High-Resolution Sex-Specific Linkage Maps of the Mouse Reveal Polarized Distribution of Crossovers in Male Germline

Eric Yi Liu,*[,1] Andrew P. Morgan,[†,1] Elissa J. Chesler,[‡] Wei Wang,[§] Gary A. Churchill,[‡] and Fernando Pardo-Manuel de Villena[†,2]

*Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27599-3175, [†]Department of Genetics, Carolina Center for Genome Sciences and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599-7264, [‡]The Jackson Laboratory, Bar Harbor, Maine 04609, and [§]Department of Computer Science, University of California, Los Angeles, California 90095-1596

ABSTRACT Since the publication of the first comprehensive linkage map for the laboratory mouse, the architecture of recombination as a basic biological process has become amenable to investigation in mammalian model organisms. Here we take advantage of high-density genotyping and the unique pedigree structure of the incipient Collaborative Cross to investigate the roles of sex and genetic background in mammalian recombination. Our results confirm the observation that map length is longer when measured through female meiosis than through male meiosis, but we find that this difference is modified by genotype at loci on both the X chromosome and the autosomes. In addition, we report a striking concentration of crossovers in the distal ends of autosomes in male meiosis that is absent in female meiosis. The presence of this pattern in both single- and double-recombinant chromosomes, combined with the absence of a corresponding asymmetry in the distribution of double-strand breaks, indicates a regulated sequence of events specific to male meiosis that is anchored by chromosome ends. This pattern is consistent with the timing of chromosome pairing and evolutionary constraints on male recombination. Finally, we identify large regions of reduced crossover frequency that together encompass 5% of the genome. Many of these "cold regions" are enriched for segmental duplications, suggesting an inverse local correlation between recombination rate and mutation rate for large copy number variants.

RECOMBINATION is a basic biological process that is shared among most sexually reproducing organisms (Morgan 1911; Gerton and Hawley 2005). It plays a key role in genome stability by ensuring the fidelity of chromosome segregation during meiosis (Sears *et al.* 1992; Hassold and Hunt 2001) and contributes to other processes such as DNA repair (Howard-Flanders and Theriot 1966; Niedzwiedz *et al.* 2004; Krejci *et al.* 2012). At the population level, recombination is an important generator of genetic diversity (Feldman *et al.* 1996; Otto and Lenormand 2002). Abnormal recombination is associated with increased aneuploidy and decreased fitness of offspring and has been associated with several human diseases (Warren *et al.* 1987; Hassold and Hunt 2001; Kong *et al.* 2004). Recombination can be exploited experimentally to map loci associated with biological traits: indeed, the construction of linkage maps is among the oldest activities in genetics (Sturtevant 1913). Finally, the recombination machinery can be co-opted for genetic engineering of many organisms (Smithies *et al.* 1985; Court *et al.* 2002).

In mammals, many factors—including sex, taxon, and genetics—are known to affect the global, as opposed to local, rate of recombination. The total number of events per meiosis and the recombination rate per unit sequence length vary widely among mammals but are strongly correlated with the fundamental number (number of chromosome arms) present in the karyotype of each species (Pardo-Manuel De Villena and Sapienza 2001). Although the molecular process of recombination can result in either a noncrossover

doi: 10.1534/genetics.114.161653
Manuscript received January 16, 2014; accepted for publication February 20, 2014; published Early Online February 27, 2014.
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161653/-/DC1.
[1]These authors contributed equally to this work.
[2]Corresponding author: Department of Genetics, University of North Carolina, 120 Mason Farm Rd., CB 7264, Chapel Hill, NC 27599-7264. E-mail: fernando@med.unc.edu

or a crossover event, to date the study of recombination in mammals has been limited almost exclusively to crossovers, which are more readily detected. Previous studies have shown that, as a general rule, the number of crossovers observed in autosomes is higher in female meiosis than in male meiosis and thus the linkage map is longer in the gametes of females (Dunn and Bennett 1967; Broman *et al.* 1998; Kong *et al.* 2002; Cox *et al.* 2009). These same studies demonstrated that the genomic distribution of crossovers between female and male meiosis is significantly different: uniform in females, but subtelomerically enhanced and pericentromerically suppressed in males. Crossover interference (Petkov *et al.* 2007) and sex-specific patterns of hotspot usage (Paigen *et al.* 2008; Kong *et al.* 2010) have been advanced as candidate explanations for these phenomena. Recently, based on evidence from flowering plants, a population genetics basis for the evolution of sex differences in recombination rates—differences in gametic selection between males and females—has been proposed (Lenormand and Dutheil 2005). Despite these advances the presence and causes of sex differences in the overall rate and spatial distribution of recombination remain the object of study and controversy.

Recent studies indicate that recombination rate also varies between closely related species and subspecies and that alleles responsible for these effects may in fact be segregating within species (Dumont *et al.* 2009; Murdoch *et al.* 2010; Dumont and Payseur 2011; Auton *et al.* 2012). In particular, crossover frequency in male mice is known to vary across different inbred strains (Koehler *et al.* 2002) and these differences have been exploited to map genetic loci affecting recombination rates in F_2 intercrosses (Murdoch *et al.* 2010; Dumont and Payseur 2011). Finally, mutations at several genes are known to lead to pathological changes in recombination (Niedzwiedz *et al.* 2004; Liebe *et al.* 2006).

Traditionally recombination has been studied in large pedigrees, using small numbers of informative markers. The first comprehensive linkage map of the laboratory mouse was developed in 1992 by genotyping hundreds of microsatellite markers in an interspecific backcross (Dietrich *et al.* 1992), work that was crucial to the success of the Mouse Genome Project (Waterston *et al.* 2002). Since then several linkage maps have been reported that have used larger experimental populations and taken advantage of denser single-nucleotide polymorphism (SNP) genotyping arrays (Shifman *et al.* 2006; Cox *et al.* 2009). Increasingly fine-grained linkage maps are an important technical resource for the research community, enabling the development of more sophisticated genetic mapping methods and the explosive growth of complex-traits analysis in the laboratory mouse (Flint and Mackay 2009; Flint and Eskin 2012). In addition, such maps provide a window into fundamental processes of transmission genetics.

The development of very dense genotyping arrays and the concurrent genotyping of large numbers of unrelated human individuals opened the door to the development of high-resolution genetic maps based on linkage disequilibrium (LD) (McVean *et al.* 2004; Myers *et al.* 2005). This type of analysis has been recently extended to mouse (Brunschwig *et al.* 2012). However, in contrast to human studies that use unrelated individuals, work in mouse has been limited to laboratory strains for which the presence of population structure and introgression between clades makes the interpretation of the data challenging (Yang *et al.* 2011; Collaborative Cross Consortium 2012). New mouse populations, in particular the highly randomized and fully traceable Collaborative Cross population, aim to mitigate these concerns (Churchill *et al.* 2004).

Even more recently the power of next-generation sequencing technologies has been applied to identify and characterize thousands of recombination hotspots by mapping meiotic double-strand breaks (DSBs) in male germ cells (Smagulova *et al.* 2011; Brick *et al.* 2012). These new approaches have focused attention on hotspots, short discrete regions of the genome with recombination frequencies that are significantly above the genome-wide average. This led to the identification of *Prdm9* (and its human homolog *PRDM9*) as the key *trans*-regulator of meiotic hotspot activity in mice and humans, respectively (Myers *et al.* 2010; Parvanov *et al.* 2010). The fact that PRDM9 is a histone 3 lysine 4 trimethyltransferase that plays a role in epigenetic modification in the germline and that there are multiple alleles in both humans and mouse with different functional characteristics has resulted in the publication of a large body of literature in only a few years (Berg *et al.* 2010; Hinch *et al.* 2011; Brick *et al.* 2012; Hussin *et al.* 2013).

Despite the accomplishments of the LD and DSB approaches to study recombination, these methods are unsuited for the characterization of sex effects: LD represents a sex average and sequencing protocols to map DSBs have been implemented only for male germ cells. In addition, these approaches cannot determine the number of crossovers within a single chromosome in a given meiosis. As we demonstrate, these limitations obscure the important properties of mammalian meiosis.

Here we applied new high-density genotyping technology to the highly informative pedigree structure of the incipient Collaborative Cross (CC) to investigate the roles of sex and genetic background in mammalian recombination. The CC is a multiparental recombinant inbred population derived from eight founder inbred *Mus musculus* strains that collectively span 90% of the segregating variation in laboratory mice (Churchill *et al.* 2004; Roberts *et al.* 2007; Chesler *et al.* 2008; Collaborative Cross Consortium 2012). Each CC line is independently derived through three generations of outbreeding followed by multiple generations of inbreeding. The unique pedigree structure of this population during the first four generations (referenced hereafter as G_0, G_1, G_2, and G_2:F_1) allows us to observe up to eight informative meioses by genotyping a single G_2:F_1 sibling pair: each crossover observed in a G_2:F_1 individual can be assigned with certainty to a specific meiosis (Figure 1). Due to

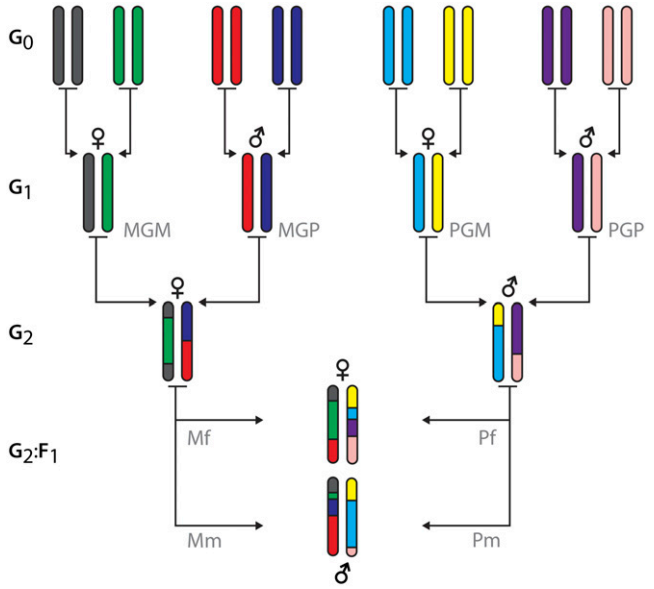


Figure 1 Pedigree structure of G_2:F_1 sibling pairs (representative funnel). Mice were bred in 237 funnels uniquely defined by the ordering of founder strains in the parental (G_0) generation, as illustrated. This design balances the contribution of each of the eight founder strains through both the maternal and the paternal lineage. Using funnel order and genotypes for both members of a sib pair, it is possible to assign crossover events inferred in members of the sib pair to one of eight meioses occurring in the germline of a specific ancestor: in the G_1 generation, maternal grandmother (MGM), maternal grandfather (MGP), paternal grandmother (PGM), or paternal grandfather (PGP); or in the G_2 generation, mother (Mf, Mm) or father (Pf, Pm), with independent G_2 meioses contributing to each sib pair. The method additionally distinguishes shared from singleton events.

randomization in the order of matings across breeding "funnels," these meioses have balanced contributions from each of the eight founder inbred strains through both female and male germlines. We genotyped 237 G_2:F_1 sib pairs from independent CC lines, using the Mouse Diversity Array platform (Yang *et al.* 2009), which allowed us to map each individual crossover to a genome interval of 35 kbp on average. This level of resolution is similar to what can be attained in LD and DSB methods while retaining the ability to determine the detailed composition of individual meioses, against a randomized and extremely diverse genetic background, capturing variants known to influence global recombination rate.

The overall conclusions of our study provide insights into the cellular and molecular mechanisms of recombination, provide new hypotheses on its evolution, and have practical consequences for the design and interpretation of mapping experiments in the laboratory mouse.

Materials and Methods

Mouse breeding

The G_2:F_1 population used in this study was bred at Oak Ridge National Laboratory (ORNL) beginning in 2005 as described in detail previously (Chesler *et al.* 2008; Collaborative Cross Consortium 2012). Briefly, eight founder strains—five classical inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, and NZO/HILtJ) and three wild-derived strains (CAST/EiJ, PWK/PhJ, and WSB/EiJ)—were intercrossed in the G_0 generation to generate G_1 hybrids (Figure 1). These G_1 progeny were intercrossed to create the four-way G_2 generation. Finally, mice from the G_2 generation were crossed to generate the eight-way progeny, known as the G_2:F_1. We refer to such a series of matings as a *funnel*. A funnel can be uniquely identified by the order of matings in the G_0 generation. Each breeding funnel constitutes a unique and independent mosaic of the eight founder genomes (Collaborative Cross Consortium 2012). Under this breeding scheme, the genetic contribution of each of the eight founder strains to each line is expected to be equivalent. The present work considers sibling pairs (each consisting of one female and one male animal) from the G_2:F_1 generation of 237 breeding funnels, for a total of 474 animals.

Two additional independent populations (to which we refer herein as the "intercrosses") were generated at the University of North Carolina at Chapel Hill (UNC) by performing intersubspecific crosses between FVB/NJ females and either (PWK/PhJ × CAST/EiJ)F_1 or (WSB/EiJ × PWK/PhJ)F_1 males. Subsets of 96 offspring from a larger total progeny from each cross were selected for this study. Throughout the article the dam is listed first and the sire last in all crosses.

G_2:F_1 and intercross mice were treated in accordance with the recommendations of the Institutional Animal Care and Use Committee of ORNL and UNC, respectively.

DNA preparation and genotyping

DNA from G_2:F_1 mice was isolated at UNC according to a protocol for isolation of high-molecular-weight DNA using proteinase K and phenol (Sambrook and Russell 2006) and genotyped using the high-density Mouse Diversity Array (Yang *et al.* 2009) at the Jackson Laboratory. DNA samples from intercross progeny were isolated at UNC and genotyped using the MegaMUGA platform (Neogen/Geneseek, Lincoln, NE), a new 77,000-probe array based on the Illumina Infinium platform (Rogala *et al.* 2014).

Haplotype reconstruction and inference of crossovers

We reconstructed founder contributions to the genomes of progeny, using microarray genotypes and pedigree information, and inferred crossovers as transitions between founder haplotypes. Crossovers are represented as intervals that span the physical genome between the nearest pair of markers that unambiguously flank the inferred event. G_2:F_1 and intercross progeny were analyzed separately; tables of all crossovers identified in both populations are included in Supporting Information, File S1 and File S2, respectively.

Collaborative Cross G_2:F_1 population: Haplotype reconstruction and localization of crossovers for CC funnels were

performed with GAIN (Liu *et al.* 2010), a hidden Markov model-based software that incorporates the CC pedigree structure in an efficient implementation of the Lander–Green algorithm (Lander and Green 1987) to obtain fully phased, highly accurate haplotype mosaics. Analysis was performed independently on each funnel but jointly on the siblings from the same funnel: crossovers are not shared across funnels but G_2:F_1 siblings from the same funnel can share crossovers from the G_1 generation, and the joint analysis can help to resolve ambiguity with respect to haplotype segment boundaries. Note that GAIN enforces all constraints imposed by the pedigree. For example, two of the strongest constraints (see Figure 1) are (1) for any G_2:F_1 mouse, the two alleles at any marker must come from different halves of the funnel, and (2) two siblings cannot inherit different alleles from one quarter-funnel at any marker. If the input data contain errors (in either genotype calls or funnel order), GAIN will infer many more crossovers than predicted to satisfy these constraints. This provides an effective indicator to identify and correct (or remove) incorrectly labeled funnels and poorly performing markers.

For each funnel, GAIN takes the funnel order, the genotypes of the eight founder strains, and genotypes of the G_2:F_1 sibling mice as input. It infers founder ancestry probabilities at each marker by building a descent model and evaluating the probability of crossover between adjacent markers, given the genotypes and a transition penalty. Founder ancestry at each marker is defined as the probability that each of $\binom{8}{2} = 28$ possible pairs of founders (*e.g.*, C57BL/6J and CAST/EiJ) are the two founders from which the two alleles at that marker are inherited. With such information, haplotype segments and crossovers can be obtained by tracing the maximum *a posteriori* ancestry probability along chromosomes, using the forward–backward algorithm as in the Lander–Green procedure. Each crossover is described by (1) proximal and distal boundaries where the probability of the most likely founder ancestry falls below a threshold, (2) proximal and distal founder ancestry on the recombining chromosome segment, and (3) the specific meiosis in the breeding funnel [maternal grandmother (MGM), maternal grandfather (MGP), paternal grandmother (PGM), paternal grandfather (PGP), mother (Mf, Mm), or father (Pf, Pm)] to which it is ascribed. The crossover interval inferred (from proximal to distal boundary) is expected to contain the crossover event with high probability. Note that there are regions where multiple founder ancestries have similar probabilities due to sparsity of markers, low genotyping quality, or similarity of DNA sequence among multiple founders (*e.g.*, due to identity-by-descent); in such cases, crossovers can be localized only to long intervals.

Before performing haplotype reconstruction, high-quality markers were identified by examining completeness (genotypes called in ≥99% of samples) and concordance of genotypes called in G_2:F_1 siblings, G_0 founder mice, and G_1 samples from previous studies, using the Mouse Diversity Array (Yang *et al.* 2009). A total of 121,504 markers (representing 15–25% of markers on each chromosome) were retained in the high-quality group (File S3) and used for initial haplotype inference. The resulting crossover intervals were refined using the remaining 549,595 mid- to low-quality markers by examining the consistency of the additional markers with their expected genotypes within and around the uncertainty limits of each crossover. On average, this reduced the width of the crossover intervals by approximately half.

Intercross population: Owing to the simplicity of the intercross pedigrees, haplotype reconstruction in the intercross population is nearly trivial. The problem reduces to identifying markers that are segregating between two paternal strains and then clustering them into haplotype blocks in a manner that minimizes the total number of crossovers. The algorithm used to reconstruct haplotypes from MegaMUGA genotype calls is described in File S7. It produces consistent results across a broad range of parameter choices and is insensitive to genotyping errors and noncrossover products.

Construction of genetic maps and estimation of local crossover density

Cumulative genetic maps for the CC funnels were computed directly from the interval representation of crossovers by integration across each chromosome to account for uncertainty in localization of crossovers. Maps were obtained separately for female and male meiosis and as a sex-averaged map. Genetic maps were computed from different subsets of the data, including separate maps for G_1 and G_2, and for genotype classes at specific loci including the *Prdm9* locus on chromosome 17. Maps were computed for each strain based on the haplotypes involved in junction formation. Cumulative maps were normalized to centimorgan units as

$$\mathrm{cM} = 100\left(\frac{\text{count of crossovers}}{\text{count of meioses} \times \text{count of funnels}}\right) \qquad (1)$$

and density estimates were obtained using a sliding window at variable widths. An R script to reproduce these analyses accompanies this article.

Pooled analysis of spatial distribution of crossovers across autosomes

To analyze the distribution of crossovers along all autosomes jointly, distinct crossovers on autosomes identified in the G_2:F_1 population and all autosomal crossoverss in the intercross population were analyzed together. Event positions were normalized against the length of the chromosome on which they occurred such that the normalized position of all events fell on the interval [0, 1]. For the purposes of this analysis, crossovers, inferred as intervals, were converted to points (at the midpoint between the proximal and the distal boundary): after normalization against chromosome length, event positions can be directly compared but interval widths are no longer meaningful. Kernel density estimates of

crossover frequency with respect to normalized chromosomal position were obtained with the KernSmooth package (Wand and Jones 1995) for the R language (R Core Team 2012: http://www.r-project.org/), using a Gaussian kernel with bandwidth selected according to the direct plug-in method (Wand and Jones 1995). Confidence intervals about the density estimate were obtained by repeating the kernel-fitting procedure on 10,000 bootstrap samples of the crossovers, taking the 2.5th and 97.5th percentiles at each point at which the original kernel was evaluated.

Sex differences in spatial distribution of crossovers were analyzed in two ways. First, we conducted a formal test of the hypothesis that the distribution of events along chromosomes is uniform in single-recombinant meiotic products. Events on single-recombinant chromosomes in each sex (which can be identified with certainty only at the G_2 generation) were compared to a uniform distribution by the Kolmogorov–Smirnov test, stratified by chromosome, and the resulting *P*-values were adjusted by the method of Benjamini and Hochberg to control the false-discovery rate. Second, we used quantile regression to test for sex differences in location at quantiles along the spatial distribution of crossovers. The following model was fitted for each chromosome,

$$y_q = \beta + \beta_{\text{male}} + \varepsilon, \tag{2}$$

where y_q is the physical position corresponding to the $(100 \times q)$th percentile of the spatial distribution of crossovers. In this model β_{male} gives the displacement of the male *vs.* the female distribution at the given quantile, such that values of $\beta_{\text{male}} > 0$ indicate a distally biased distribution in males.

The distribution of crossovers in males was compared to the distribution of recombination hotspots identified through high-resolution analysis of DSBs in Smagulova *et al.* (2011; Brick *et al.* 2012), using a simulation approach. For each chromosome, *n* DSBs (where *n* is the number of crossovers observed in the G_2:F_1 population) were sampled at random with replacement with probability proportional to hotspot strength. The distribution of DSBs and crossovers was compared by a Kolmogorov–Smirnov test. For each chromosome, a simulation *P*-value was obtained by taking the median *P*-value across 1000 random samples of DSBs. Simulation *P*-values for each chromosome were adjusted by the method of Benjamini and Hochberg to control the rate of false discovery.

Detection of cold regions

We initially identified cold regions, using a one-dimensional dynamic programming algorithm to identify regions with ≥10-fold reduction in frequency of crossovers via a generic scoring scheme (Karlin and Altschul 1990). Briefly, we first compute local crossover density ρ_i at a grid of points $i = 0, \ldots, n$ along a chromosome. Those densities are converted to initial scores e_i as

$$e_i = \lambda(1-\theta) + \rho_i \log\theta, \tag{3}$$

where λ is the mean crossover density across the chromosome and θ is a prespecified enrichment threshold (and here, $\theta = 0.1$). Then a forward pass is made over the scores to compute the dynamic-programming scores E_i,

$$E_{i+1} = \max\{0, E_i + e_{i+1}\}, \tag{4}$$

with the score at the first grid point initialized to zero ($e_0 \equiv 0$). Regions of interest are finally identified by performing a traceback on the E_i.

The dynamic programming method allows us to identify regions of reduced recombination without resorting to a fixed-width sliding window and can identify regions of any size. We applied the algorithm to male and female recombination separately and took the union of the results, which overlapped extensively. We retained only those regions of length >500 kbp and refined the boundaries of individual regions by manual inspection of their flanking crossovers.

Genomic analysis of cold regions

To capture the most up to date sequence information in all these analyses we used the newest mouse assembly, GRCm38, released by the Genome Reference Consortium in January 2012. Having obtained a list of candidate cold regions, we first determined the fraction of no calls (N's) in the reference genome sequence for a given interval and excluded those with >10% no calls. We then determined the following for each region: (1) the number and identity of crossovers in the region in the G_2:F_1 population; (2) the recombination rate in the population reported in Cox *et al.* (2009); (3) the DSB density in C57BL/6J, 9R, 13R, and $(9R \times 13R)F_1$ males based on the data reported in Brick *et al.* (2012); (4) whether the region overlaps a "recombination desert" reported in $(C57BL/10.S \times C57BL/10.F)F_1$ mice by Smagulova *et al.* (2011); (5) the number of crossovers in the 192 intercross progeny of $(PWK \times CAST)F_1$ and $(WSB \times PWK)F_1$ males; (6) the fraction of C+G base composition in the reference genome sequence; (7) the fraction of the region's sequence spanned by segmental duplications (of length >20 kbp); (8) the number of genes in the cold region; and finally (9) the number of structural variants reported in the region in the 18 strains sequenced by Yalcin *et al.* (2011).

Analysis of structural variation in cold regions

Segmental duplications were identified using dotplots generated by the software Gepard (Krumsiek *et al.* 2007). For each cold region, we compared the sequence in the region against itself to identify local segmental duplications (tandem or inverted) of length >20 kbp and computed the duplication rate as the fraction of the region spanned by such duplications. In determining the genome-wide segmental duplication rate, we used a sliding window of 1 Mbp with 500-kbp overlap between adjacent windows and calculated the average duplication rate across all windows.

To explore the relationship between segmental duplication, coldness for recombination, and structural variation we examined probe intensities for 62 representatives of the G_0

founders genotyped using the MegaMUGA platform. Each probe on this platform yields two intensity values per sample (x and y), locating samples in a two-dimensional intensity space that is partitioned into three discrete clusters (a homozygous cluster along each of the x- and y-axes and a heterozygous cluster along a 45° line drawn from the origin) for genotype calling. For a well-behaved SNP probe, the sum intensity ($x + y$) of all samples is approximately equal independent of genotype. Violations of this criterion provide evidence in support of structural variation. To capture the totality of information contained in probe intensities to provide evidence for structural variation between inbred strains, we performed principal components analysis (PCA) on the $n \times p$ matrix P of probe intensities for $n = 62$ individuals and p probes in each cold region to yield a new matrix P' containing the projection of M onto the subspace defined by the p principal components. Pairwise Minkowski distances (a generalization of the Euclidian distance metric) between samples were computed on M', and an unrooted neighbor-joining tree was constructed from the resulting $n \times n$ distance matrix, using the R package ape (http://ape-package.ird.fr/). We repeated the analysis using pairwise intensities (and thus an $n \times 2p$ data matrix).

Because off-target variation in or near the probe sequence may influence hybridization efficiency (Didion *et al.* 2012), such that samples more genetically distant from the reference used to design the array have systematically lower probe intensities, we restricted this analysis to the five closely related classical inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, and NZO/H1LtJ).

Results

Overview of the mapping population and recombination map

The genotypes of 237 G_2:F_1 sibling pairs were combined with the funnel information (see *Materials and Methods*) to infer fully phased founder haplotypes for each chromosome (Figure 1). Note that each G_2:F_1 female is the product of six meioses and her male sib is also the product of six meioses with four meioses shared between them. The four shared meioses (denoted as MGM, MGP, PGM, and PGP in Figure 1) occur at the G_1 generation. The four unique meoises (Mf, Mm, Pf, and Pm in Figure 1) occur at the G_2 generation. Crossovers are identified as transitions between founder haplotypes. Given the structure of the pedigrees, every crossover can be unambiguously assigned to one of the eight meioses, since at any given locus, any pair of founder haplotypes can be paired only in one of these eight meioses. Therefore, the identification and analysis of crossovers relies on knowledge of the order of the founder strains in the G_0 generation.

The breeding scheme ensures equal and balanced contribution of each of the eight founder strains in the autosomes and equal representation of male and female meioses. Genotyping of both male and female offspring allows the study of recombination on the X chromosome while retaining the ability to test whether the distribution of the observed crossovers is influenced by sex of the offspring analyzed.

We identified 25,038 crossovers in the 474 individual G_2:F_1 mice (Table 1). Of these, 18,948 events are observed only once and 3,045 crossovers are shared by both siblings in the pair (*i.e.*, were observed exactly twice). Therefore, we have identified 21,993 distinct crossovers in our population, 21,368 on the autosomes and 625 on the X chromosome. Because a crossover in generation G_1 can be observed only if it is transmitted to at least one G_2:F_1 individual (which occurs with probability 3/4) while all G_2 events are observable, the effective number of observed autosomal meioses per funnel is $(3/4)(4 \text{ meioses}) + 1(4 \text{ meioses}) = 7$ and the thus the sex-averaged autosomal map length is 1288 cM. The effective of number of meioses observed per funnel on the X chromosome is 3.5, by similar logic, giving an overall length of 75 cM.

We subjected the raw recombination data to a comprehensive quality-control pipeline to detect errors and to test expectations regarding the ratio at which different types of events should be observed according to Mendelian rules (File S7 and Table 2). In all cases the data conform closely to expectations, providing strong evidence for the integrity of both the breeding process and our methods for constructing the recombination map.

A key feature of our experiment is that crossovers are mapped with high precision. The median uncertainty interval is 34.9 kbp but the range is wide (19–25,150,032 bp) and the distribution bimodal (Figure S1). The uncertainty intervals in the right tail of the distribution are due to either the lack of SNPs in some genomic regions (Yang *et al.* 2009) or the lack of *informative* SNPs between the strains involved in particular crossovers, as a consequence of recent shared ancestry (Yang *et al.* 2011). All analyses of the spatial distribution of recombination account for these uncertainties (see *Materials and Methods*).

The resolution of our map is ~114 kbp (8.8 crossovers per megabase pair). The spatial distribution of crossover density estimated over large windows (for example, 5 Mbp; see panel 1 in Figure S2) in the G_2:F_1 population is qualitatively similar to that of the most current mouse linkage map (Cox *et al.* 2009). This similarity gradually disappears as the window is narrowed, likely due to a combination of technical (density and informativeness of the marker panel) and biological (sex and strain effects) differences between these experiments (panel 2 in Figure S2). The linkage map reported by Cox *et al.* (2009) was based on reanalysis of pedigrees from the eight-way heterogenous stock (HS). The HS founders are classical inbred strains; these partially overlap the CC founders but do not include the wild-derived strains. The Cox map is based on 3546 informative meioses—roughly double the number of meioses in the G_2:F_1 population—but the HS animals were genotyped at only 10,202 informative markers.

Table 1 Crossover counts in the G_2:F_1 population

			Autosomes			X chromosome			
Sex	Meiosis	Sib	Singleton	Shared	Unique	Singleton	Shared	Unique	Total
Female	Mf	Female	3,282	NA	3,282	183	NA	183	3,465
Female	Mm	Male	3,255	NA	3,255	150	NA	150	3,405
Male	Pf	Female	2,871	NA	2,871	NA	NA	NA	2,871
Male	Pm	Male	2,783	NA	2,783	NA	NA	NA	2,783
Female	MGM	Female	826	756	1,582	35	48	83	
		Male	767	756	1,523	35	48	83	
		All	1,593	756	2,349	70	48	118	2,467
Male	MGP	Female	733	730	1,463	NA	NA	NA	
		Male	768	730	1,498	NA	NA	NA	
		All	1,501	730	2,231	NA	NA	NA	2,231
Female	PGM	Female	807	766	1,573	174	NA	174	
		Male	782	766	1,548	NA	NA	NA	
		All	1,589	766	2,355	174	NA	174	2,529
Male	PGP	Female	740	745	1,485	NA	NA	NA	
		Male	757	745	1,502	NA	NA	NA	
		All	1,497	745	2,242	NA	NA	NA	2,242
		All							
Female	All	All	9,719	1,522	11,241	577	48	625	11,866
Male	All	All	8,652	1,475	10,127	NA	NA	NA	10,127
Total	All	All	18,371	2,997	21,368	577	48	625	21,993

Sex and strain effects on overall level of recombination

As expected, the total number of crossovers in autosomes is significantly smaller in male meiosis than in female meiosis (10,127 events and 11,241 events, respectively, Table 1; $P = 2.5 \times 10^{-14}$ by *t*-test, H_0: equal number of events in both sexes). Table 3 provides point estimates and 95% bootstrap confidence intervals for sex-specific map length at each generation (lack of overlap between two such intervals is asymptotically equivalent to rejection of the null hypothesis that the corresponding point estimates are equal by Student's *t*-test). The sex effect on recombination is more pronounced in G_2 than in G_1. Note that the female map expands in G_2 compared to G_1 (although the effect is not statistically significant) while the male map contracts significantly between these two generations (Figure 2). The sex difference is also observed in the distribution of total number of crossovers per individual (Figure S3). A greater number of crossovers in female meiosis are observed in both the G_1 and G_2 generations ($P = 0.019$ and $P = 1.3 \times 10^{-15}$ by *t*-test, respectively; H_0: equal number of events in both sexes). Table 3 provides point estimates and 95% bootstrap confidence intervals for sex-specific map length at each generation.

To investigate the possible causes of these differences we determined the effect of strain in the sex-specific maps at each generation. Two previous studies have mapped a strong quantitative trait locus (QTL) controlling male map length on the X chromosome (Murdoch *et al.* 2010; Dumont and Payseur 2011). Both studies reported that a QTL associated with the inheritance of the CAST/EiJ chromosome X resulted in significant expansion of the male map. To confirm and extend this observation we compared the number of crossovers in the progeny of G_1 males, which are hemizygous for chromosome X and are heterozygous for exactly two founder haplotypes on the autosomes, classified according to the subspecific origin of the X chromosome (Figure 3). We observe that the *M. m. castaneus* (CAST/EiJ, discounting regions of intersubspecific introgression) X chromosome is associated with an expansion of the male map, the *M. m. musculus* (PWK/PhJ) X chromosome is associated with contraction of the male map, and the *M. m. domesticus* X chromosomes yield intermediate male map length. We conclude that one or more loci on the X chromosome controlling the length of the male map segregate in the CC.

We additionally observed an overall effect of the autosomal genome on recombination rate that is in opposition to the effect of chromosome X. The CAST/EiJ autosomal background is associated with a contraction of the male map while the PWK/PhJ autosomal background is associated with expansion of the male map. We attempted to identify specific autosomal loci and estimate their effects by performing genome scans for both the number of crossovers observed and magnitude of crossover interference in the G_2 generation, but no locus reached genome-wide significance. The generational difference in sex-specific map lengths (Table 3) is consistent with a model in which the genotypes of the X chromosome and autosomes influence total map length additively but in opposite directions. In generation G_1 the X and autosomal loci are tightly coupled and the increasing effect of CAST/EiJ autosomal loci offsets the decreasing effects of the CAST/EiJ X chromosome. The same kind of balance occurs in the PWK/PhJ background but with directions of X and autosomal effects reversed. In generation G_2 the segregation of X and autosomal genotypes is less tightly

Table 2 Expectations regarding autosomal recombination in G_2:F_1 pedigrees based on Mendelian rules

Relationship	Expected	Observed
Observed events in G_1 *vs.* G_2	3:4	0.747 (P = 0.78)
Same event type from opposite half-funnels	1:1	1.00 (P = 1.00)
Mf *vs.* Mm		1.02
Pf *vs.* Pm		1.03
MGM *vs.* PGM		0.975
MGP *vs.* PGP		0.999
Singleton *vs.* shared events in G_1	2:1	0.673 (P = 0.17)
MGM		0.678
MGP		0.672
PGM		0.675
PGP		0.668

Except when noted otherwise, the counts to which this table refers are those in Table 1. All P-values were obtained via a χ^2-test with a single degree of freedom. A more detailed discussion of these expectations is provided in File S7.

Table 3 Cumulative autosomal map length in the G_2:F_1 by sex and generation

	Generation	
Sex	G_1 (cM)	G_2 (cM)
Female	1330 (1296, 1365)	1381 (1351, 1409)
Male	1279 (1247, 1311)	1211 (1182, 1236)

Confidence intervals were obtained from a nonparametric bootstrap procedure with 100 replicates.

coupled, giving rise to transgressive allelic combinations such as CAST/EiJ autosomal genotypes in combination with a PWK/PhJ X chromosome and leading to increased difference between the two sexes.

Sex and strain effects on spatial distribution of crossovers and interference

In addition to the sex differences in map length, we observe dramatic sex differences in the spatial pattern of crossovers along the autosomes. The raw distribution is presented in Figure 4, left, while Figure 4, right, displays a smoothed representation of their spatial distribution after standardizing chromosomes to unit length and estimating the density of crossovers on all autosomes jointly. Because mouse chromosomes are acrocentric and the function of their short arms is poorly understood, we follow the convention of taking "distal" and "telomeric" to refer to position on the long arm. Approximately half of the crossovers occur in the distal quarter of the chromosomes and almost one-third of events occur in the distal 10% of the autosomes in male meioses (Figure S4A). To demonstrate that this effect is independent of chromosome size or identity, we used quantile regression to assess the effect of sex on the cumulative spatial distribution of events along chromosomes. Figure S4B confirms that the distal portions of all 19 autosomes bear a disproportionate number of crossovers in male *vs.* female meioses.

The spatial distribution of crossovers on single- and double-recombinant chromosomes differs as expected due to interference, but preserves the excess of crossovers in the distal ends of chromosomes in males (Figure 5). Considering only single-recombinant chromosomes to reduce the confounding effect of interference, the distribution of events is indistinguishable from a uniform distribution in females for 11 of 19 autosomes (chromosomes 2, 4, 5, 6, 8, 9, 10, 11, 12, 13, and 18) at false-discovery rate 0.05 (by a Kolmogorov–Smirnov test) and is qualitatively uniform and symmetric on the remaining autosomes. The distribution in males for single-recombinant chromosomes differs from the uniform case for all chromosomes (median $P = 1.7 \times 10^{-8}$). To confirm these results using independent data we performed a similar analysis of 192 intercross offspring from two types of F_1 males [(CAST/EiJ × PWK/PhJ)F_1 and (WSB/EiJ × PWK/PhJ)F_1] that were genotyped with the 77,000-SNP MegaMUGA array. In this cross only crossovers in male meioses are observable, but their distribution (bottom row of Figure 5) closely mirrors that in the G_2:F_1 population. Notably, the spatial pattern of crossovers in male meiosis is significantly different from the distribution of DSBs and of the X chromosome effect described in the previous section (Figure S5 and File S5). This pattern is consistent with a model in which recombination is spatially polarized in the male but not the female germline (see *Discussion*).

The large-scale (>5 Mbp) recombination landscape and differences between males and females in the distribution of crossovers are conserved across all eight strain backgrounds. However, at the fine scale (<0.5 Mbp) we begin to observe heterogeneity in the distribution of crossovers with respect to strain genotypes at the recombination breakpoint (Figure S6). While our study lacks the resolution required to identify individual hotspots of recombination or recombination-associated sequence motifs, these fine-scale patterns almost certainly reflect the strain-specific nature of most hotspots. Tightly clustered strain- and sex-specific breakpoints are observed for both male and female meiosis across all chromosomes. A male-specific PWK/PhJ cluster on chromosome 15, in which a physical region of <0.1 Mbp encompasses 6 cM, is a notable example (Figure S7).

Identification of cold regions in the G_2:F_1 population

We define *cold regions* as long (>500 kbp) contiguous genomic intervals that have a >10-fold reduction in crossovers compared to the genome-wide average. Given the total number of crossovers in our experiment (21,993) we established this 500-kbp threshold in the initial identification of cold regions to minimize the number of false positives (*i.e.*, on average we expect 4.4 crossovers per 500 kbp).

We initially identified the 50 coldest regions in male and female meioses separately to allow for possible cold regions on chromosome X. Because of the substantial overlap between these sets, we took their union, totaling 68 regions, as our candidate set. Candidate regions underwent several filtering steps. First, we excluded regions in which N's in the reference genome sequence represent a large fraction (>12.5%) of the nominal length. We refined the boundaries

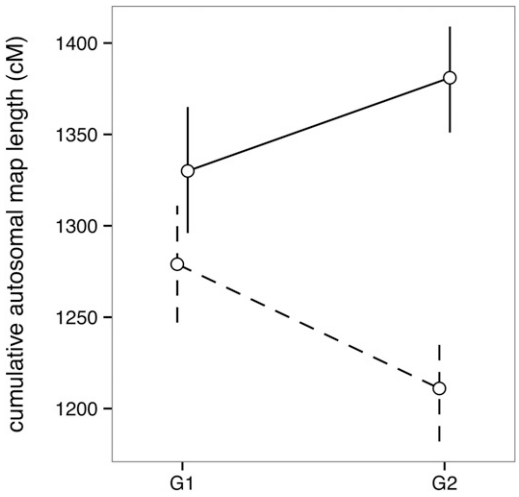


Figure 2 Sex-specific cumulative autosomal map length. Map length was computed within each generation among females (solid line) and males (dashed line). Vertical bars indicate 95% bootstrap confidence intervals.

of the 59 remaining cold regions, using the crossover uncertainty intervals in the G_2:F_1 population. After refinement, 51 of these regions have no crossovers; they are bounded proximally by the distal boundary of the proximal crossover and distally by the proximal boundary of the distal crossover (File S4). The remaining 8 cold regions have only one or two crossovers. Overall, cold regions span 124.1 Mbp (≈5% of the genome), distributed along 18 chromosomes (all chromosomes except 10 and 11), with an enrichment for proximal and distal sections of the chromosomes. On average, cold regions are almost 2.1 Mbp in length (range 582,221–10,217,265 bp). Importantly, cold regions are consistently "cold" across strains, sexes, and generations. The refined cold regions are plotted against their genomic position in Figure 6.

External validation of cold regions: To determine whether the cold regions seen in the G_2:F_1 population are cold in other populations, we examined these regions in the HS used to construct a recent linkage map of the mouse (Cox *et al.* 2009). On average, there is a fourfold reduction in crossover density in cold regions (0.14 cM/Mbp *vs.* the expected 0.5 cM/Mbp that is observed genome-wide; File S4). In fact, 57 of the 59 regions are below the genome-wide average and for 16 regions the recombination density in the HS is zero. The two exceptions are located on chromosome 7 (positions 7,231,821–12,298,098 and 110,909,130–111,734,201). The extent of validation is striking given the differences in genetic background between the two populations: of the eight strains contributing to the HS, only A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, and NZO/HILtJ—all of which are of nearly pure *M. m. domesticus* origin—are shared with the G_2:F_1 population. The strains not shared include two wild-derived strains representing subspecies (CAST/EiJ, *M. m. castaneus*; and PWK/PhJ, *M. m. musculus*, respectively) that are rare or absent in the genetic makeup of the strains in the Cox study and are separated from *M. m. domesticus* by 500,000 years of evolution (Yang *et al.* 2011). Furthermore, marker density and methods used to estimate local recombination density were quite different between the present study and the Cox study.

Recently, several maps of recombination-initiation sites in the mouse have been published (Smagulova *et al.* 2011; Brick *et al.* 2012). These studies identified regions with significant enrichment of DSBs in the male germline of mice of different genetic backgrounds [(C57BL/10.S × C57BL/10.F)F_1, 9R, 13R, C57BL/6, and (9R × 13R)F_1]. The initial study (Smagulova *et al.* 2011) identified 21 recombination deserts >3 Mbp, but noted that it was not possible to identify DSBs in some regions due to gaps in the reference sequence or highly repetitive DNA. Eleven of the cold regions identified in the G_2:F_1 population overlap with those described previously in one of these two studies. This level of concordance is even more remarkable in light of the fact that one of the Smagulova deserts was eliminated from our analysis because of a complete lack of sequence (chromosome 7, 39 Mbp; see also new GRCm38 assembly of the mouse genome), and 9 additional regions that fail to pass thresholds for inclusion in our list nonetheless show low levels of recombination in the G_2:F_1 population. More importantly, data from the second study (Brick *et al.* 2012) can be used to estimate the density of DSBs in any given region. On average, there is an 18-fold reduction in DSB density (range 14- to 24-fold; Wilcoxon's rank-sum test, $P = 1.4 \times 10^{-50}$; H_0: equal number of DSBs per 500-kbp window in cold regions *vs.* noncold regions) in cold regions compared to the genome-wide average (Figure S8), with the caveat that DSBs in repetitive sequence typically cannot be mapped via high-throughput sequencing. Across the four strain backgrounds of these two studies, an average of 9 (range 8–11) of our cold regions have no DSBs and less than half of these have low-complexity or nonunique sequence that may interfere with sequencing-based DSB identification (File S4).

We conclude that the vast majority of cold regions identified in the G_2:F_1 population represent *bona fide* regions of suppressed recombination that are independent of genetic background and strongly associated with a reduction in density of recombination-precursor sites.

Genomic analysis of cold regions: We analyzed cold regions with regard to several genomic features that have been associated with suppressed crossing over in other regions (such as centromeres): low C+G content, low gene content, and enrichment for complex duplications and repeated sequences (López-Flores and Garrido-Ramos 2012). The results are presented in File S4.

The overall C+G content in cold regions is lower than the genome-wide average (38.7% *vs.* 42%), but the distribution is multimodal (Figure S9). We observed a marked enrichment for relatively long (>20 kbp) segmental duplications, either in tandem or inverted, in cold regions. On average,

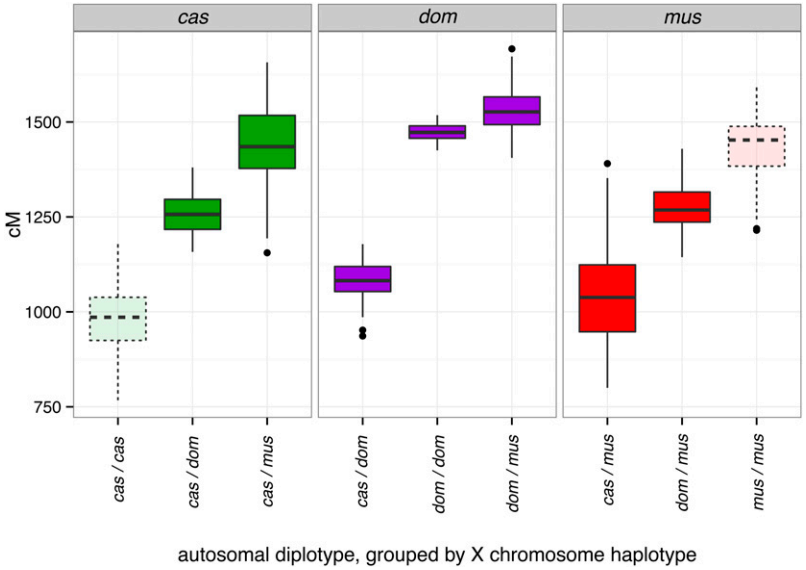


Figure 3 Effect of the X chromosome on overall recombination rate in male meiosis at G_1. Males at the G_1 generation are hemizygous for a single founder haplotype along the entire X chromosome and heterozygous for that founder haplotype everywhere else, allowing a direct test for the effect of the X chromosome against a varying genetic background. Alleles were defined on the basis of subspecific origin: CAST/EiJ, *M. m. castaneus* (*cas*); PWK/PhJ, *M. m. musculus* (*mus*); and all others, *M. m. domesticus* (*dom*), with the caveat that regions of intersubspecific introgression will be misclassified. Unique crossovers attributed to G_1 male meioses (classes MGP and PGP, Figure 1) were counted within all seven possible combinations of X chromosome haplotype and autosome diplotype (*e.g.*, *cas/dom*) for the G_1 offspring of a CAST/EiJ dam and an A/J sire. Boxplots represent map lengths computed on 100 bootstrap replicates of the observed G_1 crossovers. The two boxes with dotted outlines represent predicted values for allelic combinations not observed in the G_2:F_1 population.

almost 32% of the sequence of a single cold region is composed of segmental duplications; but in aggregate, segmental duplications account for 47.7% of the total sequence spanned by the 59 cold regions. In fact, the size and segmental duplication content of cold regions are correlated ($r^2 = 0.664$). However, 16 cold regions are completely devoid of segmental duplication.

Discussion

Using a unique multiparental intercross design, we have produced a new linkage map for the mouse genome that extends our understanding of the genetic and genomic underpinnings of meiotic recombination. Our map has three key features: very dense genotyping, balanced contributions from each sex, and the presence of uniformly high levels of genetic diversity across the genome. We observe that genetic background influences the overall level of recombination in a sex-specific manner. The pattern of genetic effects in G_1 males implicated loci on both the X chromosome and the autosomes. Interestingly, these effects worked additively and in opposition to produce relatively similar levels of recombination across founder strains. We were unable to precisely map any QTL and it appears likely that multiple unmapped autosomal loci may contribute to the overall recombination rate, a conclusion in line with previous QTL-mapping experiments (Murdoch *et al.* 2010; Dumont and Payseur 2011). However, our measure of recombination in each individual, the number of crossovers inferred in at most two offspring, is substantially noisier than cytological approaches to estimate recombination used in those previous studies. Such techniques usually involve visualization and counting of crossover-associated proteins via immunofluorescence in a sample of single spermatocytes (Anderson *et al.* 1999) and yield tens of observations (*i.e.*, independent meioses) per individual. In practice, such techniques are difficult to apply in the female germline and cannot precisely resolve the genomic location of crossovers. Our design, by contrast, allows both the counting of crossovers (through both male and female meioses) and the localization of those events with great precision, enabling us to analyze the effects of sex on both the level and the genomic distribution of recombination and to identify cold regions for recombination. Larger sample sizes and experimental designs that randomize the autosomal allelic combinations more efficiently, such as the diversity outbred (DO) population (Svenson *et al.* 2012), will be required to map loci affecting overall recombination rates.

We observe an increase in crossover density in the distal ends of every mouse autosome in male meiosis. This observation holds independent of the genetic background, chromosome size, and number of crossovers per chromosome and leads to the concentration of almost one-third of all crossover events in the distal 10th of autosomes in male meiosis. Similar patterns are observed in humans and chimpanzees (Broman *et al.* 1998; Auton *et al.* 2012).

Sex effects on overall map length in eutherian mammals (Dunn 1920; Dunn and Bennett 1967), and to a lesser extent sex differences in regional crossover density (Broman *et al.* 1998; Dumont and Payseur 2011), are not novel observations. However, the spatial and temporal precision with which we identify crossover events in this experiment allows us to link meiotic outcomes to their generating processes. We do so by exploiting the asymmetry of peaks of crossover density among chromosomes with different numbers of crossover events (Figure 4). That both the chromosomal position and the shape of the distal peak in male meiosis are identical on single- and double-recombinant chromosomes—independent of strain background, chromosome size, or sequence content—suggests that the distal event occurs via a regulated process. A process that consistently results in a concentration of crossover events in the distal portion of chromosomes must be polarized either in space or in time. Remarkably, the distribution of crossover events is

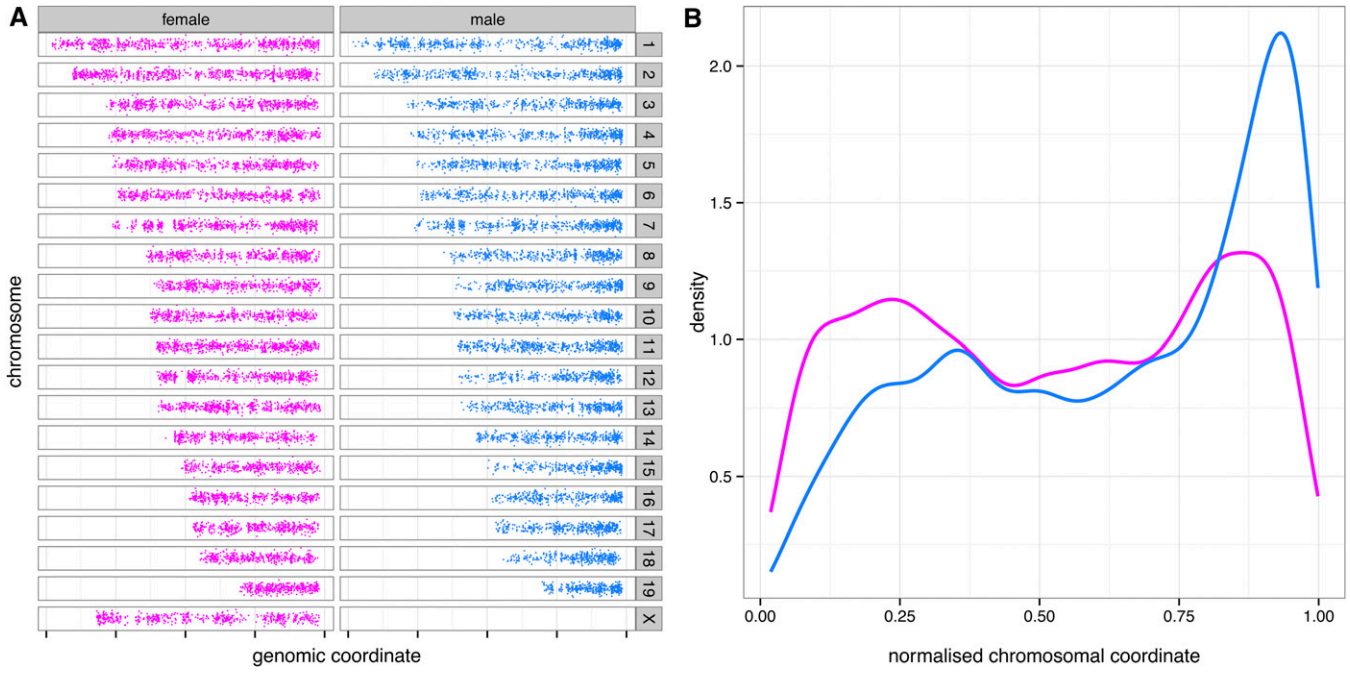


Figure 4 Spatial distribution of crossover events in male and female meioses in the CC G_2:F_1. (A) spatial distribution of crossover events in each experiment: each dot represents the midpoint between the proximal and distal boundaries of a unique crossover event occurring in a female (pink) or a male (blue) meiosis. Chromosomes are aligned at the distal ends of their long arms. (B) Kernel density estimates of crossover frequency across all autosomes in female (pink lines) or male (blue lines) meioses, after scaling all chromosomes to equal length. Note the dramatic difference in distribution of events in female *vs.* male meioses: crossover events are distributed almost uniformly across autosomes in females, but are concentrated in the distal end of autosomes in males.

clearly different from the density and individual strength of hotspots identified through high-resolution analysis of DSBs (Brick *et al.* 2012; Khil *et al.* 2012) (Figure S5). The fact that nonrecombining regions of the sex chromosomes are nonetheless marked by DSBs in male germ cells (Wojtasz *et al.* 2012) provides further evidence for an additional layer of regulation between DSB formation and their resolution as crossovers in males.

The conservation of the broad-scale features of the recombination landscape across divergent species strongly suggests that the recombination landscape in male meiosis is controlled in large part at the chromosomal level. Such a model is consistent with the observation that in many organisms, including male mammals, chromosome pairing and synapsis progress from the telomeres inward (reviewed in Hunter 2003). The early steps of meiosis thus provide a simple and universal mechanism for identifying chromosome ends independently of the number, morphology, and sequence composition of the chromosomes in a given species. It is attractive to speculate that enrichment for crossovers in the distal ends of chromosomes in males evolved as a result of the obligation for a crossover in the short and subtelomerically located region of homology between the X and Y chromosomes known as the *pseudoautosomal region* (PAR) (Burgoyne 1982; Mohandas *et al.* 1992). Once the mechanism that ensures the presence of a crossover in the PAR evolved in males, it is easy to imagine that it was co-opted to act in every other chromosome. The absence of this constraint in females would explain the sex differences in crossover density. Marsupials, an outgroup to eutherian mammals whose sex chromosomes lack a PAR, provide preliminary support for this hypothesis. The program for pairing and segregating the sex chromosomes in marsupials differs in important ways from that in eutherians (Page *et al.* 2005), and the most recent linkage maps for marsupial species including wallaby (Zenger *et al.* 2002) and opossum (Samollow 2004) show no evidence for male-specific expansion in subtelomeric regions. The evolution of eutherian sex chromosomes may thus be intimately linked to chromosome-scale behavior of the recombination machinery.

The observed sex differences in map length and spatial distribution of recombination have practical implications for the design of mapping experiments. While the female map is longer, mapping through males will provide improved resolution for loci in the distal regions of chromosomes. Furthermore, the field of mouse genetics is increasingly embracing the use of multiparental reference populations such as the CC (Collaborative Cross Consortium 2012), the HS (Valdar *et al.* 2006), and the DO (Svenson *et al.* 2012). Genetic analysis in these populations relies on the accurate identification of the founder haplotype inherited locally by each individual, a task that becomes increasingly difficult in the chromosome ends and in the presence of increasing levels of recombination. Genotyping platforms can be designed to overcome these challenges: the new MegaMUGA array spaces markers on a grid defined by genetic

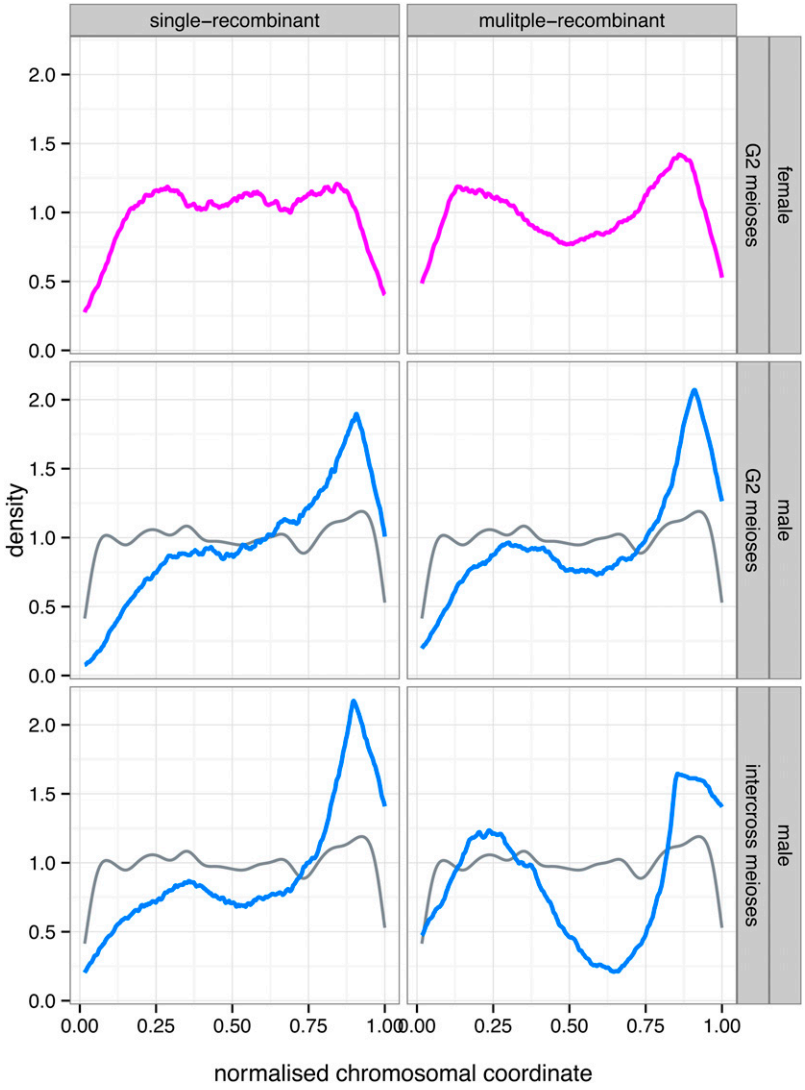


Figure 5 Asymmetric spatial distribution of crossovers in male meiosis in singly *vs.* multiply recombinant chromosomes. Shown are kernel density estimates of crossover frequency across single-recombinant (left) and double-recombinant (right) autosomes in female (pink lines) and male (blue lines) meioses. The sharp peak in crossover density observed at the distal end of autosomes in males is present in both single- and double-recombinant autosomes. An additional broader peak is present in the proximal region of double-recombinant chromosomes. This effect is independent of the distribution of double-strand breaks (DSBs) in male meiosis reported by Brick *et al.* (2012), plotted in gray.

rather than physical distance (Rogala *et al.* 2014), allowing precise and accurate detection of subtelomeric crossovers.

Finally, we have identified >50 regions of dramatically reduced recombination that we term *cold regions*. These regions represent the left tail of the sampling distribution of local recombination rate. If cold regions are purely artifacts of sampling, then we would expect their location to be random, such that they should not be replicated in independent experiments and that their number and size should decrease as an increasing number of meioses are sampled. However, a majority of our cold regions are replicated in three independent experiments—the HS, the DO, and an intercross between CAST/EiJ, PWK/PhJ, and WSB/EiJ—representing at least an order of magnitude more independent meioses than were captured in the G_2:F_1 (summarized in File S4). Furthermore, our cold regions are depleted for DSBs (Wilcoxon's rank-sum test, $P = 1.4 \times 10^{-50}$; H_0: equal number of reads per 1-Mbp window in cold regions *vs.* noncold regions) in a range of genetic backgrounds. In an F_2 or backcross design a cold region could arise due to a simple inversion, but this is unlikely to be case in our experiment: all eight G_0 founder strains would have to carry a different large inversion allele.

One-third of cold regions span genomic regions enriched for segmental duplications, and local recombination rate and density of segmental duplications are inversely correlated genome-wide [Pearson's $\rho = -0.26$, 95% C.I., (-0.28, -0.23); Figure S10 and File S6]. In fact, some of the larger and more complex tandem duplications and rearrangements in mice are among the coldest regions for recombination (Figure S11). As in pericentromeric regions, accumulation of repeated sequences could be either a cause or an effect of suppressed recombination. Either case leads to a paradox: tandem duplications are generated through unequal (*i.e.*, nonallelic) recombination, but levels of crossing over in these regions are exceedingly low. A simple explanation for this paradox is that some cold regions can be hotspots for *de novo* structural variation. This hypothesis predicts that cold regions with segmental duplications should be enriched for structural variants and that multiple independent variants should arise over short evolutionary times.

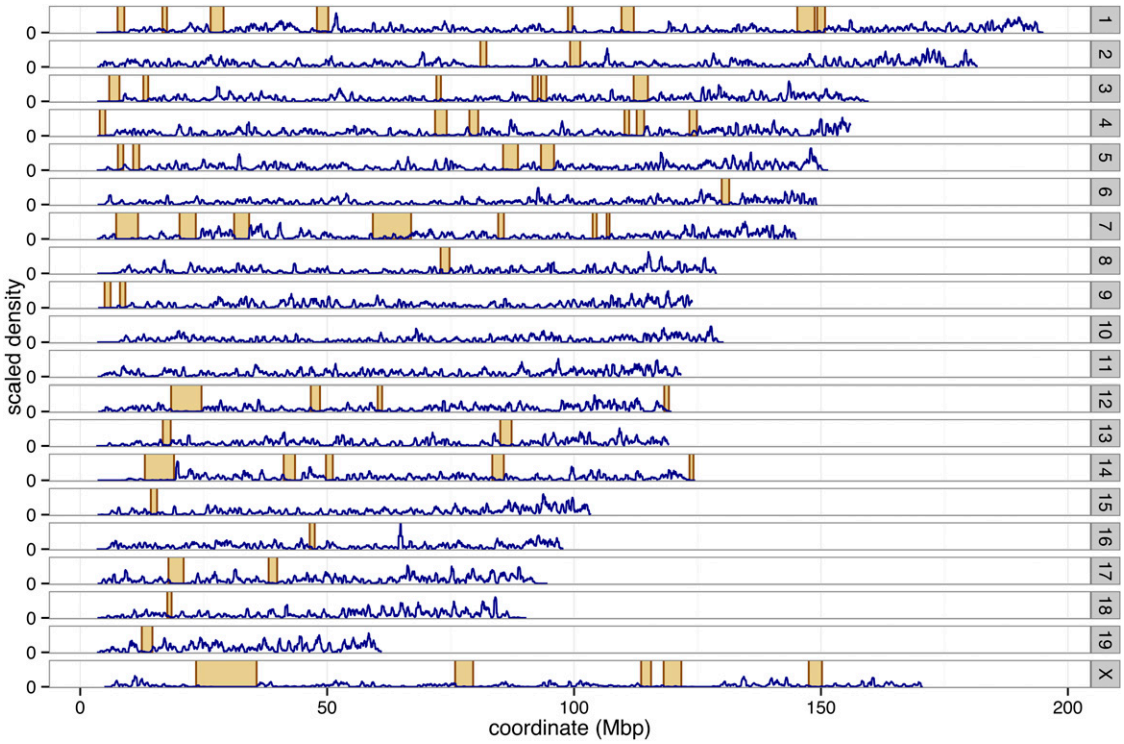


Figure 6 A catalog of 59 *cold regions* of dramatically reduced crossover frequency. Cold regions are represented as yellow-shaded boxes. Local rate of crossing over (in arbitrary units) is plotted as a dark blue line.

Support for this hypothesis can found first and foremost in a previous study that observed *de novo* structural variants within a single inbred strain, C57BL/6J, in four of our cold regions, using tiling arrays (Egan *et al.* 2007). In fact, structural variants are twice as common in cold regions as in the rest of the genome—in the eight founders of the CC; in the five founders that are classical laboratory strains; and in comparisons between two sister inbred strains, C57BL/6J and C57BL/6NJ (File S4). However, the true relative incidence of structural variants in cold regions likely exceeds this estimate. The best available catalog of structural variants in 18 laboratory inbred strains (Yalcin *et al.* 2011) has limited coverage of regions rich in segmental duplications (including our cold regions) due to the inherent difficulty in accurately aligning short reads to repeated sequences (Figure S12).

To circumvent the limitations of sequencing, we took advantage of data from the MegaMUGA genotyping array to explore the relationship between segmental duplications, coldness for recombination, and structural variation. This approach finds evidence for structural variation in segmentally duplicated portions of our cold regions. A representative example (chromosome 5: 93–96 Mbp) is displayed in Figure 7. This region has complex segmental duplication in the C57BL/6J reference sequence (see dotplot in Figure 7A). The wide variation in sum intensity at probes within the region (Figure 7B) suggests differences in copy number between strains. Inspection of the two-dimensional intensity plots (Figure 7C) demonstrates that the three strains (A/J, C57BL/6J, and NZO/HILtJ) with highest sum intensity carry at least two paralogous alleles at the first two probes, while the remaining classical inbred strains carry only one allele. At the third probe three strains (again A/J, C57BL/6J, and NZO/HILtJ) carry a different allele but in equal copy number to two wild-derived strains, CAST/EiJ and PWK/PhJ. In contrast, three strains (129S1/SvImJ, NOD/ShiLtJ, and WSB/EiJ) have intensities consistent with the presence of zero copies. Finally, the tree derived from principal components analysis of intensities in this region (Figure 7D) clearly segregates the classical inbred strains in a pattern consistent with panel sum intensities (Figure 7B) but inconsistent with local phylogeny based only on SNPs (Figure 7E) (Yang *et al.* 2011; Wang *et al.* 2012; Welsh *et al.* 2012). Taken together, these data provide additional evidence that structural variants arise multiple times in cold regions with segmental duplication. Studies in wild mice, both from *Mus* and sister clades, can provide further insights regarding the evolutionary significance of cold regions with respect to both genome stability and patterns of linkage disequilibrium.

In conclusion we report here the newest version of the linkage map for the laboratory mouse, with more precise localization of crossover events and more comprehensive coverage of segregating variation than previous maps. Although a study of this type is the most basic and traditional work in genetics, our results provide several new and important insights into the nature and spatial pattern of recombination across generations. First, we expand knowledge of sex differences in regional crossover

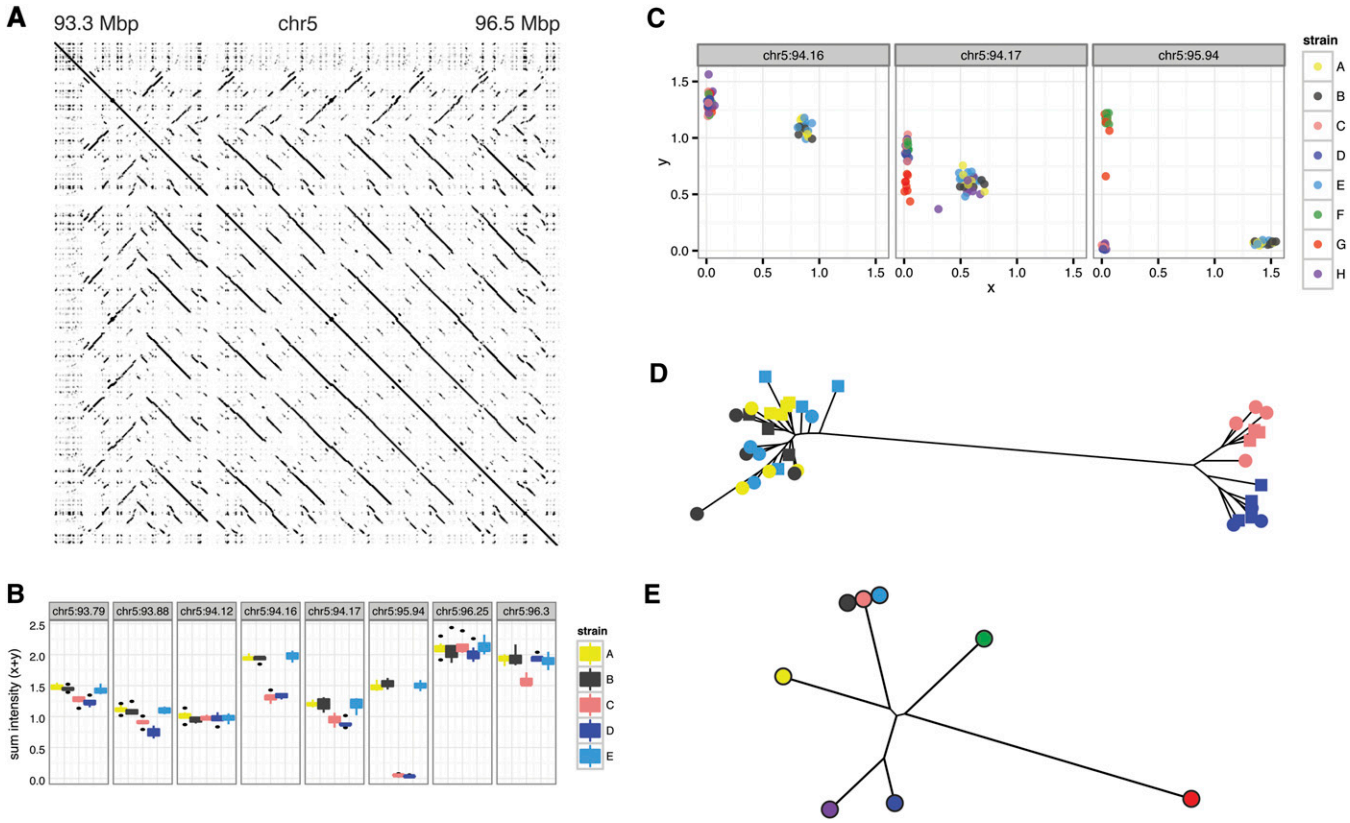


Figure 7 Cold regions enriched for complex tandem duplications and inversions show evidence of structural variation between inbred strains. Strain abbreviations are as follows: A, A/J; B, C57BL/6J; C, 129S1/SvImJ; D, NOD/ShiLtJ; E, NZO/H1LtJ; F, CAST/EiJ; G, PWK/PhJ; and H, WSB/EiJ. Color coding of strains is consistent across panels. (A) Dotplot from self-alignment of reference sequence (GRCm38/mm10 assembly) in a representative cold region. A dark dot represents >95% sequence identity between a pair of 100-bp windows. Linear clusters parallel to the main diagonal thus represent tandem duplications, and linear clusters perpendicular to the main diagonal represent duplication with inversion. (B) Profile of sum intensity at the eight probes (GRCm38/mm10 coordinates) within the region, for classical inbred strains only (strains A–E). (C) Two-dimensional intensity plots for three representative probes in the region whose nominal coordinates lie within a segmental duplication. Normalized intensity for the A and B alleles is plotted on the *x*- or the *y*-axis, respectively; definition of A and B alleles is arbitrary. Each point represents a single individual. (D) Unrooted tree based on principal components analysis on sum intensities (see *Materials and Methods*) at the eight probes in the region, for classical inbred strains only. (E) Local phylogeny based on SNPs across the compatible interval (chromosome 5: 91,477,046–96,609,347) spanning the region (Wang *et al.* 2010, 2012).

density by describing sex-specific properties of the recombination landscape that are general across all autosomes and strain backgrounds. Second, we demonstrate that by selecting a unique experimental design that combines a highly informative population with very dense genotyping we are able to infer dynamic properties from static data. Finally, we provide a catalog of cold regions for recombination that reveals a paradoxical inverse relationship between recombination and segmental duplication. This work provides a foundation for testing specific hypotheses regarding the effect of sex and genetics in recombination.

Data Accession

The recombination maps from the G_2:F_1 population—sex specific, sex averaged, strain specific, and strain averaged—will be made publicly available via the Mouse Map Converter tool (http://cgd.jax.org/mousemapconverter/) hosted by the Center for Genome Dynamics (CGD) at the Jackson Laboratory. Raw genotype data in the form of Affymetrix *.cel files will be made available for download on the CGD repository page (http://cgd.jax.org/datasets/diversityarray/CELfiles.shtml). Code to generate and analyze genetic maps is available via GitHub: http://github.com/andrewparkermorgan/cc-g2f1/.

Acknowledgments

We thank Dr. Jan Krumsiek for sharing the source code of Gepard and the mouse-husbandry staff at the University of North Carolina for their ongoing efforts in support of the Collaborative Cross. We thank Leonard McMillan, Catie Welsh, and Chen-Ping Fu (University of North Carolina Department of Computer Science) and John Calaway and John Didion (F. Pardo-Manuel de Villena laboratory) for access to unpublished MegaMUGA genotype data and analysis tools. This project was partially supported by National Institutes of Health grants P50GM076468 (to G.A.C.), R21MH096261 (to F.P.M.d.V.), and T32GM067553 and T32GM008719 (to A.P.M.).

Literature Cited

Anderson, L. K., A Reeves, L. M. Webb, and T. Ashley, 1999 Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. Genetics 151: 1569–1579.

Auton, A., A. Fledel-Alon, S. Pfeifer, O. Venn, L. Ségurel *et al.*, 2012 A fine-scale chimpanzee genetic map from population sequencing. Science 336: 193–198.

Berg, I. L., R. Neumann, K.-W. G. Lam, S. Sarbajna, L. Odenthal-Hesse *et al.*, 2010 PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nat. Genet. 42: 859–863.

Brick, K., F. Smagulova, P. Khil, R. D. Camerini-Otero, and G. V. Petukhova, 2012 Genetic recombination is directed away from functional genomic elements in mice. Nature 485: 642–645.

Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White, and J. L. Weber, 1998 Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am. J. Hum. Genet. 63: 861–869.

Brunschwig, H., L. Levi, E. Ben-David, R. W. Williams, B. Yakir *et al.*, 2012 Fine-scale maps of recombination rates and hotspots in the mouse genome. Genetics 191: 757–764.

Burgoyne, P. S., 1982 Genetic homology and crossing over in the X and Y chromosomes of mammals. Hum. Genet. 61: 85–90.

Chesler, E. J., D. R. Miller, L. R. Branstetter, L. D. Galloway, B. L. Jackson *et al.*, 2008 The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mamm. Genome 19: 382–389.

Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie *et al.*, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36: 1133–1137.

Collaborative Cross Consortium, 2012 The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics 190: 389–401.

Court, D. L., J. A. Sawitzke, and L. C. Thomason, 2002 Genetic engineering using homologous recombination. Annu. Rev. Genet. 36: 361–388.

Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell *et al.*, 2009 A new standard genetic map for the laboratory mouse. Genetics 182: 1335–1344.

Didion, J. P., H. Yang, K. Sheppard, C.-P. Fu, L. McMillan *et al.*, 2012 Discovery of novel variants in genotyping arrays improves genotype retention and reduces ascertainment bias. BMC Genomics 13: 34.

Dietrich, W., H. Katz, S. E. Lincoln, H. S. Shin, J. Friedman *et al.*, 1992 A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131: 423–447.

Dumont, B. L., and B. A. Payseur, 2011 Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet. 7: e1002116.

Dumont, B. L., K. W. Broman, and B. A. Payseur, 2009 Variation in genomic recombination rates among heterogeneous stock mice. Genetics 182: 1345–1349.

Dunn, L. C., 1920 Linkage in mice and rats. Genetics 5: 325–343.

Dunn, L. C., and D. Bennett, 1967 Sex differences in recombination of linked genes in animals. Genet. Res. 9: 211–220.

Egan, C. M., S. Sridhar, M. Wigler, and I. M. Hall, 2007 Recurrent DNA copy number variation in the laboratory mouse. Nat. Genet. 39: 1384–1389.

Feldman, M. W., S. P. Otto, and F. B. Christiansen, 1996 Population genetic perspectives on the evolution of recombination. Annu. Rev. Genet. 30: 261–295.

Flint, J., and E. Eskin, 2012 Genome-wide association studies in mice. Nat. Rev. Genet. 13: 807–817.

Flint, J., and T. F. C. Mackay, 2009 Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. 19: 723–733.

Gerton, J. L., and R. S. Hawley, 2005 Homologous chromosome interactions in meiosis: diversity amidst conservation. Nat. Rev. Genet. 6: 477–487.

Hassold, T., and P. Hunt, 2001 To err (meiotically) is human: the genesis of human aneuploidy. Nat. Rev. Genet. 2: 280–291.

Hinch, A. G., A. Tandon, N. Patterson, Y. Song, N. Rohland *et al.*, 2011 The landscape of recombination in African Americans. Nature 476: 170–175.

Howard-Flanders, P., and L. Theriot, 1966 DNA repair in genetic recombination. Genetics 53: 1137–1150.

Hunter, N., 2003 Synaptonemal complexities and commonalities. Mol. Cell 12: 533–535.

Hussin, J., D. Sinnett, F. Casals, Y. Idaghdour, V. Bruat *et al.*, 2013 Rare allelic forms of PRDM9 associated with childhood leukemogenesis. Genome Res. 23: 419–430.

Karlin, S., and S. F. Altschul, 1990 Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA 87: 2264–2268.

Khil, P. P., F. Smagulova, K. M. Brick, R. D. Camerini-Otero, and G. V. Petukhova, 2012 Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. Genome Res. 22: 957–965.

Koehler, K. E., J. P. Cherry, A. Lynn, P. A. Hunt, and T. J. Hassold, 2002 Genetic control of mammalian meiotic recombination. I. Variation in exchange frequencies among males from inbred mouse strains. Genetics 162: 297–306.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. 31: 241–247.

Kong, A., J. Barnard, D. F. Gudbjartsson, G. Thorleifsson, G. Jonsdottir *et al.*, 2004 Recombination rate and reproductive success in humans. Nat. Genet. 36: 1203–1206.

Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson *et al.*, 2010 Fine-scale recombination rate differences between sexes, populations and individuals. Nature 467: 1099–1103.

Krejci, L., V. Altmannova, M. Spirek, and X. Zhao, 2012 Homologous recombination and its regulation. Nucleic Acids Res. 40: 5795–5818.

Krumsiek, J., R. Arnold, and T. Rattei, 2007 Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23: 1026–1028.

Lander, E. S., and P. Green, 1987 Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA 84: 2363–2367.

Lenormand, T., and J. Dutheil, 2005 Recombination difference between sexes: a role for haploid selection. PLoS Biol. 3: e63.

Liebe, B., G. Petukhova, M. Barchi, M. Bellani, H. Braselmann *et al.*, 2006 Mutations that affect meiosis in male mice influence the dynamics of the mid-preleptotene and bouquet stages. Exp. Cell Res. 312: 3768–3781.

Liu, E. Y., Q. Zhang, L. McMillan, F. P.-M. de Villena, and W. Wang, 2010 Efficient genome ancestry inference in complex pedigrees with inbreeding. Bioinformatics 26: i199–i207.

López-Flores, I., and M. A. Garrido-Ramos, 2012 The repetitive DNA content of eukaryotic genomes. Genome Dyn. 7: 1–28.

McVean, G. T., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley *et al.*, 2004 The fine-scale structure of recombination rate variation in the human genome. Science 304: 581–584.

Mohandas, T. K., R. M. Speed, M. B. Passage, P. H. Yen, C. Chandley *et al.*, 1992 Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. Am. J. Hum. Genet. 51: 526–533.

Morgan, T. H., 1911 Random segregation *vs.* coupling in Mendelian inheritance. Science 34: 384.

Murdoch, B., N. Owen, S. Shirley, S. Crumb, K. W. Broman *et al.*, 2010 Multiple loci contribute to genome-wide recombination levels in male mice. Mamm. Genome 21: 550–555.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321–324.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman *et al.*, 2010 Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327: 876–879.

Niedzwiedz, W., G. Mosedale, M. Johnson, C. Y. Ong, P. Pace *et al.*, 2004 The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. Mol. Cell 15: 607–620.

Otto, S. P., and T. Lenormand, 2002 Resolving the paradox of sex and recombination. Nat. Rev. Genet. 3: 252–261.

Page, J., S. Berríos, M. T. Parra, A. Viera, J. A. Suja *et al.*, 2005 The program of sex chromosome pairing in meiosis is highly conserved across marsupial species: implications for sex chromosome evolution. Genetics 170: 793–799.

Paigen, K., J. P. Szatkiewicz, K. Sawyer, N. Leahy, E. D. Parvanov *et al.*, 2008 The recombinational anatomy of a mouse chromosome. PLoS Genet. 4: e1000119.

Pardo-Manuel de Villena, F., and C. Sapienza, 2001 Recombination is proportional to the number of chromosome arms in mammals. Mamm. Genome 12: 318–322.

Parvanov, E. D., P. M. Petkov, and K. Paigen, 2010 Prdm9 controls activation of mammalian recombination hotspots. Science 327: 835.

Petkov, P. M., K. W. Broman, J. P. Szatkiewicz, and K. Paigen, 2007 Crossover interference underlies sex differences in recombination rates. Trends Genet. 23: 539–542.

Roberts, A., F. Pardo-Manuel de Villena, W. Wang, L. McMillan, and D. W. Threadgill, 2007 The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. Mamm. Genome 18: 473–481.

Rogala, A. R., A. P. Morgan, A. M. Christensen, T. J. Gooch, T. A. Bell *et al.*, 2014 The Collaborative Cross as a resource for modeling human disease: CC011/Unc, a new mouse model for spontaneous colitis. Mamm. Genome DOI: 10.1007/s00335-013-9499-2.

Sambrook, J., and D. W. Russell, 2006 Isolation of high-molecular-weight DNA from mammalian cells using proteinase K and phenol. Cold Spring Harb. Protoc. DOI: 10.1101/pdb.prot4036.

Samollow, P. B., 2004 First-generation linkage map of the gray, short-tailed opossum, *Monodelphis domestica*, reveals genome-wide reduction in female recombination rates. Genetics 166: 307–329.

Sears, D. D., J. H. Hegemann, and P. Hieter, 1992 Meiotic recombination and segregation of human-derived artificial chromosomes in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 89: 5296–5300.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams *et al.*, 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. 4: e395.

Smagulova, F., I. V. Gregoretti, K. Brick, P. Khil, R. D. Camerini-Otero *et al.*, 2011 Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature 472: 375–378.

Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S. Kucherlapati, 1985 Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature 317: 230–234.

Sturtevant, A. H., 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14: 43–59.

Svenson, K. L., D. M. Gatti, W. Valdar, C. E. Welsh, R. Cheng *et al.*, 2012 High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics 190: 437–447.

Valdar, W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman *et al.*, 2006 Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet. 38: 879–887.

Wand, M., and C. Jones, 1995 *Kernel Smoothing*, Ed. 1. CRC Press, Boca Raton, FL.

Wang, J., K. J. Moore, Q. Zhang, F. P.-M. de Villena, W. Wang *et al.*, 2010 Genome-wide compatible SNP intervals and their properties, pp. 43–52 in *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology -BCB '10*, edited by A. Zhang and M. Borodovsky. ACM Press, New York.

Wang, J. R., F. P.-M. de Villena, and L. McMillan, 2012 Comparative analysis and visualization of multiple collinear genomes. BMC Bioinformatics 13(Suppl 3): S13.

Warren, A., A. Chakravarti, C. Wong, S. Slaugenhaupt, S. Halloran *et al.*, 1987 Evidence for reduced recombination on the nondisjoined chromosomes 21 in Down syndrome. Science 237: 652–654.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril *et al.*, 2002 Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.

Welsh, C. E., D. R. Miller, K. F. Manly, J. Wang, L. McMillan *et al.*, 2012 Status and access to the Collaborative Cross population. Mamm. Genome 23: 706–712.

Wojtasz, L., J. M. Cloutier, M. Baumann, K. Daniel, J. Varga *et al.*, 2012 Meiotic DNA double-strand breaks and chromosome asynapsis in mice are monitored by distinct HORMAD2-independent and -dependent mechanisms. Genes Dev. 26: 958–973.

Yalcin, B., K. Wong, A. Agam, M. Goodson, T. M. Keane *et al.*, 2011 Sequence-based characterization of structural variation in the mouse genome. Nature 477: 326–329.

Yang, H., Y. Ding, L. N. Hutchins, J. Szatkiewicz, T. A. Bell *et al.*, 2009 A customized and versatile high-density genotyping array for the mouse. Nat. Methods 6: 663–666.

Yang, H., J. R. Wang, J. P. Didion, R. J. Buus, T. A. Bell *et al.*, 2011 Subspecific origin and haplotype diversity in the laboratory mouse. Nat. Genet. 43: 648–655.

Zenger, K. R., L. M. McKenzie, and D. W. Cooper, 2002 The first comprehensive genetic linkage map of a marsupial: the tammar wallaby (*Macropus eugenii*). Genetics 162: 321–330.

Communicating editor: B. A. Payseur

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161653/-/DC1

High-Resolution Sex-Specific Linkage Maps of the Mouse Reveal Polarized Distribution of Crossovers in Male Germline

Eric Yi Liu, Andrew P. Morgan, Elissa J. Chesler, Wei Wang, Gary A. Churchill, and Fernando Pardo-Manuel de Villena

DOI: 10.1534/genetics.114.161653

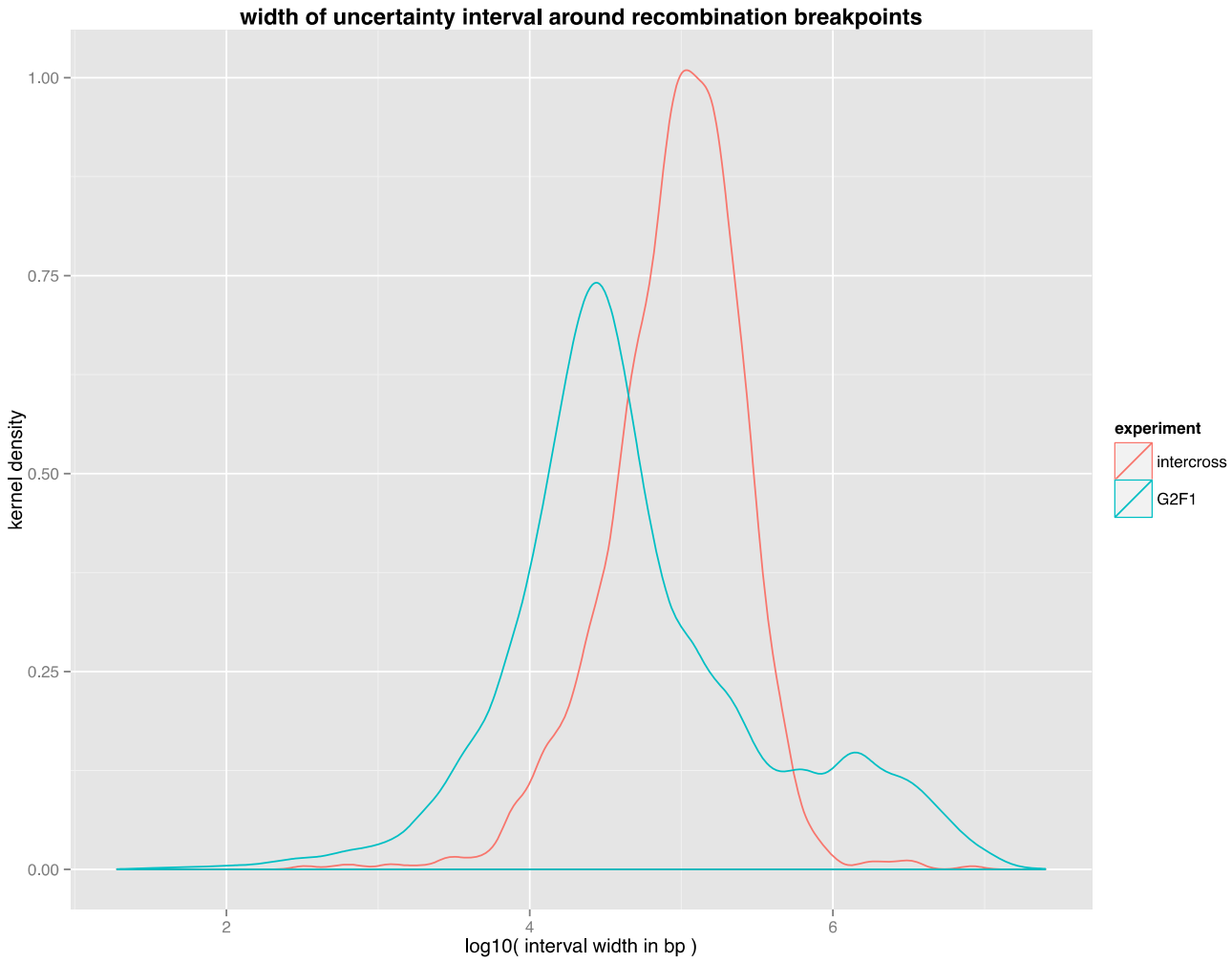
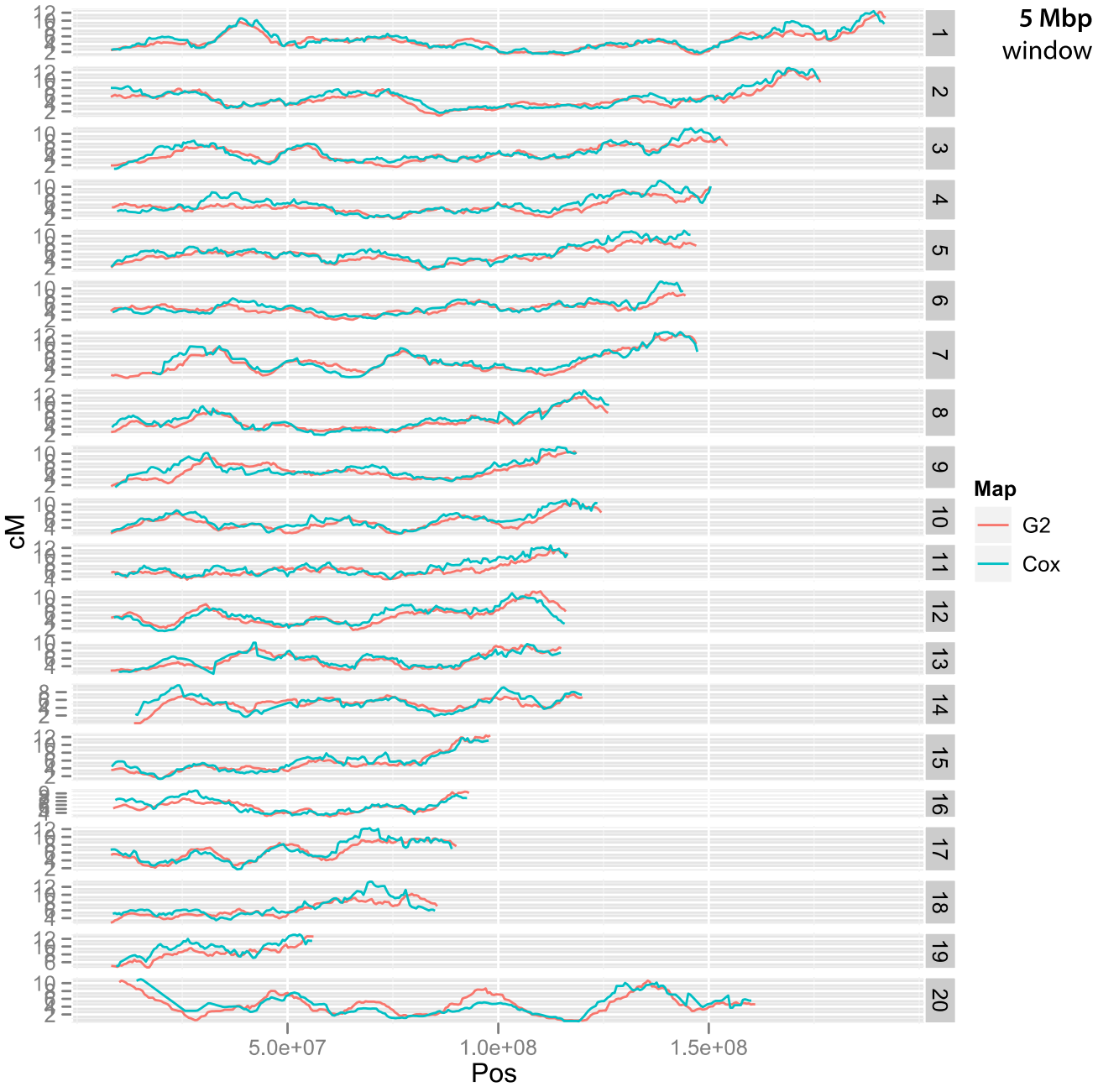


Figure S1 Distribution of width of uncertainty intervals for crossover events in the G_2:F_1 and intercross populations. The average precision in the G_2:F_1 population is approximately 35 kbp, although the distribution is skewed by the presence of a few very long uncertainty intervals.


5 Mbp
window
cM
Pos
5.0e+07
1.0e+08
1.5e+08
Map
G2
Cox

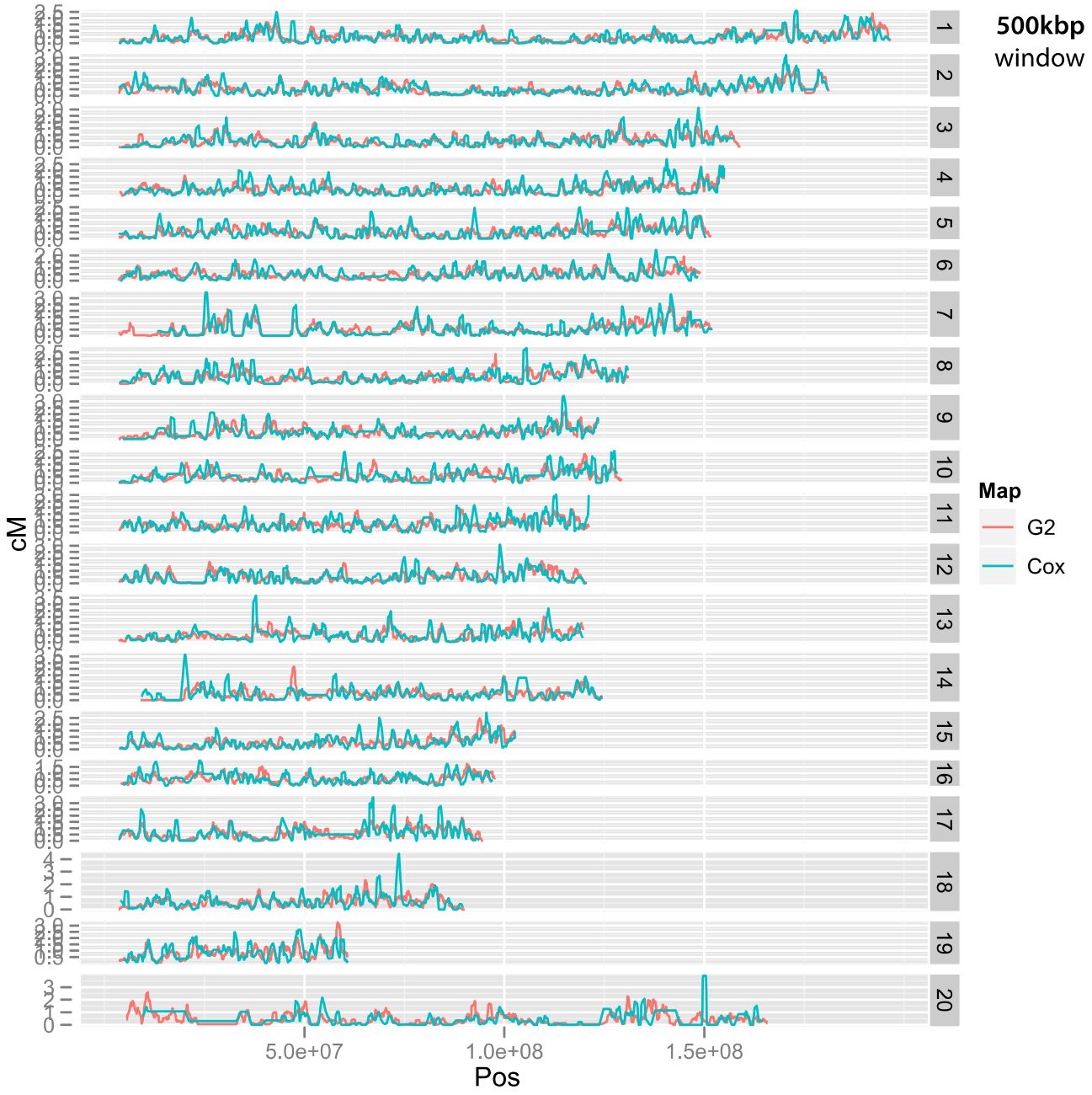


Figure S2 Sex-averaged crossover density in the G_2:F_1 (pink line) and Heterogeneous Stock (Cox *et al.* 2009, from) computed in 5 Mbp (first panel) or 500 kbp (second panel) windows along the genome.

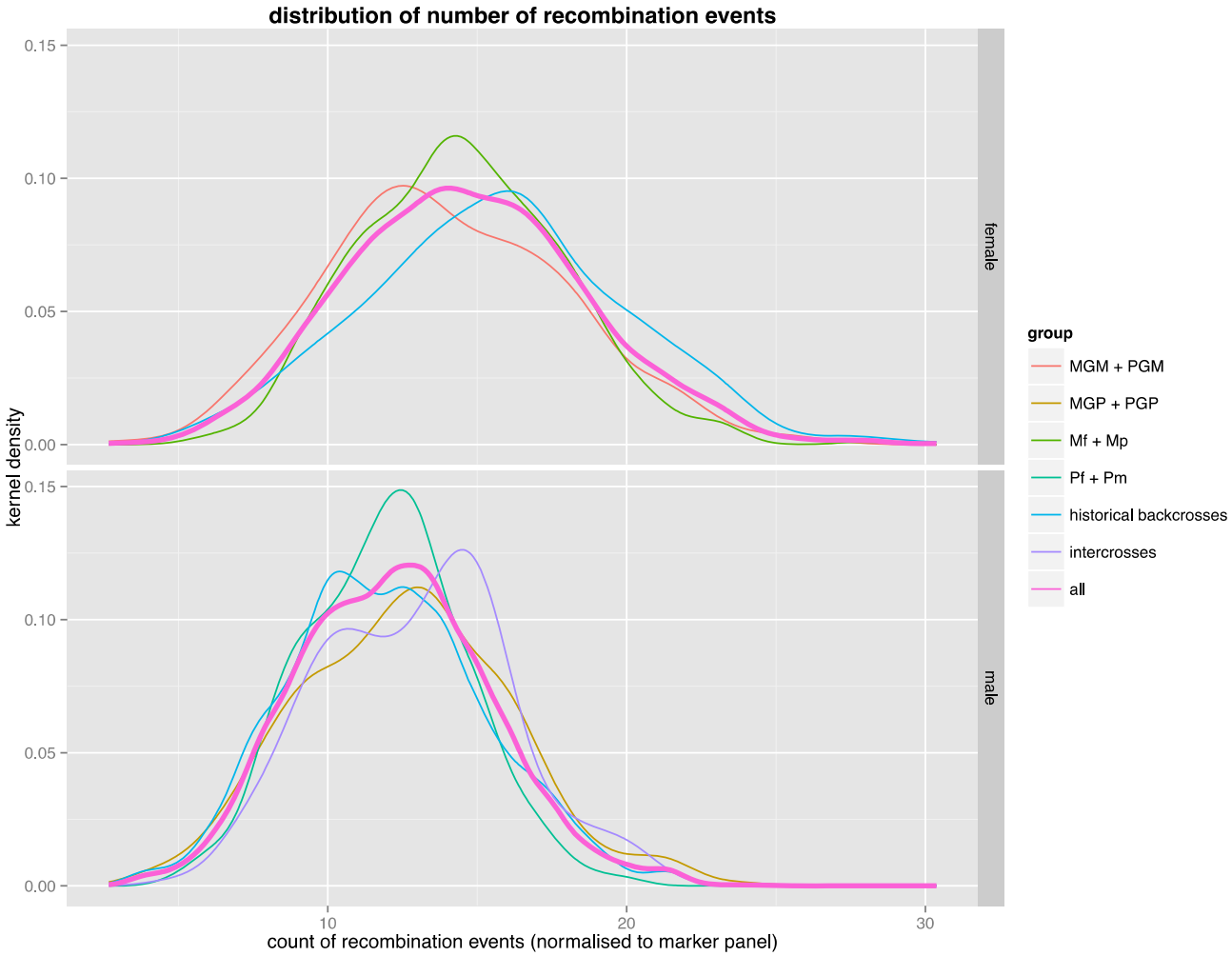
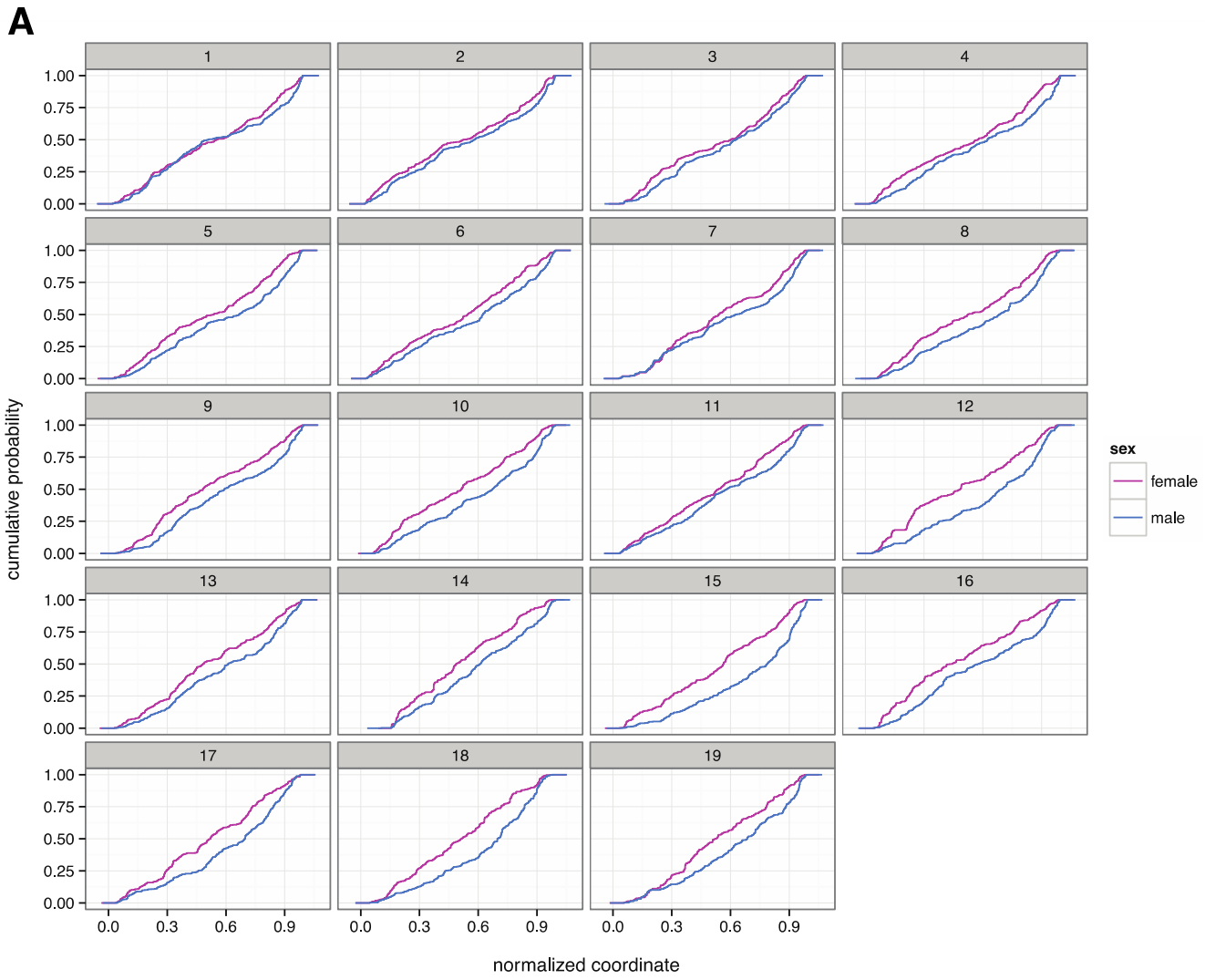


Figure S3 Distribution of the number of crossover events observed in the genome of G_2:F_1 offspring, subdivided by the meiosis and sex of the parent to which the events are ascribed (see Materials & Methods). A similar distribution computed from a panel of historical backcrosses is included for reference.


A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
1.00
0.75
0.50
0.25
0.00
0.0
0.3
0.6
0.9
cumulative probability
normalized coordinate
sex
female
male

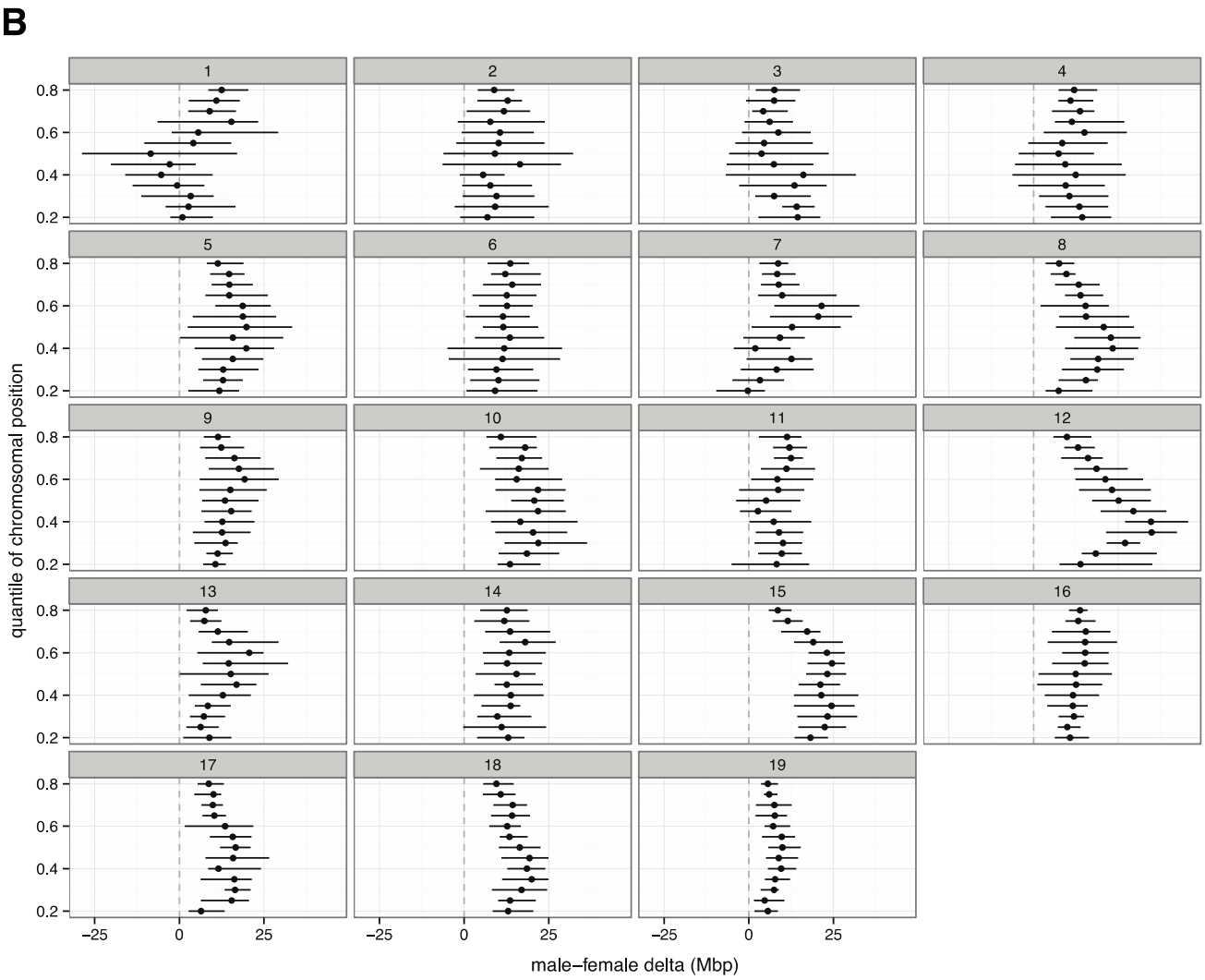


Figure S4 Sex differences in spatial distribution of crossover events. (A) Cumulative distribution of crossovers along scaled, normalized autosomes in the G_2:F_1 population. More than one-third of all autosomal crossover events occur in the distal tenth of chromosomes in males; this effect is absent in females. (B) Location shift of male distribution *vs* female distribution estimated by quantile regression. Intervals represent 95% confidence intervals for displacement of the male distribution at the specified quantiles along the empirical cumulative distribution. The male distribution is displaced towards the distal end at nearly every quantile across all autosomes.

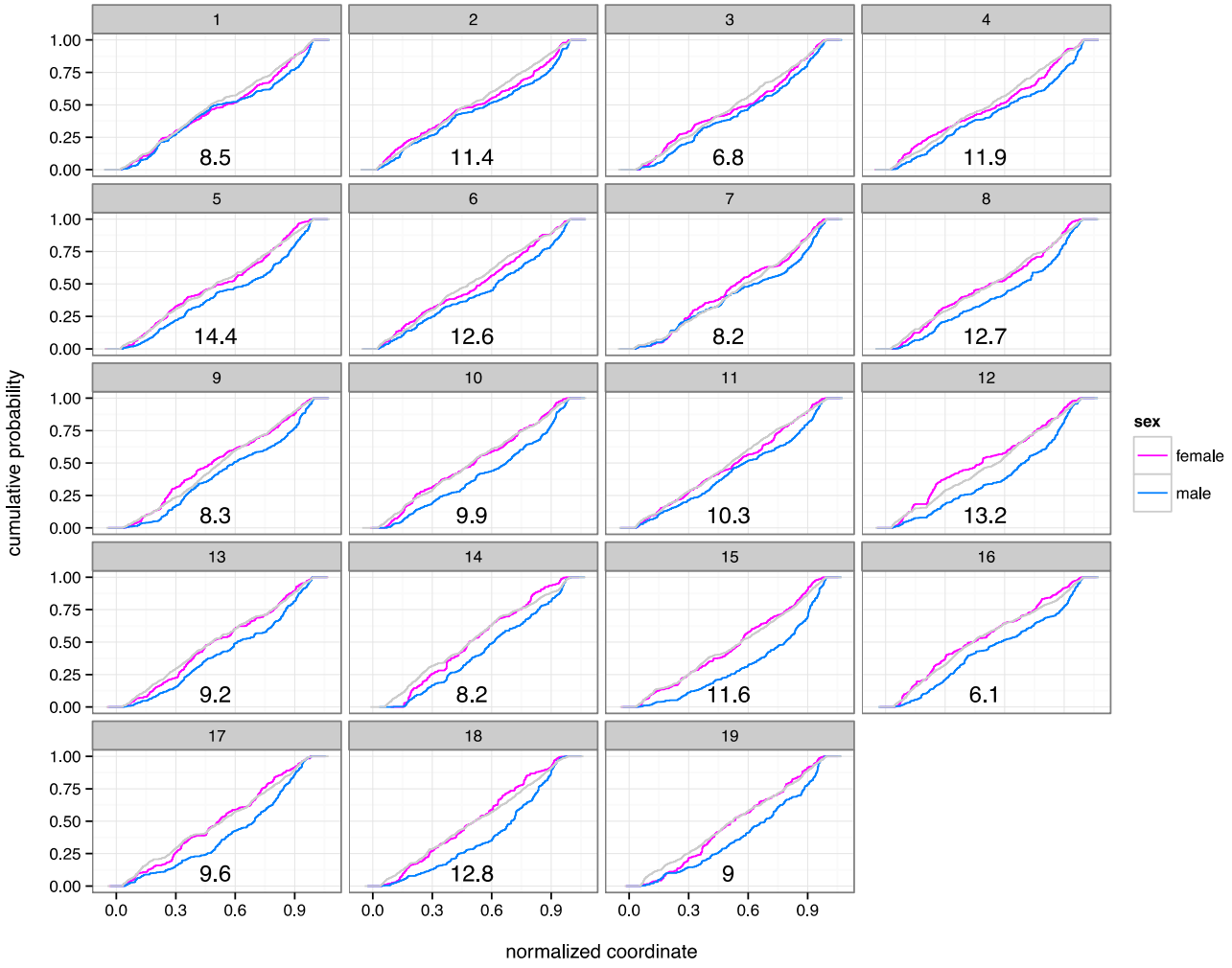


Figure S5 Spatial distribution of crossover events *vs* distribution of double-strand breaks (DSBs) reported by Smagulova *et al.* (2011); Brick *et al.* (2012). Cumulative distribution of crossover events along scaled, normalized chromosomes by sex is shown as in Figure S4. Cumulative distribution of DSBs is shown in grey. Values in each panel give -log10(*p*-values) for a two-tailed Kolmogorov-Smirnov test of difference between male distribution and DSB distribution on each chromosome.

Figure S6
Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161653/-/DC1

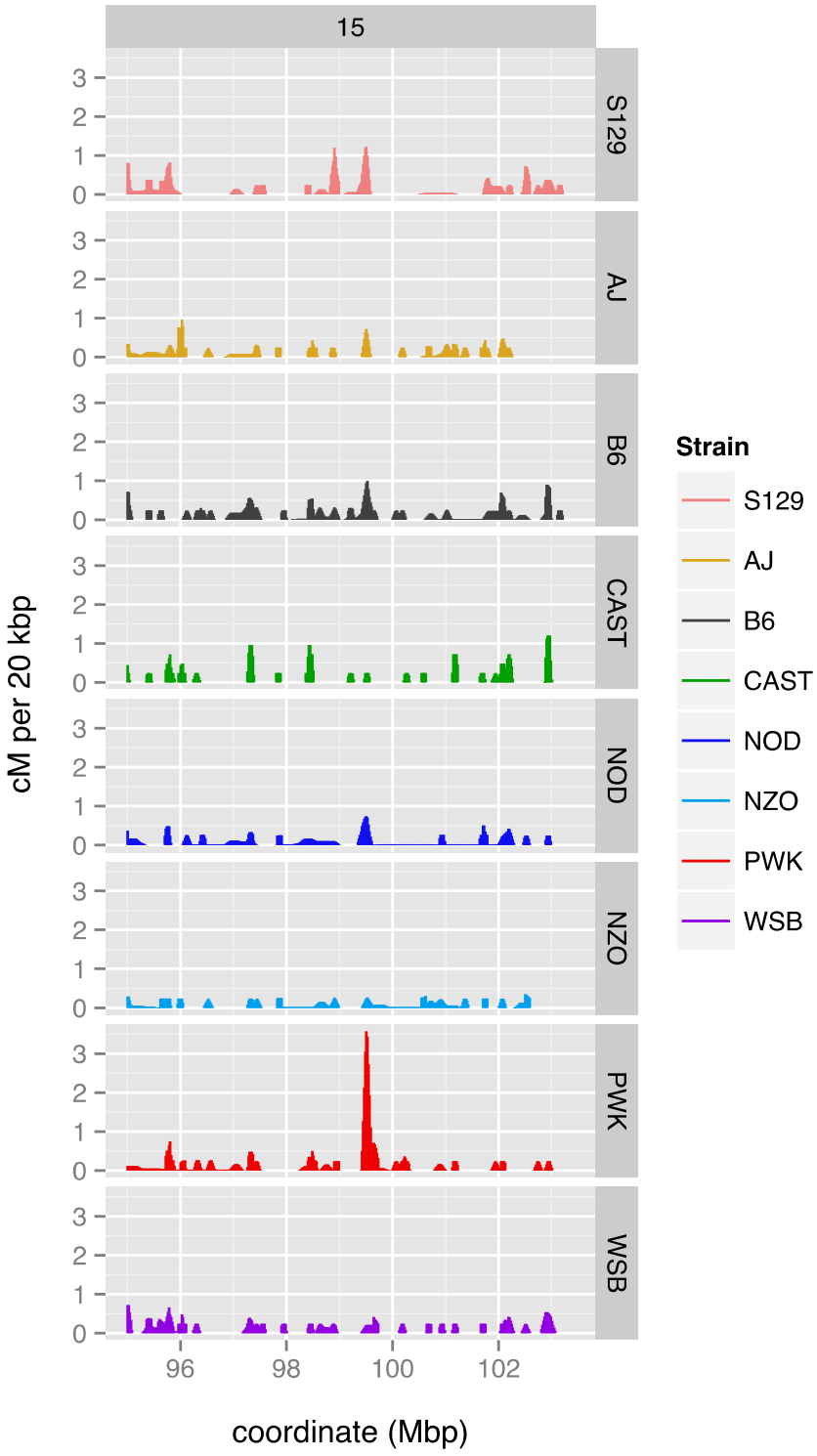


Figure S7 Example of a strain-specific recombination hotspot: a cluster of crossover events on distal chromosome 15 involving PWK/PhJ. The 100-kbp interval at the center of the plot has a genetic length of 6 cM.

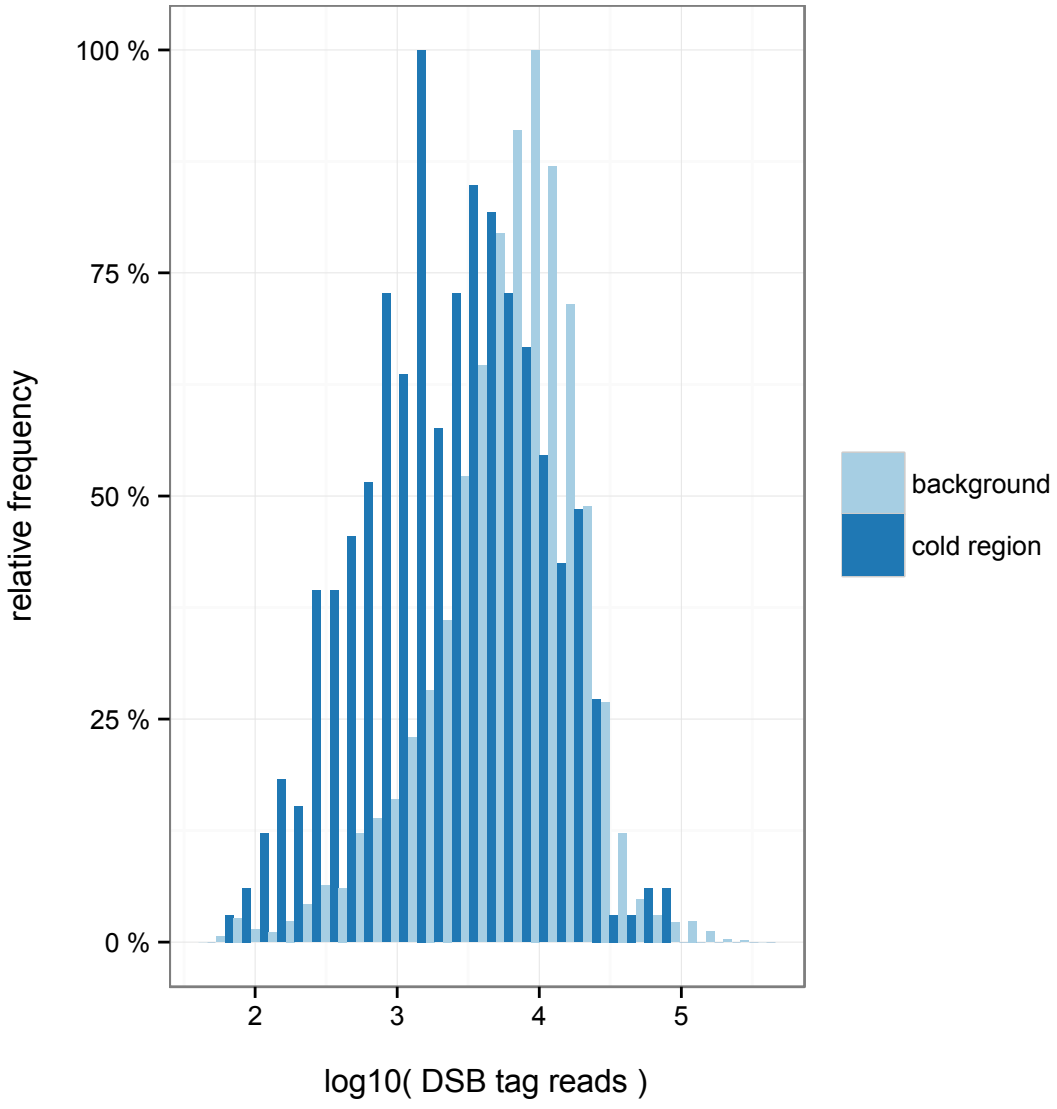


Figure S8 Frequency distribution of double-strand break (DSB) density in Smagulova *et al.* (2011); Brick *et al.* (2012) weighted by strength, measured as log10(number of tag reads) in 500 kbp windows (with 250 kbp overlap). Windows are considered to lie in a cold region if they overlap the region by more than 50% of their length; otherwise they are designated as background.

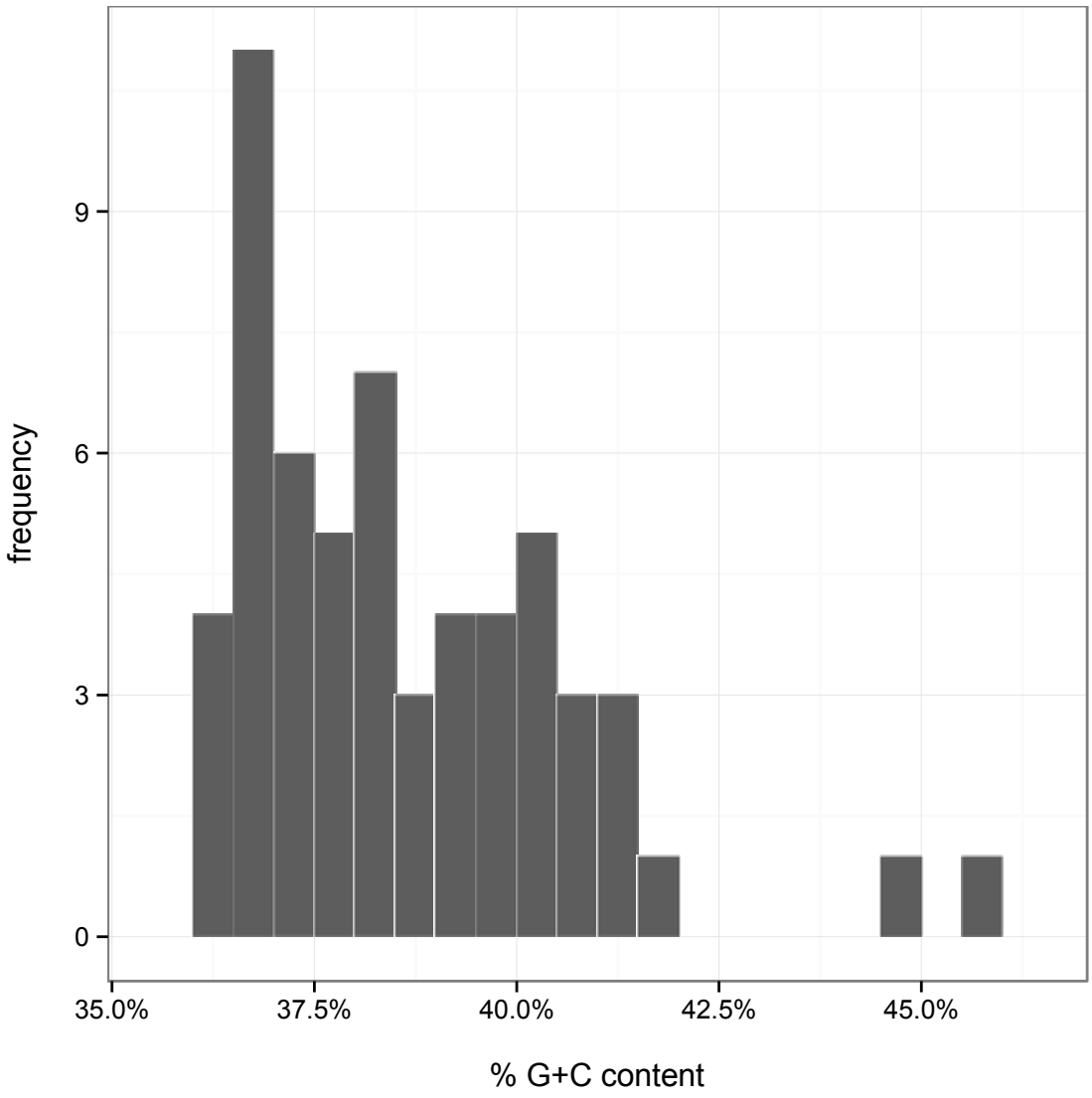


Figure S9 Distribution of C+G content of the 59 recombination-cold regions

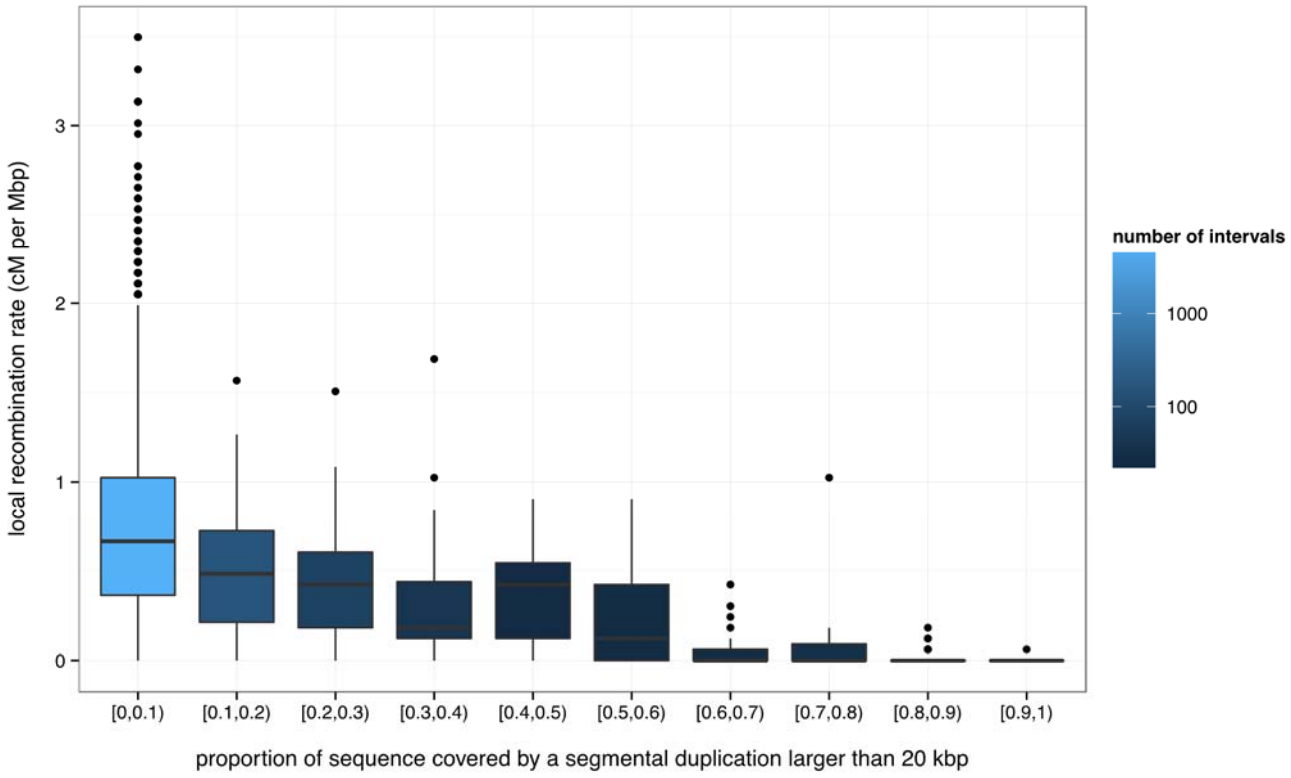


Figure S10 Correlation between segmental duplication rate and local recombination rate computed in 1 Mbp windows genome-wide.

Figure S11 Dotplots of self-alignment of reference sequence (GRCm38/mm10 assembly), as described in Materials & Methods, for all cold regions; one image per cold region, in ZIP-compressed archive. Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161653/-/DC1

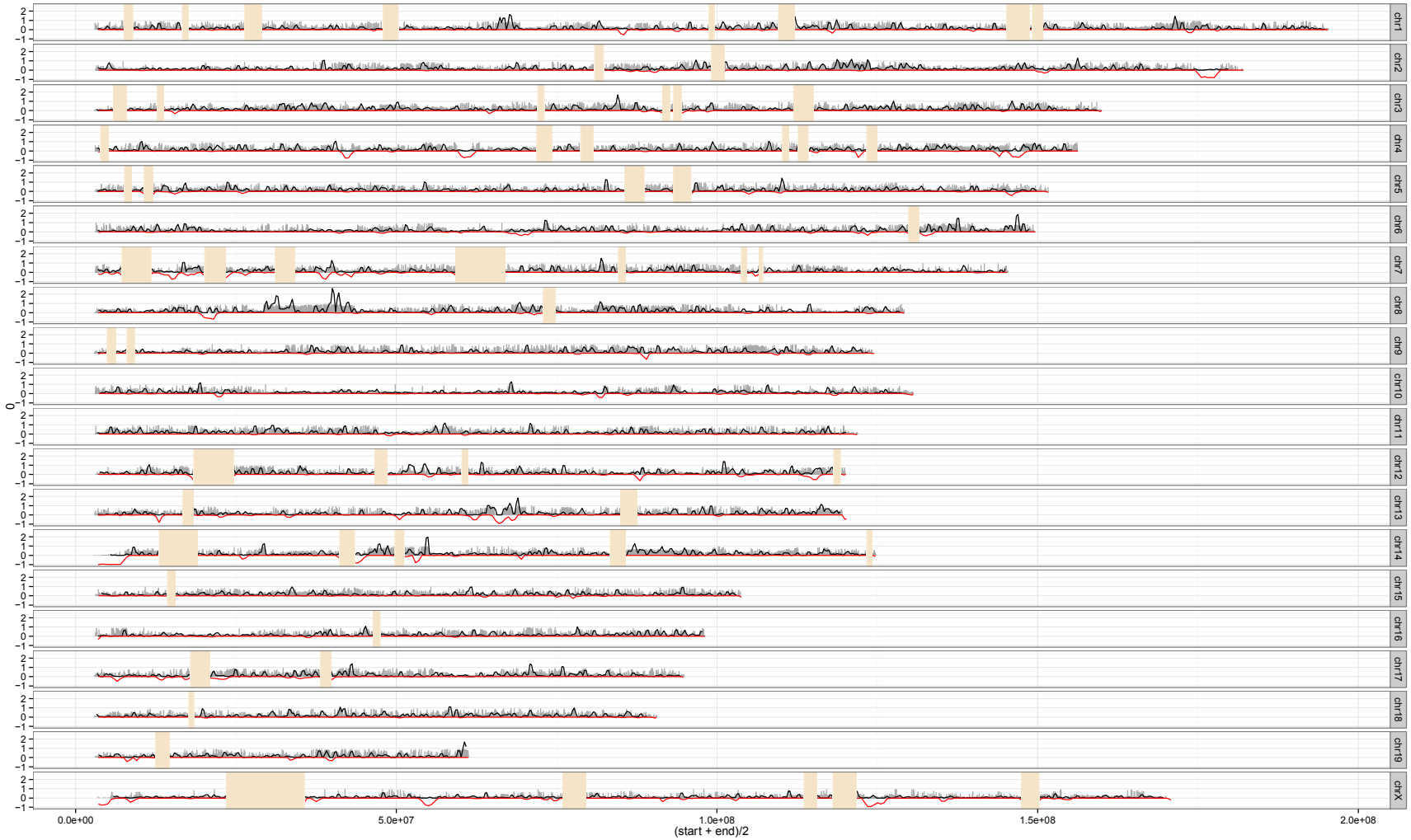


Figure S12 Genome-wide density of structural variants (SVs) in 18 laboratory mouse strains (Yalcin *et al.* 2011) (black line above abscissa) compared to segmental duplication rate (red line below abscissa). Individual SVs are plotted as grey bars with height proportional to the number of strains in which they were observed.

Files S1-S6

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161653/-/DC1

File S1 Crossover events inferred in G_2:F_1 population.

File S2 Crossover events inferred in intercross population.

File S3 Mouse Diversity Array markers used for haplotype inference in G_2:F_1 population, according to quality tier, as ZIP-compressed archive.

File S4 Coordinates and genomic features of cold regions.

File S5 Sequence coordinates of double-strand breaks identified in Smagulova *et al.* (2011); Brick *et al.* (2012), converted to GRCm38/mm10 coordinates, for use in recapitulating analyses of their spatial distribution.

File S6 Fraction of genome spanned by segmental duplications > 20 kbp in size, computed in 1 Mbp bins with 500 kbp overlap.

File S7

Starting at the ends: high-resolution sex-specific linkage maps of the mouse indicate that recombination in male germline is anchored at telomeres

Supplementary Text

1 Mendelian expectations for G_2:F_1 pedigrees

Here we derive, from first principles applied to the Collaborative Cross breeding scheme, a set of strong predictions regarding the expected number and distribution of recombination events observed in G_2:F_1 offspring. Figure 1 from the main text is reproduced for reference as Figure A. Although we restrict our discussion to the autosomes for simplicity, similar logic can be applied to the X chromosome.

1.1 Observable meioses per funnel

A recombination event occurring in one of the eight meioses comprised by each breeding funnel (4 at G_1 and 4 at G_2) can be observed only if its crossover products are transmitted to at least one member of the G_2:F_1 sib pair. Transmission of a G_1 event can occur via one or both of two G_2 meioses: for MGM or MGP events, via Mf or Mm; and for PGM or PGP events, via Pf or Pm. Each G_2 meiosis is independent, so by Mendel's first law the probability that a G_1 event is transmitted to at least one G_2:F_1 sib is $1 - P(\text{transmitted to neither sib}) = 1 - (\frac{1}{2})(\frac{1}{2}) = \frac{3}{4}$.

All (transmitted) crossover products of G_2 meioses are observable as they are transmitted directly to G_2:F_1 offspring. Thus the expected number of observable meioses per funnel is $4(\frac{3}{4}) + 4(1) = 7$.

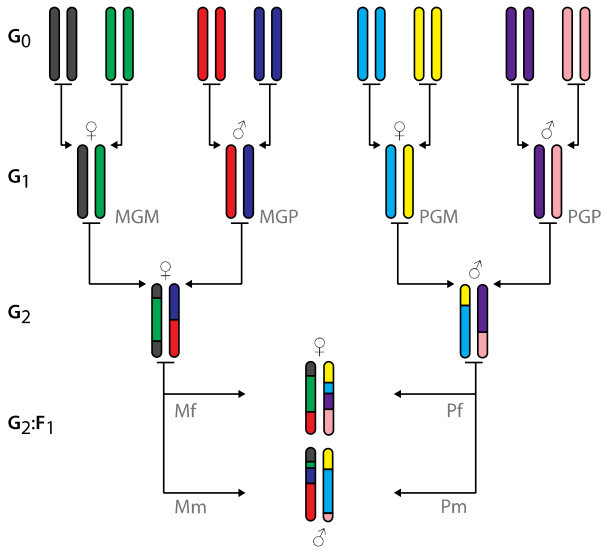


Figure A: Pedigree structure of G_2:F_1 sibling pairs (representative funnel), with each of eight independent meioses labelled.

1.2 G_1 *vs* G_2 events

It follows directly from the logic above that the expected ratio of G_1 to G_2 events is $3 : 4$.

1.3 Balance between half-funnels

Each breeding funnel can be divided into halves, the left (G_2 maternal) half comprising meioses MGM, MGP, Mf and Mm and the right (G_2 paternal) half comprising PGM, PGP, Pf and Pm. Because the halves are balanced for both the number of meioses (4) and the number of meioses per sex (2), the expected ratio of events from the left to the right half-funnel is $1 : 1$.

1.4 Singleton *vs* shared events at G_1

We call an event *shared* if it is observed in both members of a G_2:F_1 sib pair, and a *singleton* if it is observed in only one member of the pair. Each sib is

descended from one paternal G_2 meiosis (Pf or Pm for the female or male sib, respectively) and one maternal G_2 meiosis (Mf or Mm), none of which are shared between sibs. Thus all events assigned to G_2 must be singletons. (The likelihood of observing two independent but indistiguishable recombination events — that is, two events with the same strains proximal and distal to the junction, whose uncertainty intervals share the same boundaries — is negligibly small in our design.)

However, grandpaternal meioses, those occurring at G_1 (MGM, MGP, PGM, PGP), are shared between members of the sib pair. By Mendel's first law, the probability that a G_1 event transmitted to one member of the sib pair is also transmitted to the other is $\frac{1}{2}$. The expected ratio of singleton to shared events in G_1 is thus $2:1$.

2 Algorithm for haplotype reconstruction in intercross population

Intercross offspring — (FVB/NJ × (PWK/PhJ × CAST/EiJ))F_1 or (FVB/NJ × (WSB/EiJ × PWK/PhJ))F_1 — were genotyped using the MegaMUGA Array (Fu, Didion *et al.*, in preparation.) Inference of haplotype blocks from MegaMUGA genotype calls proceeds in three steps, illustrated in Figure B. As in the CC population, we seek to determine the founder ancestry of chromosomal segments, and subsequently to localize recombination events which generate transitions between founders on a given chromosome. The method described here takes advantage of the simplicity of pedigrees in the intercross: for all offspring, the maternal haplotype is always FVB/NJ and the paternal haplotype can be only one of two strains (CAST/EiJ or PWK/PhJ; WSB/EiJ or PWK/PhJ, respectively).

2.1 Founder ancestry assignment of genotypes at each marker

Using genotypes from 32 control samples representing the founder strains in the intercross population (FVB/NJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ), each genotype call from an intercross offspring is assigned to a pair of founder strains: one from the maternal side of the cross and one from the paternal side. (Markers which are not segregating between paternal founder strains

are uninformative for haplotype reconstruction.) Heterozygous calls are resolved to a pair of founder strains by considering segregation patterns between founder strains: for example, if founder strains 1 and 2 have (homozygous) genotype A and founder 3 has genotype B at a given marker, a heterozygous call at that marker is consistent with the founder assignments $(1, 3)$ or $(2, 3)$ but not $(1, 2)$.

2.2 Clustering of founder assignments into haplotype blocks

Founder assignments per marker are clustered into blocks from left to right (*i.e.* proximal to distal) along each chromosome, ignoring uninformative markers. A block is defined as a run of consecutive markers with the same founder assignment; it is described by its left and right endpoints (which may span the entire chromosome) and its founder assignment. Blocks thus define chromosome segments inherited from a given founder strain, and owing to the duality between haplotype structure and the recombination which generates it, boundaries between adjacent blocks define recombination events. The uncertainty interval for a recombination event is the distance between the distal-most marker in the proximal block and the proximal-most marker in the distal block.

2.3 Refinement of haplotype blocks

Two constrains are applied to the resulting haplotypes blocks to improve the inferred haplotypes. First, internal blocks — those for which neither boundary is a chromosome end (such as the one marked with an "X" in Figure 3, last panel) — are required to be longer than 25 Mbp. This is because these blocks are defined by multiple recombination events, which in this population must have occurred during the same meiosis; interference is expected to space such consecutive recombination events widely along the chromosome. Because the 25 Mbp threshold represents approximately the 0.1^{st} percentile of interference distance in our experiment, blocks shorter than this are biologically highly unlikely. We subsume internal blocks smaller than 25 Mbp into neighboring blocks. (Note that, because only two founder strains can be present on the paternal haplotype, it does not matter if the small block is subsumed into its right- or left-hand neighbor.) Second, each block must consist of least 10 markers, in order to mitigate effects of genotyping error

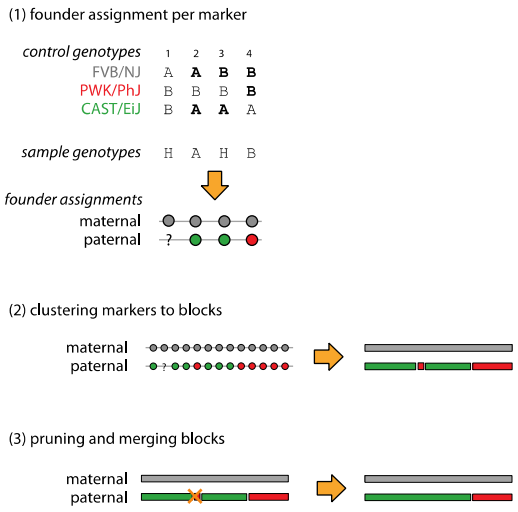


Figure B: Algorithm for haplotype reconstruction in intercross population.

or noncrossovers. The far distal regions are exempted from this constraint in order to capture subtelomeric recombination events, and such events were confirmed as true recombinations by manual inspection of probe intensity data at underlying markers.

In practice this algorithm returns identical results across a wide range of choices for the minimum-interference and minimum-markers-per-block parameters, except for a highly-repetitive, poorly-assembled region of proximal chromosome 14 for which several markers are known to be problematic in heterozygosity (Yang *et al.* 2011; Collaborative Cross Consortium 2012). Likewise, clustering of markers into blocks from right to left instead of from left to right has no effect on results.

References

Collaborative Cross Consortium, 2012 The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics **190**:

389–401.

Yang, H., J. R. Wang, J. P. Didion, R. J. Buus, T. A. Bell, *et al.*, 2011 Subspecific origin and haplotype diversity in the laboratory mouse. Nat Genet **43**: 648–55.