



High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans

Graham Coop, *et al.*
Science **319**, 1395 (2008);
DOI: 10.1126/science.1151851

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 3, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/319/5868/1395>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/1151851/DC1>

This article **cites 31 articles**, 7 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/319/5868/1395#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans

Graham Coop,^{1*} Xiaoquan Wen,¹ Carole Ober,^{1,2} Jonathan K. Pritchard,^{1*} Molly Przeworski^{1*}

Recombination plays a crucial role in meiosis, ensuring the proper segregation of chromosomes. Recent linkage disequilibrium (LD) and sperm-typing studies suggest that recombination rates vary tremendously across the human genome, with most events occurring in narrow “hotspots.” To examine variation in fine-scale recombination patterns among individuals, we used dense, genome-wide single-nucleotide polymorphism data collected in nuclear families to localize crossovers with high spatial resolution. This analysis revealed that overall recombination hotspot usage is similar in males and females, with individual hotspots often active in both sexes. Across the genome, roughly 60% of crossovers occurred in hotspots inferred from LD studies. Notably, however, we found extensive and heritable variation among both males and females in the proportion of crossovers occurring in these hotspots.

Errors in the recombination process during meiosis underlie a variety of chromosomal abnormalities and greatly increase the risk of nondisjunction (1–3). Nonetheless, the total number of recombination events varies significantly among individuals (4), and rates of genetic exchange over fine scales are known to differ in males (5–8). These observations hint at extensive variation in many aspects of the recombination process (9), the nature and extent of which have yet to be systematically characterized. In particular, because observations of

recombination between closely linked markers come from sperm typing, we still know little about fine-scale patterns of recombination in females.

With the recent advent of high-density genotyping platforms, it is now feasible to study fine-scale patterns of recombination with pedigree data. To test this approach, we analyzed genome-wide single-nucleotide polymorphism data from the Affymetrix GeneChip Mapping 500K Array Set (Affymetrix, Santa Clara, CA) in 725 related Hutterites, a population of European descent

(10). The 725 individuals form part of a known, 1650-person, 13-generation pedigree, which we broke down into a set of 82 overlapping nuclear families for purposes of analysis. Of the 82 families, 50 have between four and ten genotyped children, 18 have three, and 14 have two, allowing us to infer recombination events in a total of 364 male and 364 female gametes. Although the number of meioses is smaller than that of Kong *et al.* (11) (728 versus 1257 meioses), our marker density is nearly 100-fold higher, allowing us to define crossover locations with high spatial resolution.

To infer recombination events from genotype data in nuclear families with two or more children, we devised an algorithm that effectively phases the parental chromosomes and identifies positions where a child’s chromosome switches from copying one parental haplotype to the other [Fig. 1A and supporting online material (SOM) text]. This approach identified 24,095 autosomal crossovers in 728 meioses, of which 12,278 were localized to an interval of less than 100 kb and 4,854 to within 30 kb (Fig. 1B). We inferred a mean of 39.6 recombination events per gamete [95% confidence interval (CI):

¹Department of Human Genetics, University of Chicago, 920 East 58th Street, Cummings Life Science Center, Chicago, IL 60637, USA. ²Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: gcoop@uchicago.edu (G.C.); pritch@uchicago.edu (J.K.P.); mfp@uchicago.edu (M.P.)

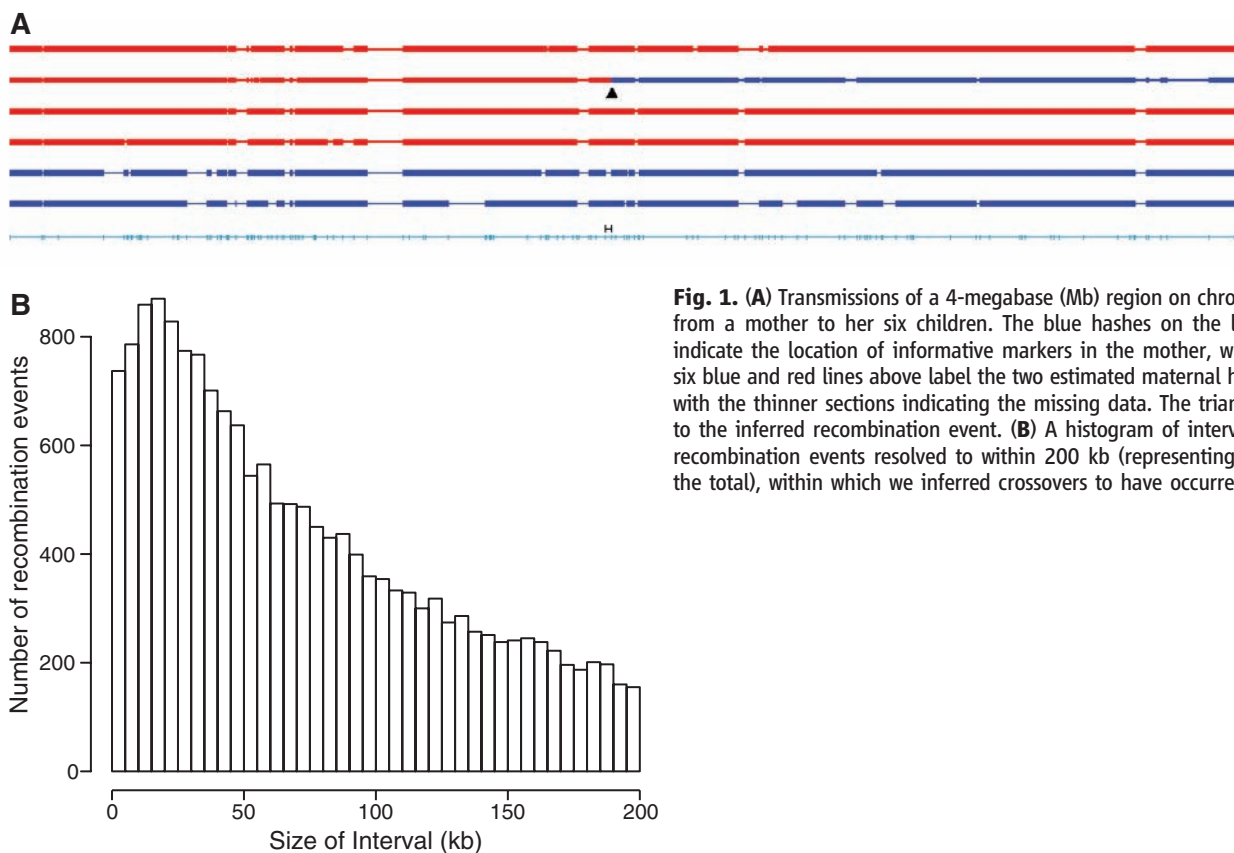


Fig. 1. (A) Transmissions of a 4-megabase (Mb) region on chromosome 5 from a mother to her six children. The blue hashes on the lowest line indicate the location of informative markers in the mother, whereas the six blue and red lines above label the two estimated maternal haplotypes, with the thinner sections indicating the missing data. The triangle points to the inferred recombination event. (B) A histogram of interval sizes of recombination events resolved to within 200 kb (representing ~70% of the total), within which we inferred crossovers to have occurred.

38.5 to 40.6] in females and a mean of 26.2 recombination events per gamete (95% CI: 25.6 to 26.7) in males. Both these estimates, as well as our recombination rate estimates at the megabase scale, agree closely with those of previous studies (11) (SOM text). This agreement implies that our algorithm calls recombination events reliably and that, at least at this scale of comparison, the Hutterite and Icelandic populations have similar overall recombination rates.

We then examined recombination among individuals, confirming the existence of significant variation ($SD = 4.71$, $P = 0.0007$) in the mean number of recombination events among females (11, 12). A previous report (13) had suggested that mothers with higher recombination rates have slightly more offspring and that viable offspring of older mothers tend to have higher recombination rates (13). We saw a similar effect in the Hutterites: We estimated that children born to mothers aged 35 years or older have, on average, an extra 3.1 maternal recombination events as compared with those born to mothers below the age of 35 (one-sided $P = 0.028$, from a stringent within-family permutation test). This maternal age effect may reflect selection against oocytes that have insufficient numbers of recombination events to overcome insults to the meiotic system accumulated over time (13, 14).

Among males, we also found significant variation in the total number of recombination events ($SD = 2.59$, $P = 0.0001$), as reported for cytogenetic studies (15) but, until recently (16), not seen in pedigree studies. Moreover, in males, we detected significant variation in the number of crossovers on individual chromosomes (especially chromosome 19), even after correcting for genome-wide recombination (SOM text). Notably, the chromosomes with significant variation in males show no such evidence in females, suggesting that there may be sex-specific modifiers of recombination rate at this scale (9, 17). Unlike in females, we did not detect an effect of paternal age on recombination.

On a broad scale, recombination rates are known to increase with gene density (11), which is consistent with a link between transcription and recombination, as found in yeast (18). At a finer scale, however, patterns of LD suggest that recombination rates are actually reduced near genes and highest at a short distance from the start positions of genes (19). Because LD patterns are shaped not only by recombination but also by natural selection, the interpretation of this finding is not clear-cut; indeed, the observation of increased LD within genes has been interpreted both as a signal of natural selection in genes (20) and as evidence for reduced recombination (19).

To resolve this issue using directly observed recombination events, we estimated the average recombination rates as a function of distance from the nearest transcription start site (TSS) (Fig. 2 and SOM text). Of the 4854 recombination events that were refined to within 30 kb, we

found that recombination rates are typically low near the TSS (both upstream and downstream) and are highest in regions tens or hundreds of kilobases from the nearest TSS. These results indicate that recombination tends to occur in more-

distant intergenic sites that may be less likely to be associated with promoter function, implying that the primary cause of increased LD near the TSS is decreased recombination rather than selection.

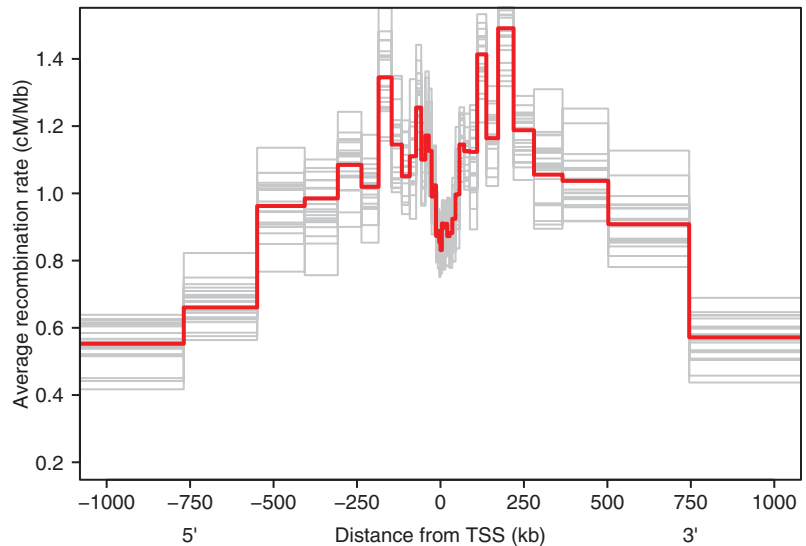


Fig. 2. Distribution of recombination relative to genes. The red line plots the estimated, average recombination rate as a function of distance from the nearest TSS, calculated with recombination events refined to within 30 kb. The physical length of each bin is indicated by the length of the horizontal line. The 20 gray lines show averages calculated from bootstrap resampling of recombination events, as a measure of the uncertainty in our estimates. cM, centimorgan.

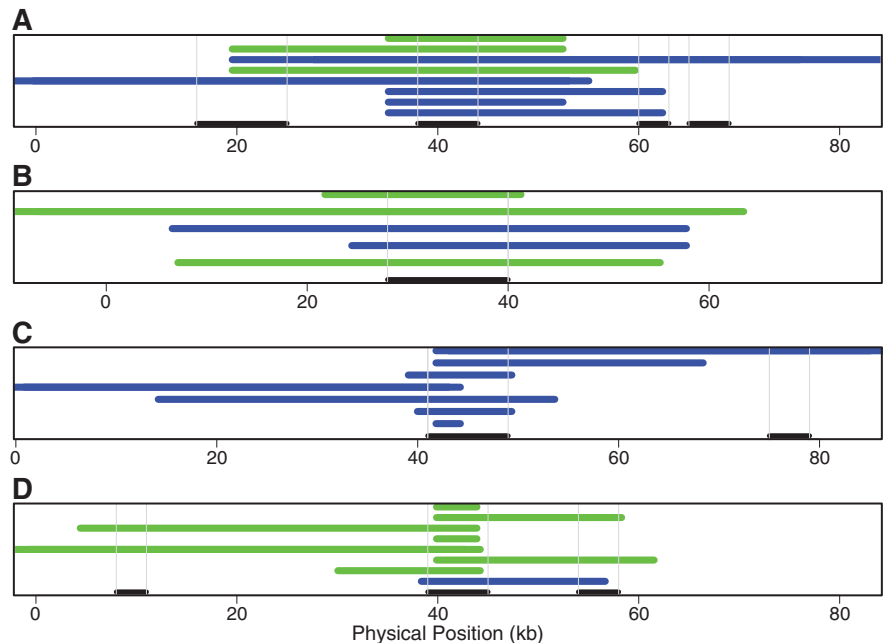


Fig. 3. Overlap of recombination events with four specific hotspots inferred from LD analyses. These regions were chosen because they contain some of the most active hotspots seen over all chromosomes (SOM text). Each panel displays results for one region, with the physical position (in kilobases) denoted on the x axis. Only recombination events localized to within 100 kb are shown, with intervals containing male and female crossovers indicated in green and blue, respectively. The locations of the hotspots estimated from LD data are shown along the bottom of each panel as black lines, and vertical light gray lines indicate their boundaries. (A) Region located at 69.5 kb on chromosome 17. (B) Region located at 58.7 kb on chromosome 19. (C) Region located at 119.5 kb on chromosome 10. (D) Region located at 132.5 kb on chromosome 11.

Sperm-typing and LD analyses suggest that most (60 to 70%) crossover events occur in about 10% of the genome (19), a heterogeneity largely due to 1- to 2-kb hotspot regions that experience sharply elevated recombination relative to that of the background (8, 9, 19, 21–23). Although such studies have vastly improved our knowledge of fine-scale rates, sperm-typing studies are labor-intensive and only informative about male rates. In turn, LD-based estimates rely on a simple population genetic model and are both sex-averaged (over both male and female ancestors of the sample) and time-averaged (over many ancestral generations); consequently, such estimates cannot be used to learn about variation in rates among individuals. In contrast, our high-resolution pedigree data allow us to directly observe crossover events in transmissions from both males and females and to examine inter-individual variation.

To learn more about the nature of hotspots, we considered all recombination events in our data whose location could be inferred to within 30 kb (2910 female and 1944 male events; see SOM for results with other cutoffs). To assess the congruence between LD- and pedigree-based estimates of recombination, we examined how often these well-resolved recombination events overlapped with 32,996 putative hotspots estimated from LD patterns in the Phase II HapMap data (24) (SOM text). We found that 72% of crossovers overlap a hotspot, when just 32% would be expected to do so by chance. We then used a likelihood method to estimate the true proportion of recombination events that takes place in hotspots, accounting for the possibility that an event overlaps simply by chance (SOM text). We found that 60% (95% CI: 58 to 61%) of recombination events occurred in hotspots, which

closely agrees with analyses of LD data (19). A number of the LD hotspots that were overlapped by our inferred recombination events appear to be extremely active: For example, three of the hotspots shown in Fig. 3 are potentially active in as many as 1% of meioses (see SOM text).

Overall, our results support the picture of recombination rate heterogeneity as suggested by LD analyses, notably in terms of the fraction of crossovers occurring in hotspots. This concordance implies that hotspots detected in extant populations have persisted for at least the time scale detectable in LD (i.e., thousands of generations). Our findings do not mean that every inference of a hotspot from LD data is true; our well-resolved recombination events only overlap a total of 3200 hotspots, leaving many predicted hotspots to be confirmed.

At broad scales, females and males are known to differ dramatically in their recombination rates (11, 12), whereas, at finer scales, very little is known about differences between sexes. One hint that recombination rate heterogeneity may be similar between the two sexes is that LD data from the X chromosome—which (outside of the pseudoautosomal regions) recombines only in females—show patterns of hotspots that are roughly similar to those on autosomes (19). Our data show that indeed overall hotspot use is quite similar in the two sexes. Across the genome, the fraction of crossovers that occur in recombination hotspots inferred from LD differs only slightly between males (62%; 95% CI: 59 to 64%) and females (57%; 95% CI: 55 to 59%). Moreover, inspection of specific hotspots revealed that they are often active in both males and females, as they coincide with well-resolved recombination events in both sexes (Fig. 3, A and B). A subset of hotspots, however, seems to be used mainly by

one sex or the other. For example, the hotspot in Fig. 3C is potentially active only in females, whereas the hotspot in Fig. 3D appears to be active mostly in males (SOM text). Our analyses indicated that the sex-specific use of individual hotspots is explained in part by differences in broad (megabase)-scale rates but that there is also considerable variation between sexes below the megabase scale. We also examined the overlap between recombination events within and between sexes, controlling for the broader-scale rate (for details, see the SOM). Together, these analyses suggest that males may use a smaller subset of hotspots than females.

Although we found no marked difference in the average hotspot use between sexes, we noted extensive variation among both males and females in the fraction of crossovers that occur in hotspots. For each parent, averaging across all their offspring, we estimated the genome-wide proportion of events that occur in LD-based hotspots (α) (Fig. 4). The variation in α among individuals is highly significant by a likelihood ratio test (P value from permutation test: $P < 0.002$, for both sexes). Moreover, the narrow-sense heritability of the fraction of crossovers in LD-based hotspots is estimated to be 0.22, which is significantly larger than 0 ($P = 0.01$, with a test that accounts for relatedness across the entire Hutterite pedigree) (25). Thus, genome-wide use of LD-based hotspots is significantly variable among individuals (males and females), and this variation is heritable.

One interpretation of this finding is that some individuals use recombination hotspots less frequently than others. However, because hotspots detected in LD data are likely to have been active for thousands of years, it may be that all individuals use hotspots equally, but some tend to use newer or weaker hotspots that are less likely to be found in analyses of LD. Regardless of the interpretation, the finding of heritable variation in LD-based hotspot use points to heritable differences among individuals in some aspect of the recombination machinery.

This result is particularly interesting in light of recent reports suggesting that hotspot locations have evolved rapidly since the split between humans and chimpanzees (26–29), because differences in trans-acting factors in humans and chimpanzees could account for the marked difference in hotspot locations between the two species. Moreover, our finding offers a possible solution to the hotspot paradox (i.e., the existence of hotspots despite biased gene conversion against alleles that promote them) (6, 7, 30). A single change in the recombination machinery could create many new hotspots in the genome, counteracting the removal of individual hotspots from the population by biased gene conversion (9, 31).

These analyses uncovered tremendous variation in recombination rates over all genomic scales considered and, in particular, heritable variation in hotspot use. It should now be possi-

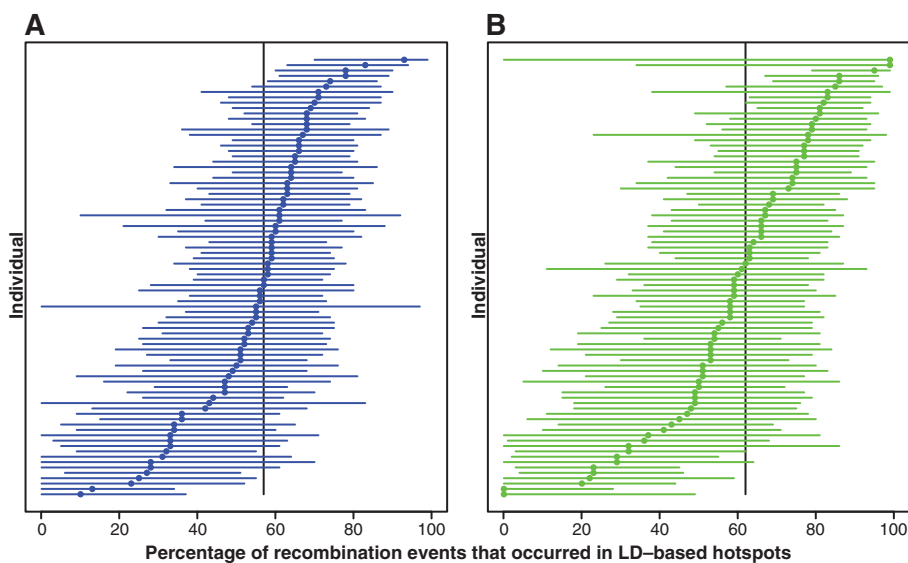


Fig. 4. The percentage of crossovers inferred to have occurred in LD-based hotspots in each individual. The maximum likelihood estimate (MLE) for each individual [females in (A) and males in (B)] is shown as a circle, and the 95% CIs are indicated by the length of the horizontal lines. Individuals are ordered by their MLE. The black vertical line in each panel shows the overall MLE.

ble to map the genetic basis for variation in different aspects of the recombination process, with high-density genotyping data. Identifying the loci that contribute to this variation will offer unparalleled insights into the genetic basis of recombination rate variation and the selective forces governing the evolution of recombination rates (9).

References and Notes

1. T. Hassold, P. Hunt, *Nat. Rev. Genet.* **2**, 280 (2001).
2. K. Inoue, J. R. Lupski, *Annu. Rev. Genomics Hum. Genet.* **3**, 199 (2002).
3. J. A. Bailey, E. E. Eichler, *Nat. Rev. Genet.* **7**, 552 (2006).
4. A. Lynn, T. Ashley, T. Hassold, *Annu. Rev. Genomics Hum. Genet.* **5**, 317 (2004).
5. J. Yu et al., *Am. J. Hum. Genet.* **59**, 1186 (1996).
6. A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267 (2002).
7. A. J. Jeffreys, R. Neumann, *Hum. Mol. Genet.* **14**, 2277 (2005).
8. I. Tiemann-Boege, P. Calabrese, D. M. Cochran, R. Sokol, N. Arnheim, *PLoS Genet.* **2**, e70 (2006).
9. G. Coop, M. Przeworski, *Nat. Rev. Genet.* **8**, 23 (2007).

10. C. Ober, M. Abney, M. S. McPeck, *Am. J. Hum. Genet.* **69**, 1068 (2001).
11. A. Kong et al., *Nat. Genet.* **31**, 241 (2002).
12. K. W. Broman, J. C. Murray, V. C. Sheffield, R. L. White, J. L. Weber, *Am. J. Hum. Genet.* **63**, 861 (1998).
13. A. Kong et al., *Nat. Genet.* **36**, 1203 (2004).
14. R. U. Vallente, E. Y. Cheng, T. J. Hassold, *Chromosoma* **115**, 241 (2006).
15. T. Hassold et al., *Cytogenet. Genome Res.* **107**, 249 (2004).
16. V. G. Cheung, J. T. Burdick, D. Hirschmann, M. Morley, *Am. J. Hum. Genet.* **80**, 526 (2007).
17. T. Lenormand, *Genetics* **163**, 811 (2003).
18. T. D. Petes, *Nat. Rev. Genet.* **2**, 360 (2001).
19. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, *Science* **310**, 321 (2005).
20. D. A. Hinds et al., *Science* **307**, 1072 (2005).
21. L. Kauppi, A. J. Jeffreys, S. Keeney, *Nat. Rev. Genet.* **5**, 413 (2004).
22. D. C. Crawford et al., *Nat. Genet.* **36**, 700 (2004).
23. G. A. T. McVean et al., *Science* **304**, 581 (2004).
24. International HapMap Consortium, *Nature* **449**, 851 (2007).
25. M. Abney, M. S. McPeck, C. Ober, *Am. J. Hum. Genet.* **66**, 629 (2000).

26. J. D. Wall, L. A. Frisse, R. R. Hudson, A. Di Rienzo, *Am. J. Hum. Genet.* **73**, 1330 (2003).
27. S. E. Ptak et al., *PLoS Biol.* **2**, e155 (2004).
28. S. E. Ptak et al., *Nat. Genet.* **37**, 429 (2005).
29. W. Winckler et al., *Science* **308**, 107 (2005).
30. A. Boulton, R. S. Myers, R. J. Redfield, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8058 (1997).
31. G. Coop, S. R. Myers, *PLoS Genet.* **3**, e35 (2007).
32. We thank S. Myers and members of the Pritchard, Przeworski, and Stephens labs for discussion. This work was supported by NIH grants GM83098 to M.P.; HG002772 to J.K.P.; and HD21244, HL56399, and HL56533 to C.O.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1151851/DC1
SOM Text
Figs. S1 to S10
Tables S1 to S9
References

17 October 2007; accepted 23 January 2008
Published online 31 January 2008;
10.1126/science.1151851
Include this information when citing this paper.

Sequence Variants in the *RNF212* Gene Associate with Genome-Wide Recombination Rate

Augustine Kong,* Gudmar Thorleifsson, Hreinn Stefansson, Gisli Masson, Agnar Helgason, Daniel F. Gudbjartsson, Gudrun M. Jonsdottir, Sigurjon A. Gudjonsson, Sverrir Sverrisson, Theodora Thorlacius, Aslaug Jonasdottir, Gudmundur A. Hardarson, Stefan T. Palsson, Michael L. Frigge, Jeffrey R. Gulcher, Unnur Thorsteinsdottir, Kari Stefansson*

The genome-wide recombination rate varies between individuals, but the mechanism controlling this variation in humans has remained elusive. A genome-wide search identified sequence variants in the 4p16.3 region correlated with recombination rate in both males and females. These variants are located in the *RNF212* gene, a putative ortholog of the *ZHP-3* gene that is essential for recombinations and chiasma formation in *Caenorhabditis elegans*. It is noteworthy that the haplotype formed by two single-nucleotide polymorphisms (SNPs) associated with the highest recombination rate in males is associated with a low recombination rate in females. Consequently, if the frequency of the haplotype changes, the average recombination rate will increase for one sex and decrease for the other, but the sex-averaged recombination rate of the population can stay relatively constant.

Recombination generates part of the diversity that fuels evolution. In humans, it has been suggested that recombination rate must be highly regulated (1), as too little recombination can lead to inaccurate disjunction and aneuploidy (2, 3), whereas ectopic exchange can lead to chromosomal rearrangements (4). Some regions in the genome, known as hotspots, have much higher recombination rate per physical distance unit than the genome as a whole. By using high-density single-nucleotide polymorphism (SNP) data, from which historical recom-

bination events can be inferred, and sperm data, substantial advances have been made in the understanding of local recombination rate (5–11). Furthermore, male and female recombination patterns are different in both genome-wide and regional recombination rates (12, 13). It is also firmly established that genome-wide recombination rate varies substantially among women (12, 13), and there have been hints that this also is true in men (14–16).

Previously, we genotyped a large number of families with a genome-wide microsatellite set of ~1000 markers. This work allowed us to estimate the recombination rate for thousands of men and women and demonstrated that maternal recombination rate increases with the age of the mother

and that there is a positive correlation between the number of children and the recombination rate of a woman (17). A common inversion on chromosome 17q21.31 was also identified that associates with recombination rate and fertility of women (18). Here, we performed a genome-wide scan for variants associated with recombination rate by genotyping with the Illumina Hap300 chip 1887 males and 1702 females with recombination rate estimates [see (19) for a description of study groups]. After quality filtering, 309,241 SNPs were tested for association with recombination frequencies. Male and female recombination rates were studied separately with weighted regression where the weight of a person was proportional to the number of children used to estimate recombination rate. We fitted an additive model with the estimated recombination rate regressed on the number of an allele (0, 1, or 2) a person carried. The results were then adjusted for relatedness between individuals and potential population stratification with the method of genomic control (20). Specifically, standard errors of the effect estimates resulting from the regressions were multiplied by a factor of 1.041 for males and 1.067 for females corresponding to dividing the chi-square test statistics by an adjustment factor of $1.084 = 1.041^2$ and $1.138 = 1.067^2$ [see (19) for quality control and statistical analysis].

For the recombination rate of males, three SNPs achieved genome-wide significance ($P < 1.6 \times 10^{-7}$, fig. S1). They were rs3796619 ($P = 1.1 \times 10^{-14}$), rs1670533 ($P = 1.8 \times 10^{-11}$), and rs2045065 ($P = 1.6 \times 10^{-11}$), which were all located within a small region in strong linkage disequilibrium (LD) on chromosome 4p16.3 (Fig. 1). The same three SNPs were also associated with the female recombination rate (Table 1). The last two SNPs achieved genome-wide significance; no

deCODE Genetics Inc, 101 Reykjavik, Iceland.

*To whom correspondence should be addressed. E-mail: kong@decode.is (A.K.); kstefans@decode.is (K.S.)