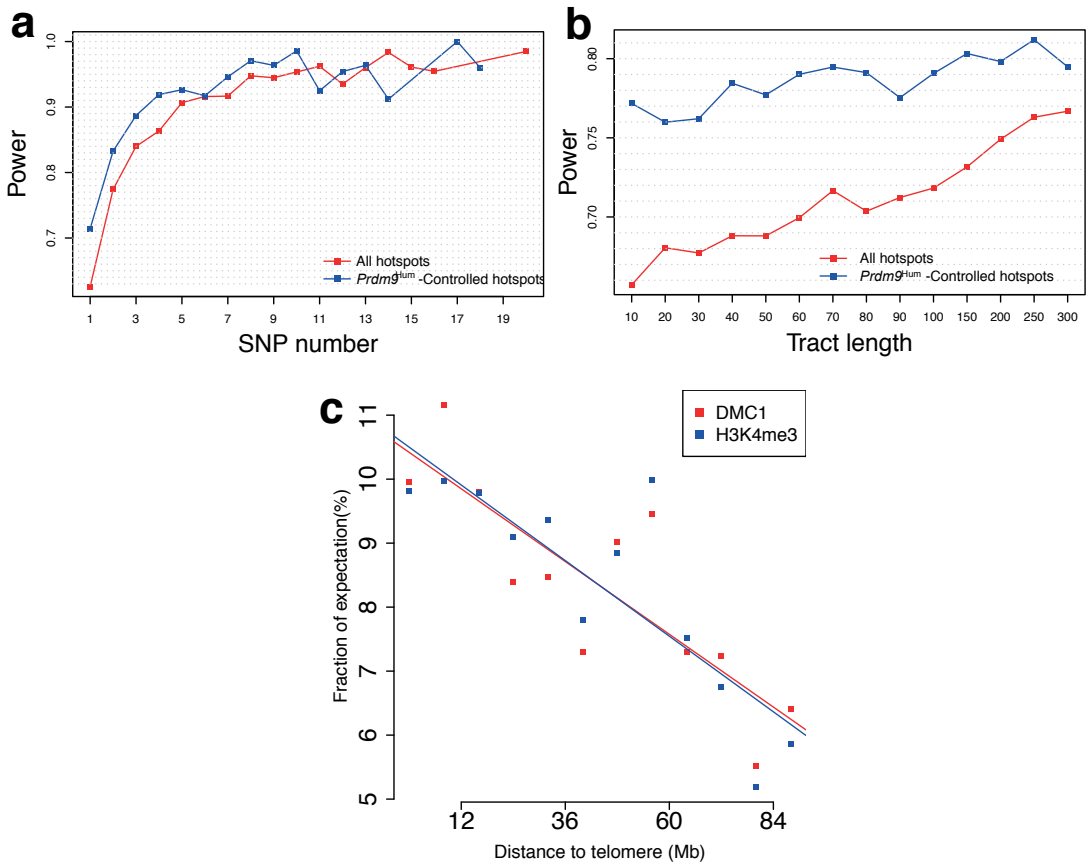
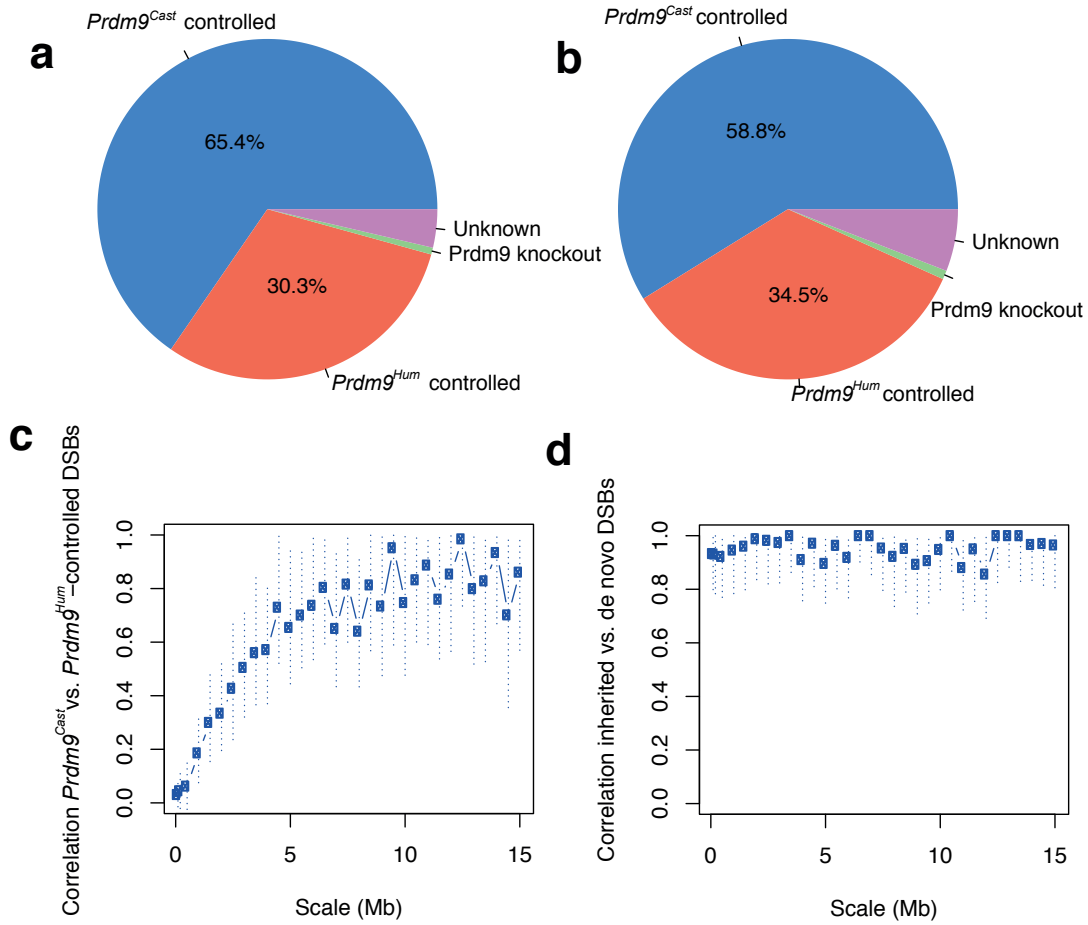
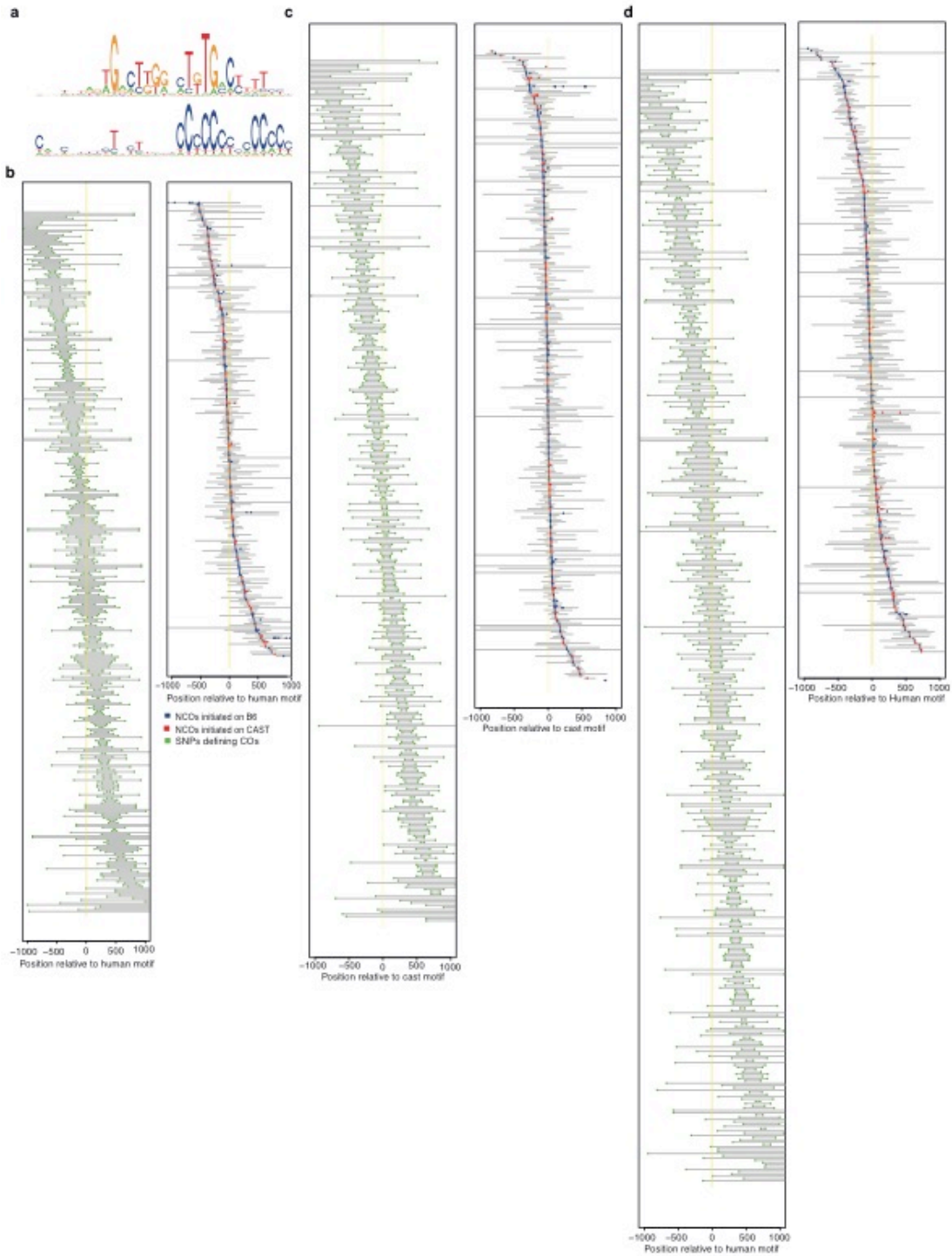


Extended Data Fig. 1| Estimation of power to detect NCO events. a, Power increases with a greater number of converted sites per NCO tract in simulations. Red line: power to detect events in F2 mice, Blue line: power to detect human-controlled events in F5 mice. **b**, Power increases with greater mean tract length in simulations. **c**, Fraction of total DMC1/H3K4me3 signal coming from bins of genomic regions at different distances to the telomere (x-axis). Both DMC1 and H3K4me3 signals are enriched in telomeric regions. Full resolution available at: https://figshare.com/s/bf883f746fd676f1edb4

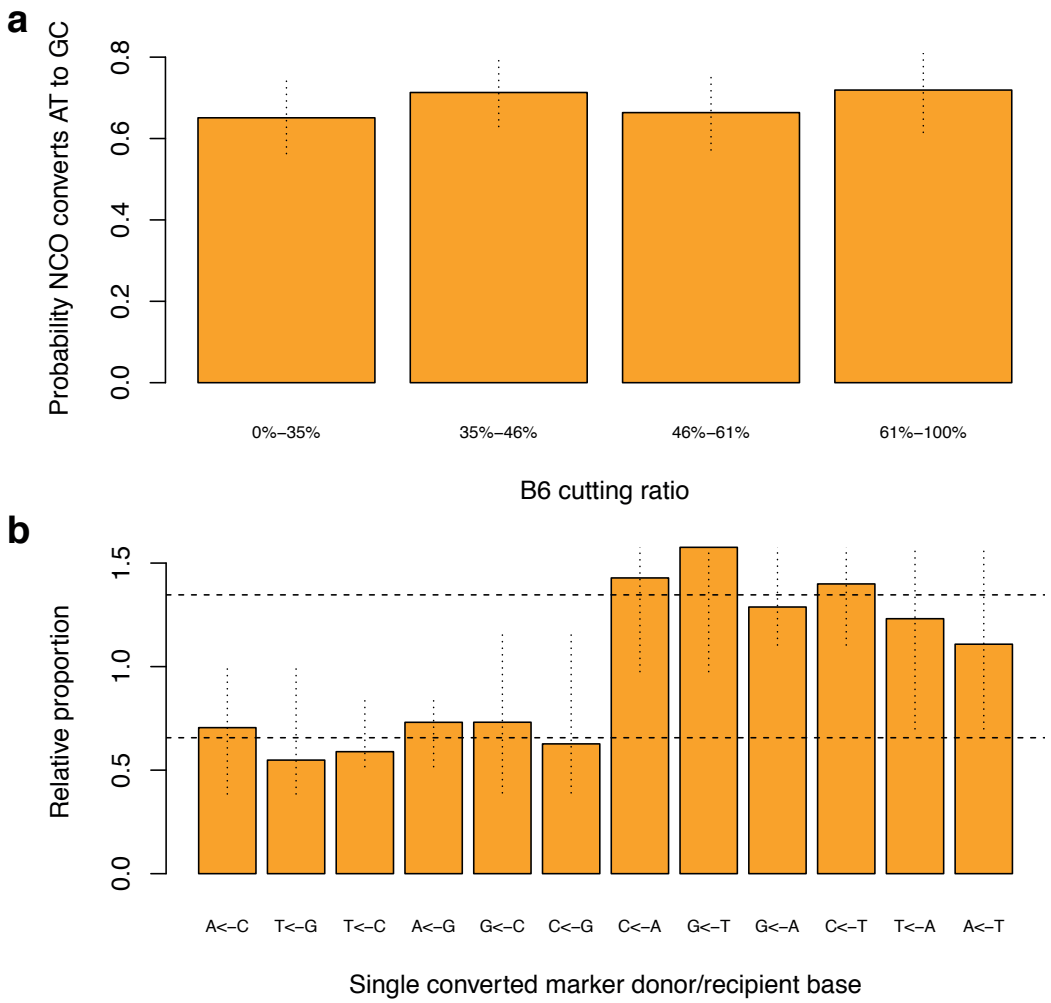


Extended Data Fig. 2| Allelic dominance and broad-scale patterns. DMC1 (**a**) and H3K4me3 (**b**) signals show dominance of *Prdm9*^{Cast}–controlled over *Prdm9*^{Hum}–controlled hotspots. **c**, Estimated underlying correlation between *Prdm9*^{Cast}-controlled recombination events and *Prdm9*^{Hum}-controlled recombination events at different scales. Details are as for Fig. 2g. Correlation is low at finer scales and higher at broader scales. **d**, Correlation between inherited recombination events and *de novo* recombination events at different scales, as for **c**. Encouragingly, estimated correlation is ~100% at all scales. Full resolution available at: https://figshare.com/s/bf883f746fd676f1edb4

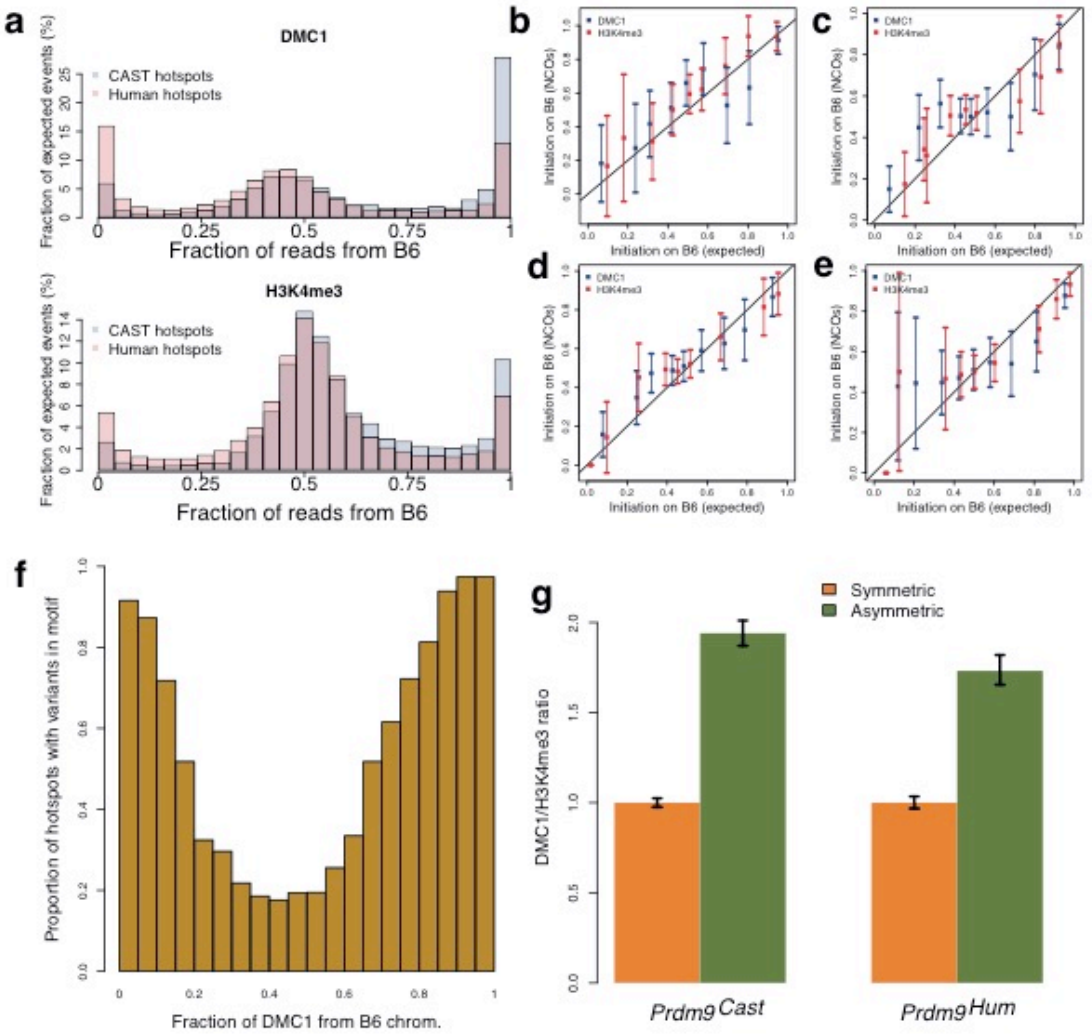


Extended Data Fig. 3| NCOs and COs distribute around PRDM9 binding motifs. **a**, We identified distinct motifs, and their locations, within 97% of hotspots controlled by *Prdm9*^{Cast} (top) and 74% of *Prdm9*^{Hum} hotspots (bottom). **b**, *De novo* COs (left) and NCOs (right) detected from the F5 humanized colony distribute around the PRDM9^{Hum} binding motif. **c**, Inherited COs (left) and NCOs (right) detected from the

F5 humanized colony that are controlled by *Prdm9*^{Cast} distribute around the PRDM9^{Cast} binding motif. **d**, Inherited COs (left) and NCOs (right) detected from the F5 humanized colony that are controlled by *Prdm9*^{Hum} distribute around the PRDM9^{Hum} binding motif. Full resolution available at: https://figshare.com/s/bf883f746fd676f1edb4

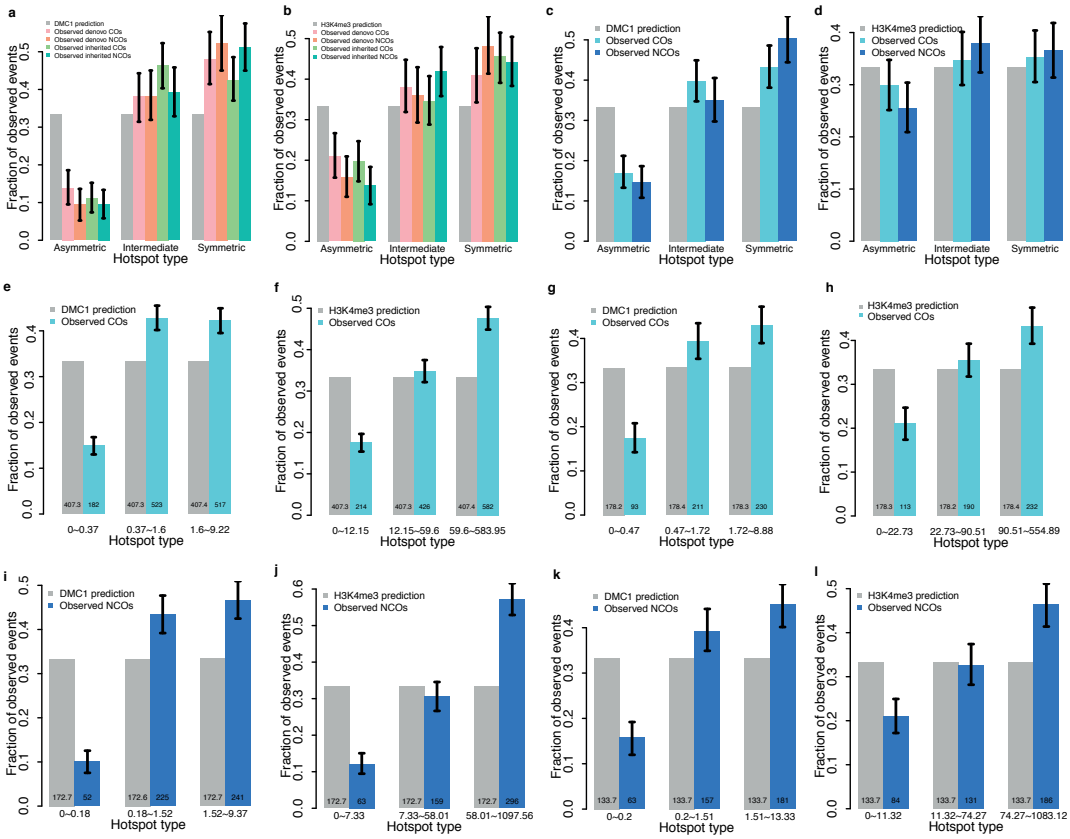


Extended Data Fig. 4| GC-bias occurs independently of hotspot symmetry. a, Hotspots with different fractions of reads coming from B6 (x-axis) show similar GC-bias (y-axis). **b**, For each of the 12 possible combinations of NCO donor/recipient alleles (x-axis; e.g. A<-C converts recipient C to donor A), we plot the proportion of observed single-SNP NCOs of that type, relative to the corresponding proportion for all SNPs within observed multiple-SNP NCOs, which lack GC-bias. Vertical lines: 95% CI's after pooling strand-equivalent pairs (e.g. A<-C and T<-G). Horizontal dotted lines: mean relative proportions for NCO events whose recipient types are G/C or A/T respectively, showing under-representation of events whose recipients are G/C could explain observed patterns. Full resolution available at: https://figshare.com/s/bf883f746fd676f1edb4



Extended Data Fig. 5| Asymmetric hotspot properties. a, The proportions of DMC1/H3K3me3 reads coming from hotspots binned according to the bias in chromosome-informative reads towards the B6 chromosome (100%: all reads from B6). *Prdm9*^{Hum}-controlled and *Prdm9*^{Cast}-controlled hotspots are shown separately. *Prdm9*^{Cast} shows bias towards B6. **b-e**, NCOs were binned according to their predicted (H3K4me3 or DMC1) B6 cutting ratio, and expected (x-axis) versus observed (y-axis) fraction of events initiating on B6 are plotted. Vertical lines: 95% confidence intervals. Plots show events from (**b**) F2, (**c**) F5 *de novo* events, (**d**) F5 inherited events controlled by *Prdm9*^{Hum} and (**e**) F5 inherited events controlled by *Prdm9*^{Cast}. Expected versus observed values are highly correlated. **f**, For hotspots binned according to the fraction of DMC1 reads from the B6 chromosome (x-axis),

the fraction containing SNP/indel variants within the PRDM9 binding motif (y-axis). **g**, The genome-wide autosomal ratio of mean DMC1 heat to mean H3K4me3 enrichment for asymmetric hotspots (fraction of reads from B6 chromosome is either larger than 0.95 or smaller than 0.05) relative to symmetric hotspots (fraction of reads from B6 chromosome is larger than 0.4 and smaller than 0.6) in both *Prdm9*^{Hum} and *Prdm9*^{Cast} controlled hotspots. Error bars: 95% bootstrap confidence intervals for the ratio of means. Asymmetric hotspots show elevated DMC1 relative to H3K4me3 signal. Full resolution available at: https://figshare.com/s/bf883f746fd676f1edb4



Extended Data Fig. 6| Recombination events avoid asymmetric hotspots. a, *Prdm9*^{Hum}-controlled hotspots are binned by their symmetry (*21*) into "asymmetric", "intermediate" and "symmetric" hotspots, so each bin has the same expected number of events according to DMC1 heat. Grey bars: expected event fraction in bins from DMC1 heat. Coloured bars: observed number of (resampled) events in each bin, in labelled categories *de novo* COs from F5, *de novo* NCOs, inherited COs controlled by *Prdm9*^{Hum} and inherited NCOs controlled by *Prdm9*^{Hum}. Vertical lines: 95% bootstrapped confidence intervals. **b**, As **a** but for recombination events controlled by *Prdm9*^{Cast}, so there are no *de novo* F5 events. **c**, As **a** except the binning and predicted events are calculated according to H3K4me3 heat. **d**, As **b** except the predicted events are calculated according to H3K4me3 heat. **e-h**, As **a** and **c** except showing CO events (no rejection sampling), now binning hotspots according to their average heats on the homologous chromosome, and for labelled alleles and measures of hotspot heat

(DMC1/H3K4me3). **i-l**, As **e-h**, except for (all) NCO events, and now binning hotspots according to their heats on the homologous chromosome (supplementary information), because the initiating chromosome is identifiable for NCOs. Across plots, note strong under-representation in asymmetric hotspots ($p<0.05$). Full resolution available at: https://figshare.com/s/bf883f746fd676f1edb4

Extended Data Table 1| Filters to identify true NCOs.

Filter	Description	Applied to
1	Remove sites that have read depth <20 in MGP version 4 data in B6 or CAST	All sites
2	Remove sites that have >1 alternative reads in B6 and sites that have >1 reference reads in CAST	All sites
3	Remove sites called heterozygous in any of the 28 strains of mice in the MGP data	All sites
4	Remove sites shared by >2 F2 animals	All sites
5	Remove sites such that within 500bp there are >28 reads whose mate pairs map (insert size) >1kb away	All sites
6	Removed sites covered by less than 10 good reads*	All sites
7	Remove sites that are called different by Platypus version 0.7.9.1	All sites
8	Remove sites that have read depth >95% quantile for the sample	Heterozygous
9	Remove sites that have <3 good reference reads or <3 good alternative reads	Heterozygous
10	Remove sites that show allelic imbalance (>70% of reads agree with non-converted background)	Heterozygous
11	Remove sites if they have genotype quality (GQ) <30	Heterozygous
12	Remove sites that have no good reads from the alternative allele	Homozygous
13	Remove sites where the nearby non-converted sites overlapping read pairs containing potential converted sites show allelic imbalance (potential allelic "dropout")	Homozygous
14	After applying the above filters, remove sites if there are >2 sites filtered within 500bp, and the fraction of removed sites in this region (<500 bp) is >50%. Iterate this process until we don't remove further sites ("guilt by association")	All sites
15	After step 14, recover potential converted sites <500bp from conversion events passing filters (avoid removal of genuine long or complex events by accidentally failing filters). Iterate until we don't recover additional sites.	All sites

* "Good" reads are defined as reads whose mate pair is not mapped to other chromosomes, and with insert size ≤1000bp. For reads containing the converted site, this site is >5bp from any indel and >10bp from the end of the read. Most properly mapped read pairs comfortably satisfy the first condition; we found empirically that alignment artefacts for reads failing the second condition led to many miscalled NCO events.

Extended Data Table 2| Summary of NCO/CO events overlapping DMC1 and/or H3K4me3 peaks.

Datasets	Total events	Overlap DMC1	Overlap H3K4me3	Overlap either	Overlap either (%)
F2 COs	295	272	271	282	95.6
F5 de-novo COs	821	646	690	728	88.7
F5 inherited COs	1384	1164	1196	1264	91.3
All COs	2500	2082	2157	2274	91.0
F2 NCOs	183	147	144	154	84.2
F5 de-novo NCOs	510	355	375	402	78.8
F5 inherited NCOs	882	730	713	771	87.4
All NCOs	1575	1232	1232	1327	84.3

Extended Data Table 3| GC-bias in NCO events.

NCOs		AT to GC	GC to AT	Probability of AT to GC	P value
Human controlled	All	554	366	0.60	6.2e-10
	F5 de-novo	261	160	0.62	1.0e-07
	F5 inherited	270	195	0.58	0.0006
CAST controlled	All	298	167	0.64	1.3e-09
	F5	218	111	0.66	3.7e-09
	F2	80	56	0.59	0.048
F5 de-novo	Paternal	71	53	0.57	0.1265
	Maternal	56	42	0.57	0.1888
	Either	127	95	0.57	0.0372

All events in this table overlap DMC1 hotspots. P-values are calculated via binomial two-sided tests.

Extended Data Table 4| GC-bias for single-SNP versus multi-marker NCO events. See separate Excel sheet. As Extended Data Table 2, except NCO events are stratified according to whether they contain a single marker, or overlap multiple markers. Results show only human-controlled NCO events in F5 mice. In addition to Extended Data Table 3 categories, we stratify NCO events according to whether they occur in asymmetric vs. symmetric hotspots, strong or weak hotspots, or nearby versus distally from an identified PRDM9 binding motif. All categories show similar results, and GC-bias specific to single-SNP NCO events.