016/j.cell.2011.02.009

[5]   de Massy B. Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. Annu Rev Genet 2013; 47:563-99; PMID:24050176; http://dx.doi.org/10.1146/annurev-genet-110711-155423

[6]   Lam I, Keeney S. Mechanism and regulation of meiotic recombination initiation. Cold Spring Harb Perspect Biol 2014; 7:a016634; PMID:25324213; http://dx.doi.org/10.1101/cshperspect.a016634

[7]   Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature 1997; 386:414-7; PMID:9121560; http://dx.doi.org/10.1038/386414a0

[8]   Keeney S, Giroux CN, Kleckner N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 1997; 88:375-84; PMID:9039264; http://dx.doi.org/10.1016/S0092-8674(00)81876-0

[9]   Neale MJ, Pan J, Keeney S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature 2005; 436:1053-7; PMID:16107854; http://dx.doi.org/10.1038/nature03872

[10]   Cao L, Alani E, Kleckner N. A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. Cell 1990; 61:1089-101; PMID:2190690; http://dx.doi.org/10.1016/0092-8674(90)90072-M

[11]   Sun H, Treco D, Szostak JW. Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. Cell 1991; 64:1155-61; PMID:2004421; http://dx.doi.org/10.1016/0092-8674(91)90270-9

[12]   Zakharyevich K, Ma Y, Tang S, Hwang PY, Boiteux S, Hunter N. Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. Mol Cell 2010; 40:1001-15; PMID:21172664; http://dx.doi.org/10.1016/j.molcel.2010.11.032

[13]   Mimitou EP, Yamada S, Keeney S. A global view of meiotic double-strand break end resection. Science 2017; 355:40-5; PMID:28059759; http://dx.doi.org/10.1126/science.aak9704

[14]   Bishop DK, Park D, Xu L, Kleckner N. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. Cell 1992; 69:439-56; PMID:1581960; http://dx.doi.org/10.1016/0092-8674(92)90446-J

[15]   Shinohara A, Ogawa H, Ogawa T. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell 1992; 69:457-70; PMID:1581961; http://dx.doi.org/10.1016/0092-8674(92)90447-K

[16]   Schwacha A, Kleckner N. Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. Cell 1997; 90:1123-35; PMID:9323140; http://dx.doi.org/10.1016/S0092-8674(00)80378-5

[17]   Cloud V, Chan YL, Grubb J, Budke B, Bishop DK. Rad51 is an accessory factor for Dmc1-mediated joint molecule formation during meiosis. Science 2012; 337:1222-5; PMID:22955832; http://dx.doi.org/10.1126/science.1219379

[18]   Hunter N. Meiotic recombination: the essence of heredity. Cold Spring Harb Perspect Biol 2015; 7; PMID:26511629; http://dx.doi.org/10.1101/cshperspect.a016618

[19]   Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 2010; 327:836-40; PMID:20044539; http://dx.doi.org/10.1126/science.1183439

[20]   Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 2010; 327:876-9; PMID:20044541; http://dx.doi.org/10.1126/science.1182363

[21]   Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. Genetic recombination is directed away from functional genomic elements in mice. Nature 2012; 485:642-5; PMID:22660327; http://dx.doi.org/10.1038/nature11089

[22]   Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. Genome Res 2014; 24:724-32; PMID:24604780; http://dx.doi.org/10.1101/gr.170167.113

[23]   Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. Recombination initiation maps of individual human genomes. Science 2014; 346:1256442; PMID:25395542; http://dx.doi.org/10.1126/science.1256442

[24]   Davies B, Hatton E, Altemose N, Hussin JG, Pratto F, Zhang G, Hinch AG, Moralli D, Biggs D, Diaz R, et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature 2016; 530:171-6; PMID:26840484; http://dx.doi.org/10.1038/nature16931

[25]   Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. The evolutionary turnover of recombination hot spots contributes to speciation in mice. Genes Dev 2016; 30:266-80; PMID:26833728; http://dx.doi.org/10.1101/gad.270009.115

[26]   Lange J, Yamada S, Tischfield SE, Pan J, Kim S, Zhu X, Socci ND, Jasin M, Keeney S. The landscape of mouse meiotic double-strand break formation, processing, and repair. Cell 2016; 167:695-708.e16; PMID:27745971; http://dx.doi.org/10.1016/j.cell.2016.09.035

[27]   Grey C, Barthes P, Chauveau-Le Friec G, Langa F, Baudat F, de Massy B. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. PLoS Biol 2011; 9:e1001176; PMID:22028627; http://dx.doi.org/10.1371/journal.pbio.1001176

[28]   Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature 2011; 472:375-8; PMID:21460839; http://dx.doi.org/10.1038/nature09869

[29]   Baker CL, Kajita S, Walker M, Saxl RL, Raghupathy N, Choi K, Petkov PM, Paigen K. PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination. PLoS Genet 2015; 11:e1004916; PMID:25568937; http://dx.doi.org/10.1371/journal.pgen.1004916

[30]   Wu H, Mathioudakis N, Diagouraga B, Dong A, Dombrovski L, Baudat F, Cusack S, de Massy B, Kadlec J. Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. Cell Rep 2013; 5:13-20; PMID:24095733; http://dx.doi.org/10.1016/j.celrep.2013.08.035

[31]   Eram MS, Bustos SP, Lima-Fernandes E, Siarheyeva A, Senisterra G, Hajian T, Chau I, Duan S, Wu H, Dombrovski L, et al. Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. J Biol Chem 2014; 289:12177-88; PMID:24634223; http://dx.doi.org/10.1074/jbc.M113.523183

[32]   Powers NR, Parvanov ED, Baker CL, Walker M, Petkov PM, Paigen K. The meiotic recombination activator PRDM9 trimethylates both H3K36 and H3K4 at recombination hotspots in vivo. PLoS Genet 2016; 12:e1006146; PMID:27362481; http://dx.doi.org/10.1371/journal.pgen.1006146

[33]   Walker M, Billings T, Baker CL, Powers N, Tian H, Saxl RL, Choi K, Hibbs MA, Carter GW, Handel MA, et al. Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. Epigenetics Chromatin 2015; 8:31; PMID:26351520; http://dx.doi.org/10.1186/s13072-015-0024-6

[34]   Petes TD. Meiotic recombination hot spots and cold spots. Nat Rev Genet 2001; 2:360-9; PMID:11331902; http://dx.doi.org/10.1038/35072078

[35]   Kauppi L, Jeffreys AJ, Keeney S. Where the crossovers are: recombination distributions in mammals. Nat Rev Genet 2004; 5:413-24; PMID:15153994; http://dx.doi.org/10.1038/nrg1346

[36]   Thacker D, Mohibullah N, Zhu X, Keeney S. Homologue engagement controls meiotic DNA break number and distribution. Nature 2014; 510:241-6; PMID:24717437; http://dx.doi.org/10.1038/nature13120

[37]   Ohta K, Shibata T, Nicolas A. Changes in chromatin structure at recombination initiation sites during yeast meiosis. EMBO J 1994; 13:5754-63; PMID:7988571;

[38]   Wu TC, Lichten M. Factors that affect the location and frequency of meiosis-induced double-strand breaks in Saccharomyces cerevisiae. Genetics 1995; 140:55-66; PMID:7635308;

[39]   Jessop L, Allers T, Lichten M. Infrequent co-conversion of markers flanking a meiotic recombination initiation site in Saccharomyces cerevisiae. Genetics 2005; 169:1353-67; PMID:15654098; http://dx.doi.org/10.1534/genetics.104.036509

[40]   Mohibullah N, Keeney S. Numerical and spatial patterning of yeast meiotic DNA breaks by Tel1. Genome Res 2017; 27:278-88; PMID:27923845; http://dx.doi.org/10.1101/gr.213587.116

[41]   Keeney S, Lange J, Mohibullah N. Self-organization of meiotic recombination initiation: general principles and molecular pathways. Annu Rev Genet 2014; 48:187-214; PMID:25421598; http://dx.doi.org/10.1146/annurev-genet-120213-092304

[42]   Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nat Rev Mol Cell Biol 2013; 14:197-210; PMID:23486281; http://dx.doi.org/10.1038/nrm3546

[43]   Lange J, Pan J, Cole F, Thelen MP, Jasin M, Keeney S. ATM controls meiotic double-strand-break formation. Nature 2011; 479:237-40; PMID:22002603; http://dx.doi.org/10.1038/nature10508

[44]   Zhang L, Kim KP, Kleckner NE, Storlazzi A. Meiotic double-strand breaks occur once per pair of (sister) chromatids and, via Mec1/ATR and Tel1/ATM, once per quartet of chromatids. Proc Natl Acad Sci U S A 2011; 108:20036-41; PMID:22123968; http://dx.doi.org/10.1073/pnas.1117937108

[45]   Garcia V, Gray S, Allison RM, Cooper TJ, Neale MJ. Tel1(ATM)-mediated interference suppresses clustered meiotic double-strand-break formation. Nature 2015; 520:114-8; PMID:25539084; http://dx.doi.org/10.1038/nature13993

[46]   de Massy B, Rocco V, Nicolas A. The nucleotide mapping of DNA double-strand breaks at the CYS3 initiation site of meiotic recombination in Saccharomyces cerevisiae. EMBO J 1995; 14:4589-98; PMID:7556102;

[47]   Liu J, Wu TC, Lichten M. The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. EMBO J 1995; 14:4599-608; PMID:7556103;

[48]   Murakami H, Nicolas A. Locally, meiotic double-strand breaks targeted by Gal4BD-Spo11 occur at discrete sites with a sequence preference. Mol Cell Biol 2009; 29:3500-16; PMID:19380488; http://dx.doi.org/10.1128/MCB.00088-09

[49]   Fowler KR, Sasaki M, Milman N, Keeney S, Smith GR. Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. Genome Res 2014; 24:1650-64; PMID:25024163; http://dx.doi.org/10.1101/gr.172122.114

[50]   Nichols MD, DeAngelis K, Keck JL, Berger JM. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. EMBO J 1999; 18:6177-88; PMID:10545127; http://dx.doi.org/10.1093/emboj/18.21.6177

[51]   Patel A, Horton JR, Wilson GG, Zhang X, Cheng X. Structural basis for human PRDM9 action at recombination hot spots. Genes Dev 2016; 30:257-65; PMID:26833727; http://dx.doi.org/10.1101/gad.274928.115

[52]   Garcia V, Phelps SE, Gray S, Neale MJ. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 2011; 479:241-4; PMID:22002605; http://dx.doi.org/10.1038/nature10515

[53]   Cannavo E, Cejka P. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature 2014; 514:122-5; PMID:25231868; http://dx.doi.org/10.1038/nature13771

[54]   Anand R, Ranjha L, Cannavo E, Cejka P. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. Mol Cell 2016; 64:940-50; PMID:27889449; http://dx.doi.org/10.1016/j.molcel.2016.10.017

[55]   Deshpande RA, Lee JH, Arora S, Paull TT. Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. Mol Cell 2016; 64:593-606; PMID:27814491; http://dx.doi.org/10.1016/j.molcel.2016.10.010

[56]   Striedner Y, Schwarz T, Welte T, Futschik A, Rant U, Tiemann-Boege I. The long zinc finger domain of PRDM9 forms a highly stable and long-lived complex with its DNA recognition sequence. Chromosome Res 2017; PMID:28155083; http://dx.doi.org/10.1007/s10577-017-9552-1

[57]   Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002; 420:520-62; PMID:12466850; http://dx.doi.org/10.1038/nature01262

[58]   She X, Cheng Z, Zollner S, Church DM, Eichler EE. Mouse segmental duplication and copy number variation. Nat Genet 2008; 40:909-14; PMID:18500340; http://dx.doi.org/10.1038/ng.172

[59]   Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 2009; 10:451-81; PMID:19715442; http://dx.doi.org/10.1146/annurev.genom.9.081307.164217

[60]   Sasaki M, Lange J, Keeney S. Genome destabilization by homologous recombination in the germ line. Nat Rev Mol Cell Biol 2010; 11:182-95; PMID:20164840; http://dx.doi.org/10.1038/nrm2849

[61]   Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 2016; 17:224-38; PMID:26924765; http://dx.doi.org/10.1038/nrg.2015.25

[62]   Sasaki M, Tischfield SE, van Overbeek M, Keeney S. Meiotic recombination initiation in and around retrotransposable elements in Saccharomyces cerevisiae. PLoS Genet 2013; 9:e1003732; PMID:24009525; http://dx.doi.org/10.1371/journal.pgen.1003732

[63]   Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, Bourc'his D. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev 2015; 29:1256-70; PMID:26109049; http://dx.doi.org/10.1101/gad.257840.114

[64]   Kauppi L, Barchi M, Lange J, Baudat F, Jasin M, Keeney S. Numerical constraints and feedback control of double-strand breaks in mouse meiosis. Genes Dev 2013; 27:873-86; PMID:23599345; http://dx.doi.org/10.1101/gad.213652.113

[65]   Goldman AS, Lichten M. Restriction of ectopic recombination by interhomolog interactions during Saccharomyces cerevisiae meiosis. Proc Natl Acad Sci U S A 2000; 97:9537-42; PMID:10944222;

[66]   Shinohara M, Shinohara A. Multiple pathways suppress non-allelic homologous recombination during meiosis in Saccharomyces cerevisiae. PLoS One 2013; 8:e63144; PMID:23646187; http://dx.doi.org/10.1371/journal.pone.0063144

[67]   Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 2009; 323:373-5; PMID:19074312; http://dx.doi.org/10.1126/science.1163601

[68]   Flachs P, Mihola O, Simecek P, Gregorova S, Schimenti JC, Matsui Y, Baudat F, de Massy B, Pialek J, Forejt J, et al. Interallelic and intergenic incompatibilities of the Prdm9 (Hst1) gene in mouse hybrid sterility. PLoS Genet 2012; 8:e1003044; PMID:23133405; http://dx.doi.org/10.1371/journal.pgen.1003044

[69]   Balcova M, Faltusova B, Gergelits V, Bhattacharyya T, Mihola O, Trachtulec Z, Knopf C, Fotopulosova V, Chvatalova I, Gregorova S, et al. Hybrid Sterility Locus on Chromosome X Controls Meiotic Recombination Rate in Mouse. PLoS Genet 2016; 12:e1005906; PMID:27104744; http://dx.doi.org/10.1371/journal.pgen.1005906

[70]   Faieta M, Di Cecca S, de Rooij DG, Luchetti A, Murdocca M, Di Giacomo M, Di Siena S, Pellegrini M, Rossi P, Barchi M. A surge of late-occurring meiotic double-strand breaks rescues synapsis abnormalities in spermatocytes of mice with hypomorphic expression of SPO11. Chromosoma 2016; 125:189-203; PMID:26440409; http://dx.doi.org/10.1007/s00412-015-0544-7

[71]   Kauppi L, Jasin M, Keeney S. How much is enough? Control of DNA double-strand break numbers in mouse meiosis. Cell Cycle 2013; 12:2719-20; PMID:23966150; http://dx.doi.org/10.4161/cc.26079

[72]   Faisal I, Kauppi L. Sex chromosome recombination failure, apoptosis, and fertility in male mice. Chromosoma 2016; 125:227-35; PMID:26440410; http://dx.doi.org/10.1007/s00412-015-0542-9

[73]   Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif associated with recombination hot spots and genome instability in humans. Nat Genet 2008; 40:1124-9; PMID:19165926; http://dx.doi.org/10.1038/ng.213

[74]   Munoz-Fuentes V, Di Rienzo A, Vila C. Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes. PLoS One 2011; 6:e25498; PMID:22102853; http://dx.doi.org/10.1371/journal.pone.0025498

[75]   Axelsson E, Webster MT, Ratnakumar A, Consortium L, Ponting CP, Lindblad-Toh K. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res 2012; 22:51-63; PMID:22006216; http://dx.doi.org/10.1101/gr.124123.111

[76]   Buard J, Rivals E, Dunoyer de Segonzac D, Garres C, Caminade P, de Massy B, Boursot P. Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. PLoS One 2014; 9:e85021; PMID:24454780; http://dx.doi.org/10.1371/journal.pone.0085021

[77]   Schwartz JJ, Roach DJ, Thomas JH, Shendure J. Primate evolution of the recombination regulator PRDM9. Nat Commun 2014; 5:4370; PMID:25001002; http://dx.doi.org/10.1038/ncomms5370

[78]   Baker Z, Schumer M, Haba Y, Holland C, Rosenthal GG, Przeworski M. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. bioRxiv 2017; http://dx.doi.org/10.1101/093096

[79]   Adkins NL, Niu H, Sung P, Peterson CL. Nucleosome dynamics regulates DNA processing. Nat Struct Mol Biol 2013; 20:836-42; PMID:23728291; http://dx.doi.org/10.1038/nsmb.2585

[80]   Verver DE, van Pelt AM, Repping S, Hamer G. Role for rodent Smc6 in pericentromeric heterochromatin domains during spermatogonial differentiation and meiosis. Cell Death Dis 2013; 4:e749; PMID:23907463; http://dx.doi.org/10.1038/cddis.2013.269

[81]   Scherthan H. Telomere attachment and clustering during meiosis. Cell Mol Life Sci 2007; 64:117-24; PMID:17219025; http://dx.doi.org/10.1007/s00018-006-6463-2

[82]   Bois PR. A highly polymorphic meiotic recombination mouse hot spot exhibits incomplete repair. Mol Cell Biol 2007; 27:7053-62; PMID:17709383; http://dx.doi.org/10.1128/MCB.00874-07

[83]   Kauppi L, Jasin M, Keeney S. Meiotic crossover hotspots contained in haplotype block boundaries of the mouse genome. Proc Natl Acad Sci U S A 2007; 104:13396-401; PMID:17690247; http://dx.doi.org/10.1073/pnas.0701965104

[84]   Cole F, Keeney S, Jasin M. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. Mol Cell 2010; 39:700-10; PMID:20832722; http://dx.doi.org/10.1016/j.molcel.2010.08.017

[85]   Getun IV, Wu ZK, Khalil AM, Bois PR. Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. EMBO Rep 2010; 11:555-60; PMID:20508641; http://dx.doi.org/10.1038/embor.2010.79

[86]   Wu ZK, Getun IV, Bois PR. Anatomy of mouse recombination hot spots. Nucleic Acids Res 2010; 38:2346-54; PMID:20081202; http://dx.doi.org/10.1093/nar/gkp1251

[87]   Getun IV, Wu ZK, Bois PR. Organization and roles of nucleosomes at mouse meiotic recombination hotspots. Nucleus 2012; 3:244-50; PMID:22572955; http://dx.doi.org/10.4161/nucl.20325

[88]   de Boer E, Jasin M, Keeney S. Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. Genes Dev 2015; 29:1721-33; PMID:26251527; http://dx.doi.org/10.1101/gad.265561.115

[89]   Jeffreys AJ, Neumann R. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nat Genet 2002; 31:267-71; PMID:12089523; http://dx.doi.org/10.1038/ng910

[90]   Arnheim N, Calabrese P, Tiemann-Boege I. Mammalian meiotic recombination hot spots. Annu Rev Genet 2007; 41:369-99; PMID:18076329; http://dx.doi.org/10.1146/annurev.genet.41.110306.130301

[91]   Cole F, Baudat F, Grey C, Keeney S, de Massy B, Jasin M. Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. Nat Genet 2014; 46:1072-80; PMID:25151354; http://dx.doi.org/10.1038/ng.3068

[92]   Billings T, Parvanov ED, Baker CL, Walker M, Paigen K, Petkov PM. DNA binding specificities of the long zinc-finger recombination protein PRDM9. Genome Biol 2013; 14:R35; PMID:23618393; http://dx.doi.org/10.1186/gb-2013-14-4-r35

[93]   Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 1998; 14:48-54; PMID:9520501; http://dx.doi.org/10.1093/bioinformatics/14.1.48

Figure Legends

Figure 1. Spatial relationships between H3K36 trimethylation and DSBs

(A) H3K36me3 (data from ref.[32]) has a similar profile as H3K4me3 (data from ref.[22]) around SPO11-oligo hotspots. Data were locally normalized by dividing the signal at each base pair (bp) by the mean signal within each 2,001-bp window, then were averaged across hotspots. The SPO11-oligo profile was smoothed with a 51-bp Hann filter.

(B) H3K36me3 signal is often highly asymmetric around hotspots in a manner similar to H3K4me3. Heat maps (data in 5-bp bins after local normalization) were ordered according to H3K4me3 asymmetry. Because data in each hotspot were locally normalized, color-coding reflects the local spatial pattern, not relative signal strength between hotspots.

(C) Similar asymmetric patterns between H3K4me3 and H3K36me3 at SPO11-oligo hotspots. Each panel shows the mean of locally normalized profiles (51-bp Hann filter for SPO11-oligo data) across the 20% of hotspots with the most asymmetric H3K4me3 patterns (left > right in top panel; right > left in bottom panel).

(D) In three classes of PRDM9 motifs previously defined according to local SPO11-oligo pattern,[26] H3K36me3 patterns are similar. The sequence logo shows the 12-bp core PRDM9 motif identified by MEME within SPO11-oligo hotspots;[26] the light gray bar denotes the larger 36-bp segment of DNA thought to be bound by PRDM9.[21,29] SPO11-oligo profiles were smoothed with a 15-bp Hann filter.

(E) Similar H3K36me3 signal strength for hotspots in each of the three PRDM9 motif classes, as observed for H3K4me3. H3K4me3 tag counts and H3K36me3 coverage were summed in 1,001-bp windows around hotspot centers. In the box plots, thick horizontal lines indicate medians, box edges show the 25th and 75th percentiles, and whiskers indicate lowest and highest values within 1.5-fold of the interquartile range; outliers are not shown. A value of 1 was added to each hotspot to permit plotting of hotspots with no H3K4me3 or H3K36me3 signal.

(F) H3K36me3 is an imperfect predictor of DSB frequency. SPO11-oligo counts and H3K36me3 coverage were summed in 1,001-bp windows around hotspot centers. One H3K36me3 count was added to each hotspot to permit plotting of hotspots with no H3K36me3 signal.

(G–H) The effect of ATM deficiency on hotspot activity is independent of H3K4me3 levels. SPO11-oligo counts in B6, *Atm* null and *Atm* wt and H3K4me3 tag counts were summed in 1,001-bp windows around B6 hotspot centers. The ratio of SPO11-oligo counts in *Atm* null to *Atm* wt was plotted against SPO11-oligo counts in B6 (G) or H3K4me3 counts (H). One count was added to each hotspot in *Atm* null and *Atm* wt SPO11-oligo data and H3K4me3 data to permit plotting of hotspots with no signal.

(I) Fit of a multiple regression model predicting SPO11-oligo counts in hotspots from H3K4me3, H3K36me3, and *Atm* null:*Atm* wt ratio (**Table S1**).

Panels depicting SPO11-oligo and H3K4me3 ChIP data were adapted from ref [26] with permission. In panels showing H3K36me3 signal around SPO11-oligo hotspots and PRDM9 motifs, data were plotted starting from values of 0.56 to facilitate side-by-side comparison with H3K4me3 data. All R^2 values were determined by linear regression on log-transformed data.

Figure 2. Sequence composition at sites of SPO11-oligo formation and 3′ nicking

(A) (top) Schematic of a SPO11 DSB. Staggered cuts (arrows) by a SPO11 dimer generate two-nucleotide 5′ overhangs, in the middle of which is a two-fold axis of rotational symmetry. (bottom) Non-random dinucleotide composition around SPO11 cleavage sites in 9,060 SPO11-oligo hotspots. At each position, deviation of dinucleotide frequencies from local average was summed.

(B) Mononucleotide composition around the 5′ ends of uniquely mapped SPO11 oligos within hotspots. Purple, all cleavage sites; blue, SPO11 oligos with the 5′-mapped end in a PRDM9

motif and without a 5′ C; red, SPO11 oligos without an overlap to a PRDM9 motif and without a 5′ C. Plots are truncated for values outside of the range from -1 to 1.

(C) Dinucleotide base composition is not rotationally symmetric around the dyad axis. The dinucleotide composition on the left of cleavage sites was compared to the composition on the right for each of the 256 possible dinucleotide pairs. The correlation coefficients (Pearson's *r*) were then rank-ordered and displayed as a bar plot with reverse-complementary dinucleotides colored in red. The mean ± SD of the *r* values is shown for reverse-complementary and non-reverse-complementary dinucleotides.

(D) The distributions of 3′ ends of SPO11 oligos relative to 5′ ends are similar for the three classes of PRDM9 motifs defined by local SPO11-oligo pattern (see **Fig. 1D**). The 5′ and 3′ ends of SPO11-oligo profiles were smoothed with a 15-bp Hann filter.

Figure 3. DSBs in repeated sequences

(A) Five categories of SPO11-oligo mappability. 67.2% of SPO11 oligos could be assigned unambiguously to a unique location in the reference genome within either the non-repeat class (dark gray) or the repeat class (dark turquoise). The remaining 32.8% mapped to multiple places in the reference genome; half of these could be placed in either the non-repeat class (i.e., all mapped positions are in non-repeat sequences; light gray) or the repeat class (i.e., all mapped positions are in the repeat class; turquoise). However, the remaining half of the multi-mappers (16.3% of all mapped reads) could have derived from either class because reads mapped to both repeated and non-repeated sequences (light turquoise).

(B) (top) Histogram of SPO11-oligo read lengths (adapted from ref [26] with permission). (bottom) Percentages of the five categories as a function of read length. Short reads (< 20 nt) were highly enriched for multi-mapped reads, suggesting that many such SPO11 oligos were mapped to multiple places solely because of their short length.

(C) Percentages of SPO11 oligos mapped to repeat sequences, estimated by four methods. Imputed: Reads were assigned in proportion to the number of unique reads in the neighboring area (method described in ref.[26]). Normalized: Each multi-mapped read was divided evenly among its mapped positions. Stringent: All multi-mapped reads that could have derived from either repeat or non-repeat sequences were assigned entirely to that respective repeat class. This approach likely overestimates the DSB frequency in repeat sequences and therefore provides the most conservative estimate of the degree to which DSB formation is suppressed in repeats. Imputed read length: Because the mappability of short reads is less certain (B), we included from the imputed map only reads with lengths of 25–50 nt and recalculated the fraction of imputed reads in repeated sequences. This method represents the highest confidence estimate of the relative burden of DSBs in sequences annotated as repeats.

(D) DSB frequencies in two sub-groups of repeat sequences. SPO11-oligo frequencies were lower than expected in interspersed repeats and low complexity DNA sequences ("RepeatMasker") and in segmental duplications.

(E) DSB frequencies in families of repeat sequences annotated in RepeatMasker. Overlapping repeats of the same family were merged before calculation. For each repeat family, SPO11-oligo counts per base pair were summed and their enrichment was calculated relative to expected values (see Materials and methods). The most extreme examples are labeled, as are L1 and ERVK families, which include elements that are putative DNMT3L targets.[63] The identities, calculated enrichment values, and fractions of the genome occupied for all of the families analyzed are provided in **Table S2**.

Figure 4. Factors influencing DSB processing

(A) Spatial correlation of SSDS coverage (data from ref. [21]) with methylated nucleosomes. Each panel shows the mean of locally normalized profiles across the 20% of hotspots with the most

asymmetric H3K4me3 patterns (left > right in top panel; right > left in bottom panel). The same local normalization factor was applied to Watson and Crick SSDS reads for each hotspot so that signal strengths on Watson and Crick were comparable.

(B) Ratio of sum of locally normalized SSDS coverage in 3,001-bp windows around the centers of the subsets of hotspots from panel A.

(C) SSDS coverage was concordant with H3K4me3 signal, albeit weakly. Each panel shows the mean of locally normalized profiles on Watson (top) and Crick (bottom) strands in the hotspots subsetted by H3K4me3 asymmetry. Local normalization was applied separately to Watson and Crick reads so that signal strengths were comparable between the subsets.

(D) SPO11 oligos and SSDS coverage display similar density patterns in centromere-proximal regions (left) but not in centromere-distal regions (right). Points are densities of SPO11 oligos (reads per million (RPM) per Mb) and SSDS tag counts (tags per million (TPM) per Mb), within coordinates defined by SSDS hotspots, in 1-Mb windows, averaged across all 19 autosomes. In the right panel, one SPO11-oligo outlier is not shown.

(E) Ratios of SSDS tag counts to SPO11-oligo read counts in SSDS hotspots differ between by autosomal sub-chromosomal domain. "Sub-telomeric" is defined as the centromere-distal 5 Mb of each autosome. "Interstitial" is all other autosomal regions. One tag count and one SPO11-oligo read count were added to each hotspot. Boxplot outliers are not shown. P value is from Wilcoxon rank sum test.

Figure 5. Crossover asymmetry in hotspots with polymorphisms in putative PRDM9 binding sites

(A) Model for crossover asymmetry. A sequence polymorphism in a PRDM9 binding site may affect relative DSB frequencies on the two hotspot alleles, manifested as asymmetry in the locations of crossover breakpoints. If DSBs are preferentially formed on the B6 chromosome of a B6×DBA hybrid mouse, crossover breakpoints will tend to lie to the left of the hotspot center when recombinant products are assayed after PCR amplification in the B6-to-DBA orientation, and will tend to lie to the right when amplified in the DBA-to-B6 orientation.

(B–C) Examples of crossover hotspots with (B) or without (C) crossover asymmetry. (i) B6 (top) and DBA (bottom) sequences of putative 36-bp binding sites for PRDM9[B6] at hotspot centers. The nucleotides shaded in yellow in *HS59.5* highlight a polymorphism between the B6 and DBA haplotypes. In *HS61.1*, the PRDM9 motif shown is on the Crick strand. (ii) SPO11-oligo maps. Red lines indicate SPO11-oligo hotspots. (iii) Crossover breakpoints (densities expressed as centiMorgans (cM) per Mb) mapped by allele-specific PCR on sperm DNA in the B6-to-DBA (top) and DBA-to-B6 (bottom) orientation.[86,87] Ticks represent tested polymorphisms. (iv) Cumulative distributions of crossover breakpoints with fitted Gaussian curves. The number indicates the distance between the two curves at the midpoint for each cumulative plot. Vertical dashed lines indicate hotspot centers. For hotspot *HS61.1*, zero values in both orientations at outlier position -1130 bp are not shown.

(D) Crossover asymmetry is associated with presence of polymorphisms in putative PRDM9 binding sites at hotspots (**Table S3**). Crossover asymmetry was defined for each locus as the absolute difference between the midpoints of cumulative crossover breakpoint maps in the two orientations.
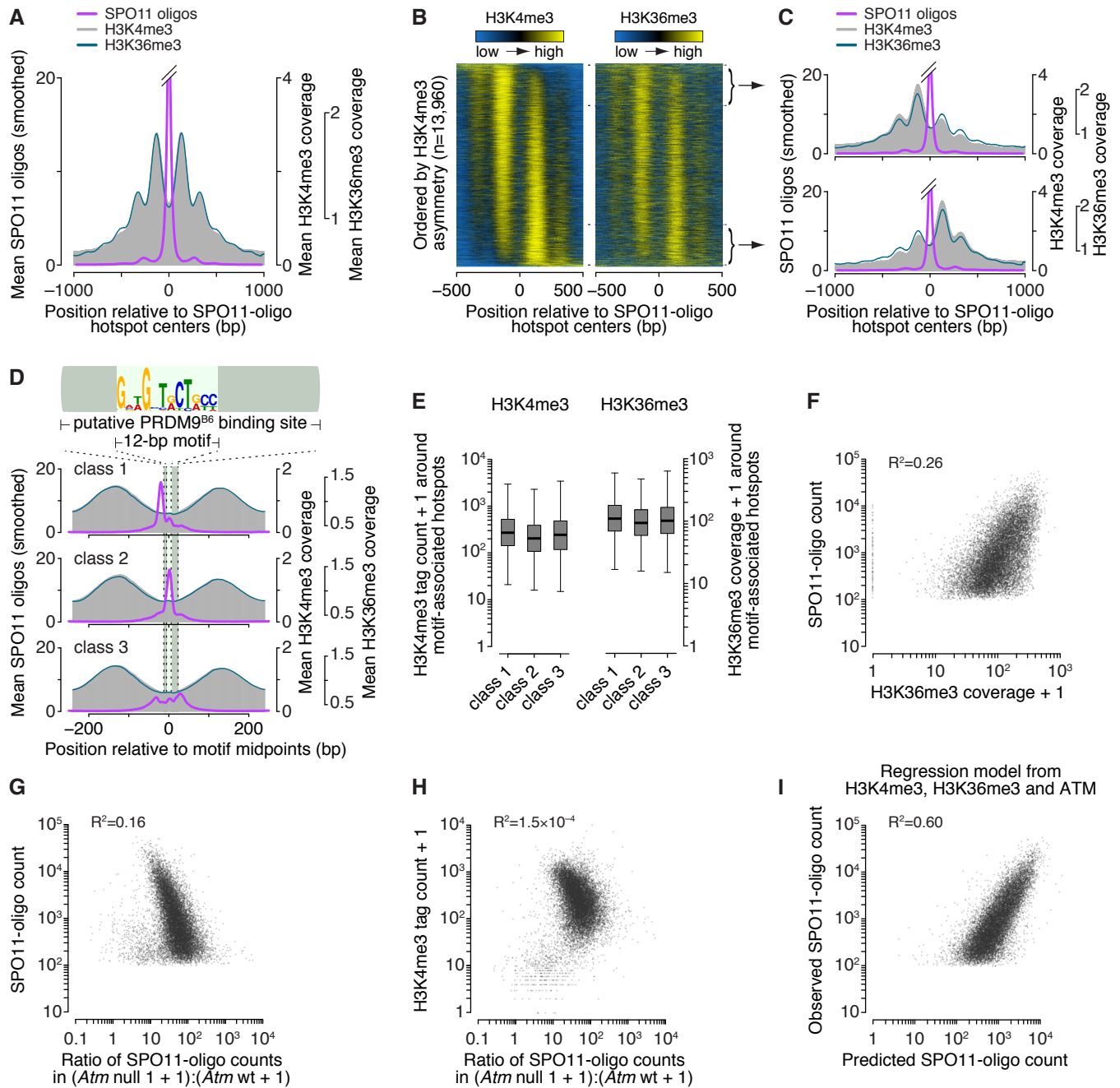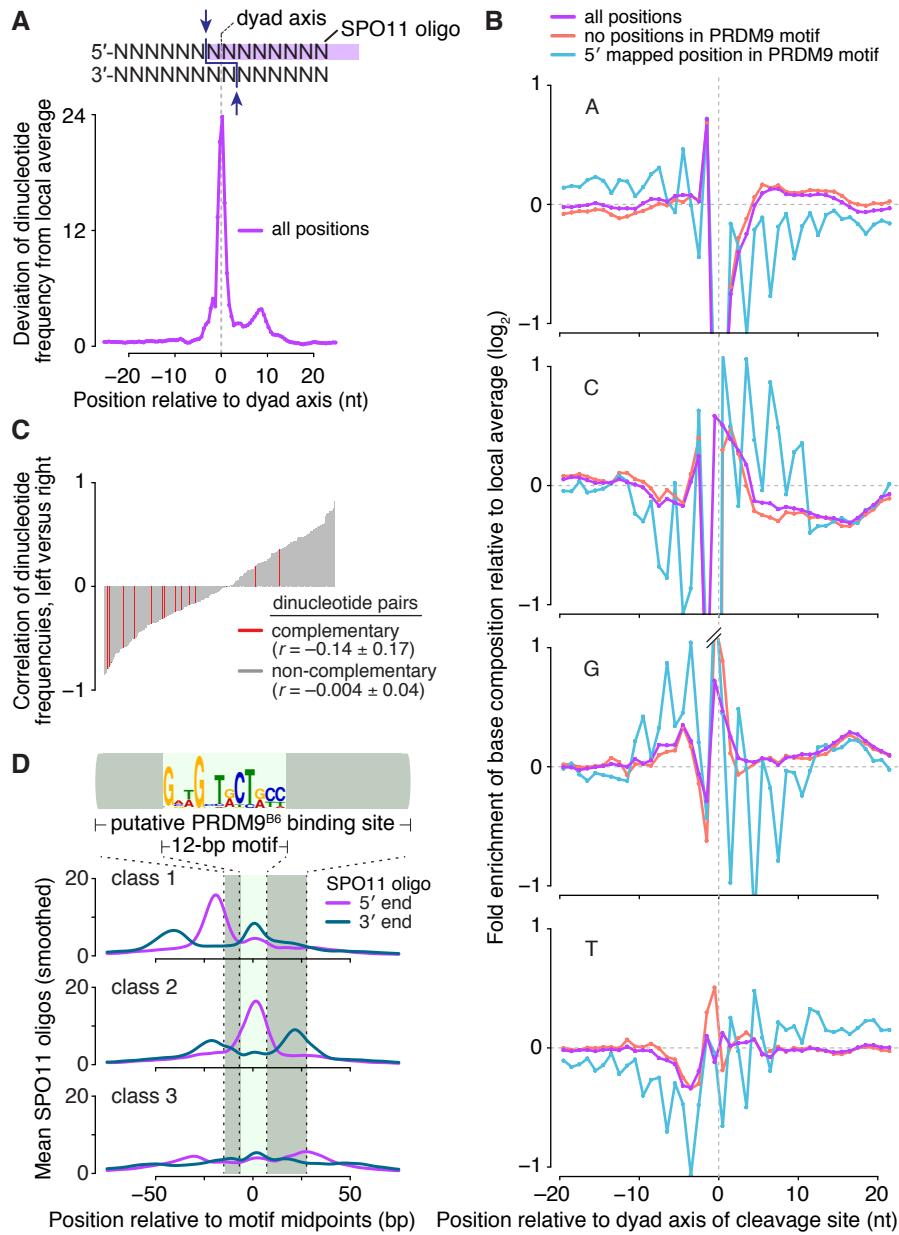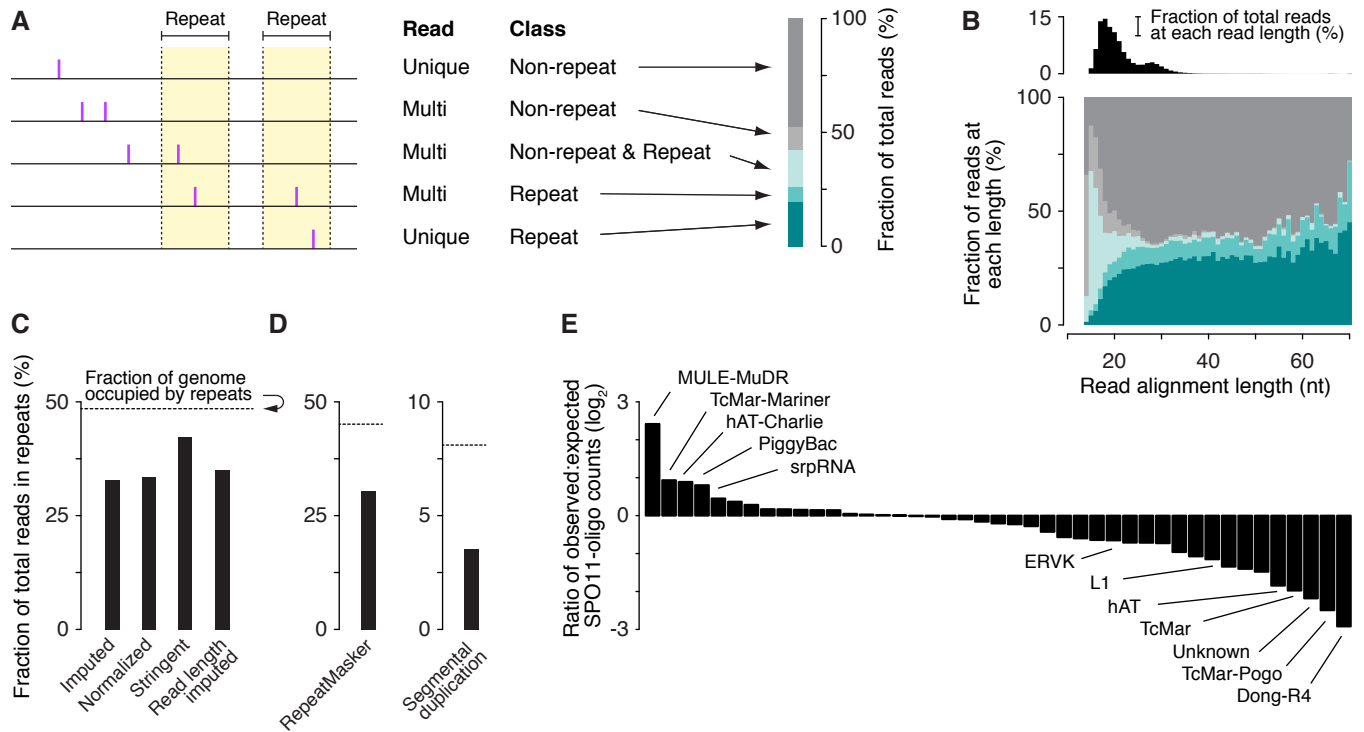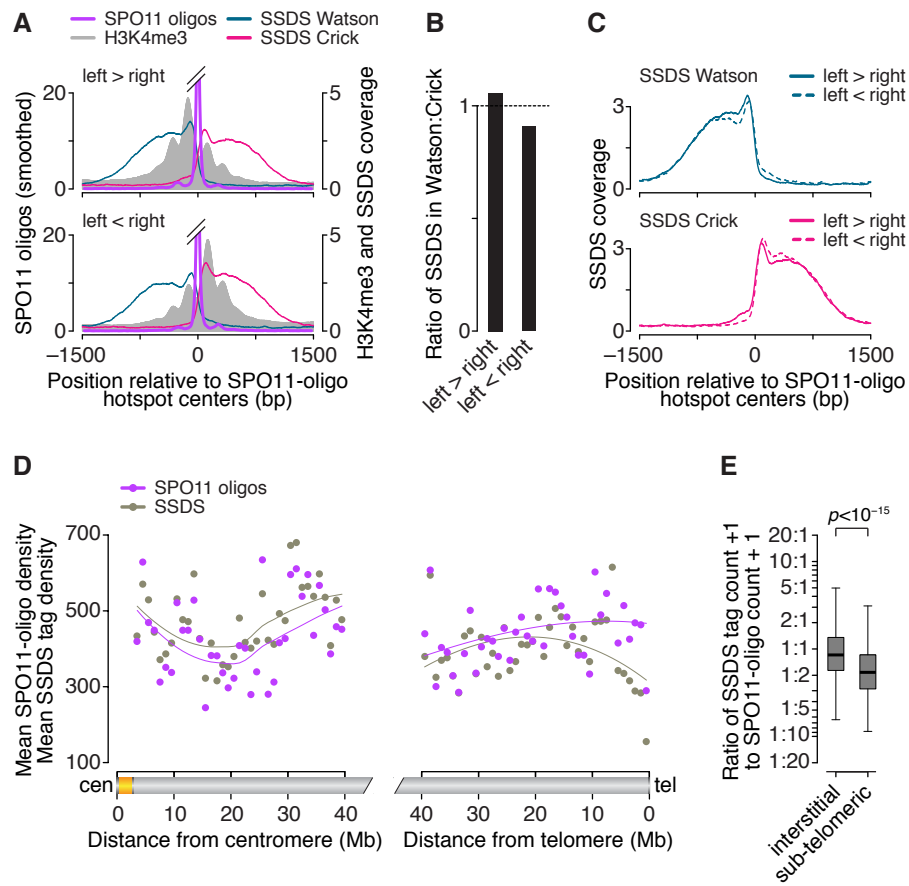
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Yamada S, *et al*. 2017
Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice

Supplemental Table 1. Multiple regression results

Multiple linear regression was performed with log-fold change in SPO11-oligo density around hotspot centers as the dependent variable and the indicated features as the independent variables. SPO11-oligo counts in B6, *Atm* null and *Atm* wt, H3K4me3 tag counts, and H3K36me3 coverage were summed in 1,001-bp windows around B6 hotspot centers. One count was added to each hotspot in *Atm* null and *Atm* wt SPO11-oligo counts, H3K4me3 tag counts, and H3K36me3 coverage counts. The regression analyses were performed on the log-transformed data. Estimates of the standardized regression coefficients (β) are shown, along with t statistics and P–values based on the standardized coefficients. For the "H3me only" model (H3K4me3 and H3K36me3), multiple R^2 = 0.4, adjusted R^2 = 0.4, F = 5,430 on 2 and 13,957 degrees of freedom, and P < 2.2×10^{-16}. For the "H3me + ATM" model, multiple R^2 = 0.6, adjusted R^2 = 0.6, F = 6,748 on 3 and 13,956 degrees of freedom, and P < 2.2×10^{-16}. Both models were different from the "H3K4me3 only" model (ANOVA P–value < 2.2×10^{-16} for each of the two models). Their AIC (akaike information criterion, a measure of a quality of models) values are 47,331 and 42,885, respectively, indicating that the ATM factor improved the fit of the model.

Model	Variable	β (standard error)	t statistic	P–value
H3me only	H3K4me3	0.49 (0.007)	66	< 2.2×10^{-16}
	H3K36me3	0.26 (0.007)	36	< 2.2×10^{-16}
	Intercept	0 (0.006)	0	1
H3me + ATM	H3K4me3	0.53 (0.006)	84	< 2.2×10^{-16}
	H3K36me3	0.20 (0.006)	31	< 2.2×10^{-16}
	ATM[1]	-0.40 (0.005)	-73	< 2.2×10^{-16}
	Intercept	0 (0.005)	0	1

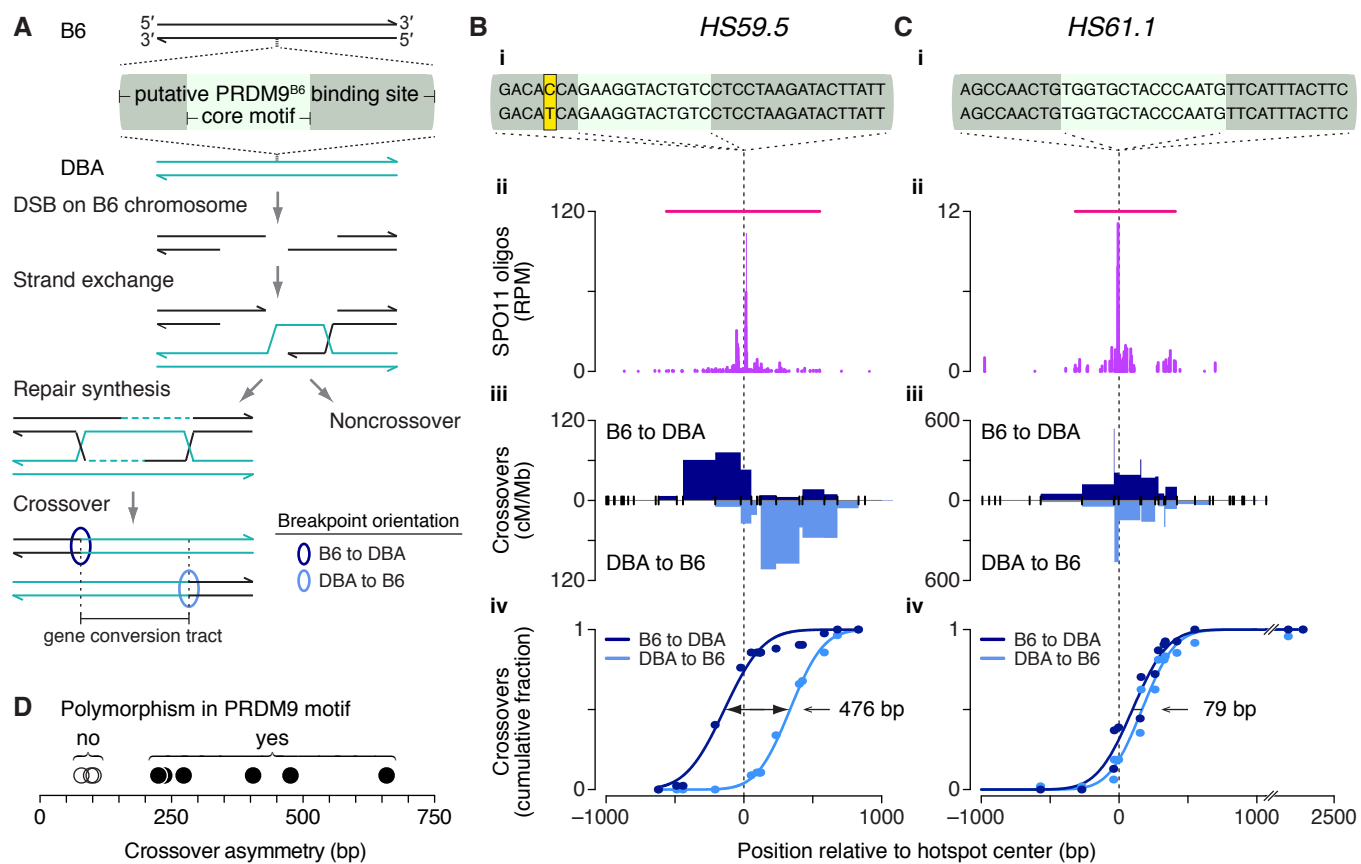[1] The variable ATM represents the ratio of SPO11-oligo counts in *Atm* null to *Atm* wt.

Yamada S, *et al*. 2017
Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice

Supplemental Table 2. DSB frequencies in interspersed repeats
For each repeat annotated in RepeatMasker, SPO11-oligo counts per base pair were summed and their enrichment was calculated relative to expected values (see Materials and methods). ERVK and L1 repeat family elements whose meiotic transcription is suppressed by DNMT3L[63] are indicated.

Repeat family – repeat name	Ratio of observed:expected SPO11-oligo counts	Fraction of genome occupied by repeat family (%)
MULE-MuDR	5.3	0.0042
TcMar-Mariner	1.9	0.0086
hAT-Charlie	1.9	0.71
PiggyBac	1.7	0.0020
srpRNA	1.4	0.0018
tRNA	1.3	0.018
ERVL	1.2	1.2
B4	1.1	2.3
ERV1	1.1	1.0
MIR	1.1	0.56
RTE-X	1.1	0.012
hAT-Tip100	1.1	0.069
Satellite	1.0	0.17
ERVL-MaLR	1.0	4.7
RNA	1.0	0.0046
TcMar-Tigger	1.0	0.26
ID	1.0	0.17
Deu	1.0	0.0063
L2	0.94	0.42
Other	0.93	0.29
Alu	0.89	2.6
Y-chromosome	0.87	0.0083
snRNA	0.85	0.0095
CR1	0.82	0.069
TcMar-Tc2	0.74	0.014
Gypsy	0.67	0.012
B2	0.66	2.3
hAT-Blackjack	0.64	0.032
ERVK	0.64	4.8
– IAPEz-int	0.026	0.45
– IAPLTR1_Mm	0.014	0.021
– MMERVK10C-int	0.16	0.11
– RLTR10C	0.21	0.029
LTR	0.61	0.0056
scRNA	0.61	0.025
DNA	0.60	0.0050
MuDR	0.51	0.0014
Low_complexity	0.47	0.77
L1	0.45	20
– L1Md_A	0.040	1.2
– L1Md_Gf	0.048	0.069
– L1Md_T	0.040	1.8
RTE-BovB	0.39	0.00085
Helitron	0.38	0.0026
Simple_repeat	0.36	2.2
hAT	0.28	0.0091
TcMar	0.25	0.00032
Unknown	0.22	0.065
TcMar-Pogo	0.18	0.00010
Dong-R4	0.13	0.00093

Supplemental Table 3. Putative PRDM9 binding sites in crossover hotspots

Using MAST[93] to query 15 known crossover hotspots[82-88] with previously published primary and secondary core PRDM9 motif sequences,[26] we identified putative 36-bp PRDM9 binding sites within 250 bp of hotspot centers. Our method detected a match to the primary or secondary PRDM9 motif in 13 of the 15 crossover hotspots.

Hotspot	F_1 Hybrid	Center[1]	Skew (bp)[2]	Core motif[4]	Putative 36-bp binding site[3] Start	End	Strand		Sequence[5]
Chromosome 1									
central	B6×A/J	78590926	97	primary	78590944	78590979	+	B6 A/J	GGAATAAGAGGCTACTGTCCCCTTTATTTTGCCAAA GGAATAAGAGGCTACTGTCCCCTTTATTTTGCCAAA
M1	B6×C3H	143877054	n.a.	primary	143877041	143877076	+	B6 C3H	AGGTAGGGGTGCTACTACCACCATGTCAGAGATCCT AGGTAGGGGTGCTACTACCACCATGTCAGAGATCCT
A3	B6×DBA	160025733	273	secondary	160025729	160025764	+	B6 DBA	ACCATGTCCAGAAGCTGCTACCATG--CGCCACACCCA ACCATGTCCAGAAGCTGCTACCATGTTCGCCACACCCA
				primary	160025732	160025767	+	B6 DBA	ATGTCCAGAAGCTGCTACCATG--CGCCACACCCACAG ATGTCCAGAAGCTGCTACCATGTTCGCCACACCCACAG
distal	B6×A/J	185267182	102	secondary	185267135	185267170	+	B6 A/J	TTACCAAAAGGATGCTGTTACCCTCTTGTGTCTTCC TTACCAAAAGGATGCTGTTACCCTCTTGTGTCTTCC
				primary	185267138	185267173	+	B6 A/J	CCAAAAGGATGCTGTTACCCTCTTGTGTCTTCCAGA CCAAAAGGATGCTGTTACCCTCTTGTGTCTTCCAGA
Chromosome 19									
HS18.2	B6×DBA	18146826	236	primary	18146854	18146889	+	B6 DBA	ATATCTAGAAGATACTGCACTTTTAAAGGTTAGCTT ATATCCAGAAGATACTGCACTTTTAAAGGTTAGCTT
				secondary	18146558	18146593	+	B6 DBA	GAAATGGCTAGGTGCTACACTTTTGTACATTGAGTC GAAATGGCTAGGTGCTACACTTTTGTACATTGAGTC
HS22	B6×DBA	22961739	659	secondary	22961603	22961638	−	B6 DBA	GAATGCCTCTGCCGCTGTCCTCTGAGCACCAGGGGA GAATGCCTCTGCCTGCTGTCCTCTGAGCACCAGGGGA
				secondary	22961721	22961756	+	B6 DBA	ATGGCAGAAGGAAGCTACCTCCCTTCTCAGTGTTCT ATGGCAGAAGGAAGCTACCTCCCTTCTCAGTGTTCT
HS23.9	B6×DBA	23842536	405	primary	23842513	23842548	−	B6 DBA	ACAGCATGGTGGTACCACCTGATGGTCACAGGATCT ACAACATGGTGGTACCACCTGATGGTCACAGGATCT
				secondary	23842412	23842447	+	B6 DBA	ACATTCAAATGGTTCTGCATCCTTTTAAAAAATCAA ACATTCAAATGGTTCTGCATCCTTTTAAAAAATCAA
				secondary	23842516	23842551	−	B6 DBA	GTGACAGCATGGTGGTACCACCTGATGGTCACAGGA ATGACAACATGGTGGTACCACCTGATGGTCACAGGA

Supplemental Table 3 continued

Hotspot	F$_1$ Hybrid	Center[1]	Skew (bp)[2]	Putative 36-bp binding site[3]			Strand		Sequence[5]
				Core motif[4]	Start	End			
Chromosome 19									
HS37	B6×DBA	38678273	n.a.	secondary	38678232	38678267	–	B6	GCAACAGTTTGGTGCTGTCATTCTATTATTTATTGG
								DBA	GCAACAGTTTGGTGCTGTCATTCTATTATTTATTGG
				secondary	38678253	38678288	+	B6	GCACCAAACTGTTGCTGTCTTATCTATGTTTCCTTC
								DBA	GCACCAAACTGTTGCTGTCTTATTTATGTTTCCTTC
HS44.2	B6×DBA	44148993	n.a.	primary	44148971	44149006	–	B6	AGGTCCAGAAGTTGCTGTCAGCCTATAGATTACCTT
								DBA	AGGTCCAGAAGTTGCTGTCAGCCTATAGATTACCTT
HS59.4	B6×DBA	59349707	n.a.	primary	59349687	59349722	–	B6	CAGAAGAGAGGGTACTATCAATGAGTTGTTGCCAGG
								DBA	CAGAAGAGAGGGTACTATCAATGAGTTGTTGCCAGG
HS59.5	B6×DBA	59442973	476	primary	59442946	59442981	+	B6	GACACCAGAAGGTACTGTCCTCCTAAGATACTTATT
								DBA	GACATCAGAAGGTACTGTCCTCCTAAGATACTTATT
HS61.1	B6×DBA	61012565	79	secondary	61012580	61012615	–	B6	AGCCAACTGTGGTGCTACCCAATGTTCATTTACTTC
								DBA	AGCCAACTGTGGTGCTACCCAATGTTCATTTACTTC
HS61.2	B6×DBA	61153474	225	primary	61153492	61153527	+	B6	ATTGTATGATGCTGCTGCTAACTTCATAGTTTCTTC
								DBA	ATTGCATGATGCTGCTGCTGACTTCATAGTTTCTTC

[1] Center is the position of the matched SPO11-oligo hotspot center, except for *HS18.2*, where Center is the middle of the crossover breakpoint distribution.

[2] Skew is based on previously published crossover data remapped to B6. n.a., not applicable (see Materials and methods)

[3] Putative 36-bp binding site was determined by extending the identified 12-bp primary core motif or 15-bp secondary core motif by 7 bp to the left and 17 bp to the right or by 9 bp to the left and 12 bp to the right, respectively.

[4] Primary and secondary motif sequence logos were previously published.[26]

[5] In Sequence column, core motifs are highlighted in green; polymorphisms are highlighted in yellow.