

Fine-scale recombination rate differences between sexes, populations and individuals

Augustine Kong¹, Gudmar Thorleifsson¹, Daniel F. Gudbjartsson¹, Gisli Masson¹, Asgeir Sigurdsson¹, Aslaug Jonasdottir¹, G. Bragi Walters¹, Adalbjorg Jonasdottir¹, Arnaldur Gylfason¹, Kari Th. Kristinsson¹, Sigurjon A. Gudjonsson¹, Michael L. Frigge¹, Agnar Helgason^{1,2}, Unnur Thorsteinsdottir^{1,3} & Kari Stefansson^{1,3}

Meiotic recombinations contribute to genetic diversity by yielding new combinations of alleles. Recently, high-resolution recombination maps were inferred from high-density single-nucleotide polymorphism (SNP) data using linkage disequilibrium (LD) patterns that capture historical recombination events^{1,2}. The use of these maps has been demonstrated by the identification of recombination hotspots² and associated motifs³, and the discovery that the *PRDM9* gene affects the proportion of recombinations occurring at hotspots^{4–6}. However, these maps provide no information about individual or sex differences. Moreover, locus-specific demographic factors like natural selection⁷ can bias LD-based estimates of recombination rate. Existing genetic maps based on family data avoid these shortcomings⁸, but their resolution is limited by relatively few meioses and a low density of markers. Here we used genome-wide SNP data from 15,257 parent–offspring pairs to construct the first recombination maps based on directly observed recombinations with a resolution that is effective down to 10 kilobases (kb). Comparing male and female maps reveals that about 15% of hotspots in one sex are specific to that sex. Although male recombinations result in more shuffling of exons within genes, female recombinations generate more new combinations of nearby genes. We discover novel associations between recombination characteristics of individuals and variants in the *PRDM9* gene and we identify new recombination hotspots. Comparisons of our maps with two LD-based maps inferred from data of HapMap populations of Utah residents with ancestry from northern and western Europe (CEU) and Yoruba in Ibadan, Nigeria (YRI) reveal population differences previously masked by noise and map differences at regions previously described as targets of natural selection.

To perform a large, family-based recombination study, one challenge is to phase the genotypes of the parents when the grandparents are not genotyped. One solution is to use genotyped nuclear families with two or more offspring, which in essence uses the children to phase the parents. However, resolution can be diminished and difficulties can arise when two or more offspring have recombinations that are close to each other. We capitalized on recent methodological advances that led to the successful determination of parental origins of over 97% of the heterozygous genotypes of 38,167 Icelanders typed on Illumina SNP arrays, many of them with ungenotyped parents^{9,10}. Parental origins provide phase. We used phased haplotypes of 8,850 mother–offspring pairs (6,041 distinct mothers) and 6,407 father–offspring pairs (4,389 distinct fathers) to identify recombinations (Fig. 1) for 15,257 meioses (Supplementary Table 1).

Recombinations were determined using 289,658 and 8,411 SNPs on the autosomal and X chromosomes respectively. The data only allowed us to assign a recombination to the region spanned by the two closest flanking heterozygous markers in the parent (Fig. 1). Treating this as a missing-data problem, the EM algorithm¹¹ was used to calculate likelihood-based estimates of recombination rates for males and females (Supplementary Information and Supplementary Table 2). Also, results from the E-step of the EM algorithm were used to calculate the estimated recombination count in each marker interval for each meiosis. In addition to genetic distances between SNPs, maps for various uniformly spaced grids were calculated by linear interpolation.

Existing genetic maps include the 2002 deCode family-based map⁸ and the most commonly used LD-based maps¹² (Methods Summary), referred to here as the CEU, YRI and COMBINED maps. The COMBINED map is essentially the average of the CEU and YRI maps.

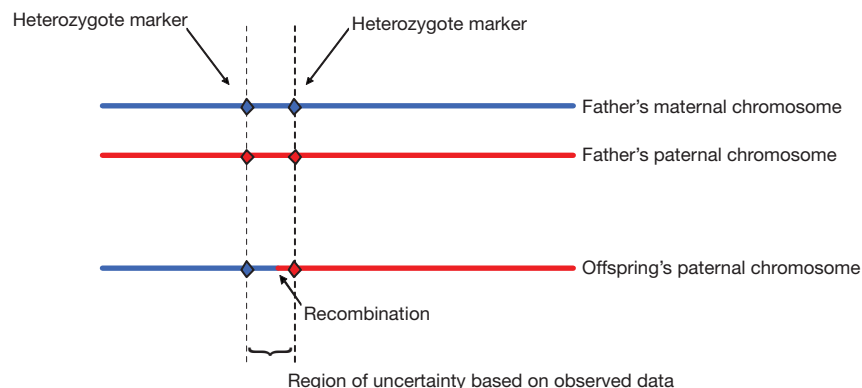


Figure 1 | Determining recombination locations. Here it is assumed that genotypes of parent and offspring have been phased, with parental origin determined^{9,10}. The parent shown is a father, but the same method applies to a mother–offspring pair. For an SNP that is heterozygous for the parent, it can be determined whether the allele passed on to the offspring is from the parent's

maternal or paternal chromosome. The location of the recombination can hence be localized to the region spanned by the two closest flanking heterozygous markers in the parent. (Details are in Supplementary Information.)

¹deCODE genetics, Sturlugata 8, 101 Reykjavik, Iceland. ²Department of Anthropology, University of Iceland, Sæmundargötu 2, 101 Reykjavik, Iceland. ³Faculty of Medicine, University of Iceland, Sæmundargötu 2, 101 Reykjavik, Iceland.

These maps have similar lengths, because the 2002 deCode map was used to scale the other maps which only provide information about relative recombination rates. By comparison, our newly constructed sex-averaged map is 3% shorter. This is probably because we tabulated only recombinations considered highly reliable and some recombinations were missed. Also, the 2002 deCode map could be slightly inflated because of genotyping errors. Supporting the assumption that the dropped recombinations are approximately randomly distributed and have minimal impact on the relative recombination rate estimates, the correlation between the sex-averaged map and the 2002 deCode Map is 0.945 at 3-megabase (Mb) resolution, roughly the limit of resolution for the older map. This correlation is stronger than that between the 2002 deCode map and the LD-based maps ($r = 0.920$, 0.914 and 0.927 for the CEU, YRI and COMBINED maps, respectively). Correlation between the sex-averaged map and the COMBINED map is 0.977 (Supplementary Table 3).

Recombination maps partitioned into 10-kb bins were calculated for each sex. For subsequent investigations, we excluded the X chromosome and 5-Mb regions at the ends of autosomal chromosomes relative to the SNP coverage, locations where the determination of recombinations is less reliable. We also excluded 10,254 bins covering unsequenced regions (Human Map build 36), of which 8,891 were centromeric. These bins generally have low recombination rates and a fraction of them include intervals without recombination rates assigned by the COMBINED-map. Genetic distances of those bins are clearly biased downwards in all three LD-based maps. In total, the studied regions covered 2,444.46 Mb or 244,446 bins. For these bins, the estimated average genetic distance is 0.0155 cM (sum = 3,790.1 cM) for females and 0.0077 cM (sum = 1886.7 cM) for males. At this resolution, the correlation between the male and female maps is 0.659.

A standardized recombination rate (SRR) was calculated for males and females separately, by dividing the genetic distance of each bin by the overall average. Defining recombination hotspots as those bins with an SRR greater than 10, we observed 4,762 hotspots for males and 4,129 hotspots for females, with an overlap of 1,953. The male hotspot bins covered 1.9% of the physical distance of the studied region but accounted for 36.2% of its recombinations. Corresponding numbers for females were 1.7% and 28.0%, respectively. Despite similarities of sexes, 718 and 125 of the 4,762 male hotspots have an SRR less than 3 and 1, respectively, in females. A permutation test (Supplementary Information) showed that these male-specific hotspots have a false-discovery rate of approximately 1.9% and 0%, respectively. Thus approximately 704 of the 718, and all of the 125, identified bins correspond to true sex differences, indicating that about 14.8% (704/4,762) of the male hotspots are sex specific. Correspondingly, of the 4,129 female hotspots, 624 (false-discovery rate 2.8%) and 166 (false-discovery rate 0.7%) have an SRR smaller than 3 and 1, respectively, in males. About 14.7% (606/4,129) of female hotspots are sex specific.

Sex-specific hotspots tend to occur in clusters. Fig. 2a shows a region harbouring the *Basonuclin-2* gene¹³ where recombinations are dominated by those resulting from male meioses. This region contains five male-specific hotspots, the two most striking being at 16.649 Mb (male SRR = 29.1) and 16.829 Mb (male SRR = 27.7). However, even though the female SRR is substantially smaller for these two bins (0.5 and 2.6, respectively), they do correspond to local peaks. This is typical for other male-specific hotspots. The same trend applies to regions where recombinations are dominated by females: that is, local peaks for male recombination rate at female-specific hotspots (Supplementary Fig. 1). Thus, even though hotspots are defined for narrow intervals (noting that the 10-kb resolution hotspots examined here could often be driven by intervals much shorter in length), they are determined by interactions between factors both local and regional, the latter concerning regions that are hundreds of kilobases to many megabases in length (Fig. 2b). If the local factors, but not the regional ones, are supportive of recombination, a local peak that is not a hotspot would result. Moreover, the regional forces influencing male and females are only

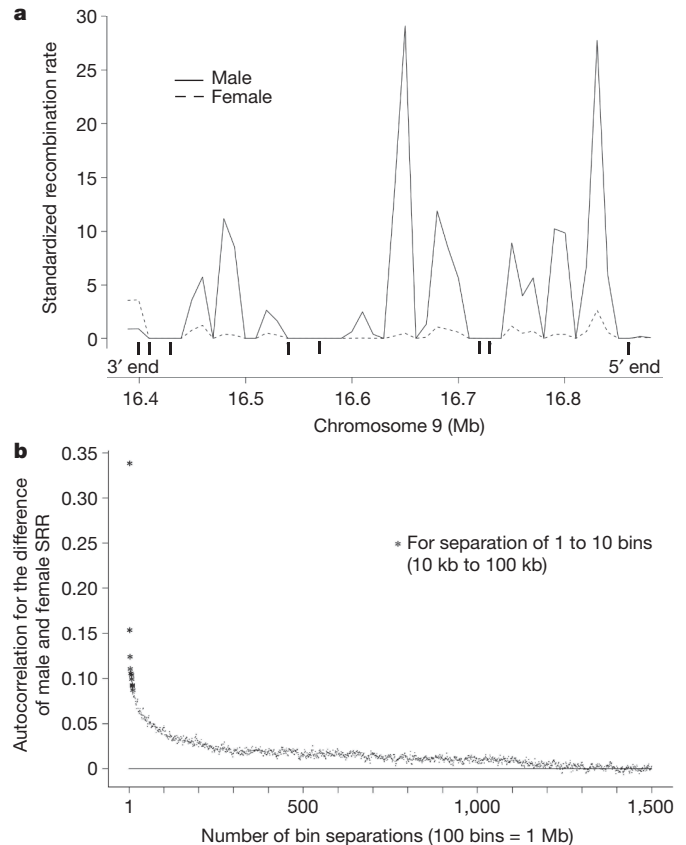


Figure 2 | Sex differences in recombinations. **a**, Male and female SRRs at the *Basonuclin-2* gene. Recombinations in this region are dominated by those resulting from male meioses. It is, however, noted that, although female recombination rates are generally low here, locations of male recombination hotspots often correspond to local peaks for female recombination rate. **b**, Autocorrelations of the difference between male and female SRR as a function of the number of bin separations. Note that, albeit small (~ 0.007), the correlation is clearly positive for bins that are 10 Mb (1,000 bins) apart.

partly correlated. Indeed, the correlation between the male and female maps is less at 3-Mb (0.649) than at 10-kb resolution (0.659), even though the former is less affected by sampling variation.

We classified the 10-kb bins as genic, intergenic or at gene boundaries (Table 1). On average, the recombination rate is lower in genic regions than in intergenic ones, a difference that is greater for females (average SRR = 0.898 and 1.053, respectively) than males (average SRR = 0.992 and 1.012, respectively). For both sexes, the recombination rate tends to be lower at genic bins containing exons, and higher for those containing only introns, particularly those where the closest exon is more than three bins away. This latter difference is much greater for males (SRR = 0.868 and 1.284, respectively) than females (SRR = 0.843 and 1.013, respectively). In fact, intron bins far from exons exhibit the greatest difference between male and female SRR (0.270, $P = 2.2 \times 10^{-7}$) among the bin categories studied. At intergenic regions, for both sexes, the recombination rate first increases with distance from the first or last exon of genes, peaking at approximately three to four bins away, then decreases. The changes are more dramatic in females than males (Table 1). For intergenic bins that are ten bins or less from genes, the average SRR for males and females is 1.119 and 1.256, respectively ($P = 1.3 \times 10^{-19}$). Hence, although more male recombinations participate in shuffling exons within genes, female meioses are characterized more by gene shuffling.

For both sexes, similar differences in SRR exist between the 5' and 3' ends of genes. For intergenic regions within 100 kb of the nearest gene, the average SRR at the 5' ends is approximately 0.15 lower than that at the 3' ends ($P = 2.7 \times 10^{-7}$). The difference disappears for distances greater than 100 kb. In contrast, bins containing the first exon of a gene

Table 1 | Sex-specific standardized recombination rate and genomic regions

Bin type (10 kb)	Number of bins	Male recombination rate [§]	Female recombination rate [§]	Male recombination rate – female recombination rate (<i>P</i> value)
Exon* (>99% genic)	40,538	0.868	0.843	0.024 (0.25)
Intron (n.e.† = 1)	20,535	1.010	0.903	0.106 (1.5 × 10 ⁻⁶)
Intron (n.e. = 2)	8,718	1.047	0.958	0.089 (0.0029)
Intron (n.e. = 3)	5,009	1.095	0.919	0.177 (3.9 × 10 ⁻⁵)
Intron (n.e. = 4)	12,580	1.284	1.013	0.270 (2.2 × 10 ⁻⁷)
Intron (all)	46,842	1.099	0.945	0.155 (1.1 × 10 ⁻¹¹)
Genic (exon + intron)	87,380	0.992	0.898	0.094 (1.9 × 10 ⁻⁸)
Gene boundary‡	24,498	0.963	1.079	-0.117 (1.3 × 10 ⁻⁵)
Intergenic (n.e. = 1)	15,912	1.120	1.196	-0.075 (0.0025)
Intergenic (n.e. = 2)	10,935	1.118	1.242	-0.124 (2.5 × 10 ⁻⁶)
Intergenic (n.e. = 3)	8,245	1.149	1.356	-0.207 (4.8 × 10 ⁻¹²)
Intergenic (n.e. = 4)	6,495	1.175	1.349	-0.174 (6.6 × 10 ⁻⁷)
Intergenic (n.e. = 5)	5,385	1.138	1.291	-0.153 (4.6 × 10 ⁻⁵)
Intergenic (n.e. = 6)	4,602	1.142	1.277	-0.135 (0.00062)
Intergenic (n.e. = 7)	4,017	1.126	1.242	-0.116 (0.010)
Intergenic (n.e. ≤ 10)	65,380	1.119	1.256	-0.136 (1.3 × 10 ⁻¹⁹)
Intergenic (n.e. ≥ 20)	48,057	0.845	0.773	0.072 (0.011)
Intergenic (all)	132,568	1.012	1.053	-0.041 (0.00092)

* An exon bin is one that is more than 99% genic and contains either part or full exons. † n.e., number of bins to nearest exon. ‡ A gene boundary bin is one that is partly genic (that is, it contains either part of or the full first or last exon of genes) and partly intergenic (≥1%). § Recombination rates are in standardized units; that is, over the 244,446 10-kb bins studied, they average to 1 for males and females, respectively. || *P* values have been adjusted for correlation among bins close to each other using a randomization procedure (see Supplementary Information).

have a higher average SRR than those containing the last exon (~ 0.11 , $P = 7.2 \times 10^{-4}$). This difference, however, does not extend further into the genic regions, which suggests that the first intron has a higher recombination rate than the last intron at the immediate neighbourhoods of the first and last exons. Figure 3 summarizes the relationships between sex-specific recombination rate and genes.

Differences in recombination rate exist between individuals of the same sex^{8,14,15}. Recently, the *PRDM9* gene was shown to be a major determinant of hotspots in humans⁴⁻⁶. This gene is highly polymorphic, with most of its sequence variants clustering in the zinc-finger domain of the gene. The Human Genome assembly (hg18) ascribes 13 zinc-finger repeats to the *PRDM9* gene. These repeats are invariant except at positions -1, 3 and 6 of each of the zinc-finger α -helices. Variations in the number of repeats within the human population have been described. In the Hutterite population, carriers of a rare version of the gene with 16 zinc-finger repeats were shown to have substantially fewer recombinations in hotspots than non-carriers⁴. To investigate comprehensively variants that could affect hotspots, we performed a genome scan, separately, for the 6,041 mothers and 4,389 fathers studied, correlating the fraction of recombinations in hotspots (henceforth referred to as the hotspot phenotype) with SNPs on the Illumina 1M chip (Methods Summary).

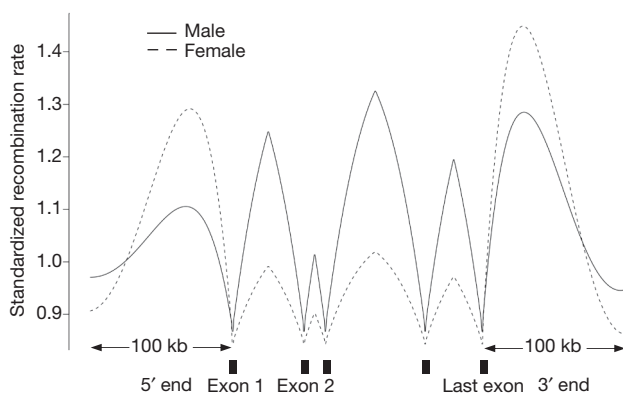


Figure 3 | Sex-specific recombination rates and genes. Schematic picture summarizing general trends (see Table 1); it is not meant to reflect the recombination rate pattern around a specific gene. Male SRR, although low at exons, tends to be high at intronic regions that are distant from exons. Male and female SRRs both tend to be high at intergenic regions around 40 kb from the first or last exon of a gene, but it is higher for females. Also, for both sexes, intergenic regions close to 3' ends tend to have higher recombination rates than those close to 5' ends.

For both sexes, many SNPs around *PRDM9* achieved genome-wide significance ($P < 5 \times 10^{-8}$, Supplementary Figs 2 and 3). Most significant was rs2914276, with the minor allele G (frequency = 3.9%) associating with fewer recombinations in hotspots ($P < 10^{-100}$ for females and $P < 10^{-50}$ for males). Determining the zinc-finger repeat number of 575 Icelanders, enriching for carriers of rs2914276-G, we observed variations in repeat numbers from 12 to 15 (Supplementary Information, Supplementary Fig. 4 and Supplementary Table 4). Imputing this polymorphism into others, 12 to 15 repeats were estimated to have frequencies of 0.1%, 96.0%, 3.2% and 0.6%, respectively. Rs2914276-G correlates substantially with either 14 or 15 repeats ($r = 0.83$), whereas the major allele A is correlated with 12 or 13 repeats. No significant difference was seen between 12 and 13 repeats with respect to hotspots. Individuals carrying only 12 or 13 repeats have 28.6% and 37.1% of their recombinations in hotspots for females and males, respectively. Fourteen and 15 repeats are associated with significantly lower fractions to 19.1% and 25.4%, and for 15 repeats to 20.5% and 27.3%. Although the higher fractions for the 15 repeats than the 14 repeats are barely significant when results from both sexes are combined ($P = 0.018$), it emphasizes that the fraction of recombinations in hotspots does not decrease monotonically with number of repeats. The number of repeats has a stronger association with the hotspot phenotype than rs2914276, but the latter remains highly significant after accounting for the former. By sequencing the zinc-finger repeats from 55 Icelandic chromosomes covering the 12 to 15 repeat spectrum, a variant rs6875787 leading to an amino-acid change in the sixth zinc finger, also noted previously⁴, was seen (Supplementary Fig. 4a). Additional sequencing and further investigations (Supplementary Information) showed that the minor allele of rs6875787 is in about 5.3% of the chromosomes with 13 repeats, and confirmed a previous suggestive finding⁴ that it lowers the fraction of recombinations in hotspots. However, the effect is only about one-tenth that of the 14 or 15 repeats. Thus it could be that many polymorphisms in the *PRDM9* locus affect hotspots, but the repeat polymorphism alone captures most (>90%) of the association currently observed between variations in *PRDM9* and the hotspot phenotype. Because the 14 and 15 repeats do not behave that differently, we collapsed them into a single 14/15 allele with a frequency of about 3.9%. We estimated that the differences between the 14/15 and 12/13 repeats alone can account for 60% and 44% of the total systematic component of the hotspot phenotype for males and females, respectively (Supplementary Information). The 16 repeat found in the Hutterites⁴ was not observed in the Icelandic samples examined, but the described differences between the 14/15 and the 12/13 repeats are novel.

Because few parents carry more than one 14/15 allele, we grouped heterozygous and homozygous carriers together (351 males and 429 females, contributing 502 and 612 meioses, respectively) to construct two sex-specific carrier recombination maps. The same was done for non-carriers: that is, those with only 12 or 13 repeats. Although carriers have fewer recombinations in hotspots, they remain in substantial excess of the genome average (average SRR = 13.1 and 11.4 for male and female carriers, respectively, compared with an average SRR of 19.0 and 16.9, respectively, for non-carriers). Moreover, the binding motif corresponding to 13 repeats is associated with increased recombination rate and hotspots in carriers and non-carriers, although the effect is slightly stronger for non-carriers (Supplementary Information and Supplementary Table 5). The binding motif predicted for 14 repeats is also associated with increased recombination rate for both sets of individuals, but here the effect is stronger for carriers. In addition to motif intensities at a bin itself, motif intensities at nearby bins also appear to have an effect (Supplementary Table 6). However, the magnitudes of all these correlations are low, and the motifs alone provide very modest power for predicting hotspots.

For the 10-kb bins, the correlation between the CEU and YRI maps is 0.716 (Supplementary Table 3). The correlation between our overall sex-averaged map is stronger with the COMBINED map (0.729) than with the CEU (0.700) and YRI (0.643) maps, which indicates that a substantial part of the difference between the CEU and YRI maps is noise. Nonetheless, by examining the variations of *PRDM9* in the HapMap YRI samples, we found zinc-finger repeat lengths of 12 to 15 and 17 to 19 (Supplementary Table 4 and Supplementary Fig. 4b). Grouping different repeat lengths into three composite alleles, the 12/13, 14/15 and 17/18/19 alleles have frequencies of 65.8%, 26.7% and 7.5%, respectively. The 14/15 allele is much rarer in the CEU samples, where only 13 and 14 repeats are found, at frequencies of 96.6% and 3.4%, respectively. We standardized all maps, including sex-averaged maps constructed separately for carriers and non-carriers of the 14/15 repeat (referred to as mapC and mapNC, respectively), in the same way as with our sex-specific maps. When regressing the difference between the YRI and CEU maps, on mapC and mapNC jointly, the coefficient of mapC was positive (0.089, $P < 10^{-100}$) and the coefficient of mapNC was negative (-0.307 , $P < 10^{-300}$). When regressing the CEU and YRI maps separately on mapC and mapNC jointly, all coefficients were positive. For CEU, the coefficient of mapC was approximately 5.4% of the coefficient of mapNC, compared with 33.2% for YRI. Thus true differences between Europeans and Africans, explained by differences in frequencies of *PRDM9* variants, are identified.

There are 4,006 hotspot bins (10-kb bins with SRR >10) in our overall sex-averaged map, compared with 4,010 for the COMBINED map. The overlap is 2,139. Based on mapC and mapNC, 18.3% of the recombinations of carriers of the 14/15 repeat are in hotspots of our overall sex-averaged map, whereas the figure is 27.0% in non-carriers. Simulations that adjust for increased variation in maps estimated based on a reduced sample size show that, genome-wide, mapC is not smoother than mapNC, which suggests that the reduced recombination rate at hotspots defined by the overall map is compensated for by hotspots elsewhere. Among the 5,034 bins with an SRR greater than 10 in mapC, 1,380 and 371 have an SRR less than 3 and 1, respectively, in mapNC. A permutation test shows that 350 and 38 bins with such properties are expected by chance, suggesting that 1,030 of the 1,380 (74.6%), and 333 of the 371 (89.8%), are true hotspots specific to the 14/15 repeat carriers. Further support comes from examining the differences in SRR between the YRI and CEU maps, wherein 545 of the 1,380 (39.5%) and 150 of the 371 (40.2%) identified bins fall into the top fifth percentile of all bins studied.

In the genic regions, our sex-averaged map and the COMBINED map have an average SRR of 0.929 and 0.904, respectively. The difference is significant ($P = 3.1 \times 10^{-5}$) and is mainly accounted for by bins containing exons and intronic bins within 10 kb of an exon. One possibility

is that regions around exons are more likely to have been subject to natural selection, resulting in a lower number of recombinations detectable from LD and consequently leading to an underestimation of recombination rate in LD-based maps⁷. However, for regions (23,614 10-kb bins) that have been proposed as targets of selection by at least two of nine genome scans compiled by Akey¹⁶, the difference between the COMBINED map and our map is opposite to that observed for regions around exons. Specifically, although both maps assign low recombination rates to these regions, the average SRR in the COMBINED map (0.650) is significantly higher ($P = 5.7 \times 10^{-9}$) than that in our sex-averaged map (0.593). Whether these differences are a result of some novel bias that affects estimates of LD-based recombination rate at regions under selection, or partly reflect properties exhibited by the statistical methods used to identify regions under selection that are currently poorly understood, warrants further investigation.

Polymorphisms at the *RNF* gene that influence total genome-wide recombination rates of males and females in opposite directions¹⁵ have little impact on the fraction of recombinations in hotspots (Supplementary Information). Variations at the *PRDM9* gene influence recombination locations in a similar manner for both sexes, but have little effect on total genome-wide recombination rate. An inversion on chromosome 17 reported to associate with increased fertility and genome-wide recombination rate¹⁷ also appears to increase the fraction of recombinations in hotspots, but the effect is limited to females ($P = 2.9 \times 10^{-5}$ and 0.49 for females and males, respectively) and is modest (Supplementary Information). These polymorphisms, together with the systematic regional and local differences in recombination rates between the sexes, provide a glimpse of nature's ingenuity in building diversity and flexibility into the system. The maps constructed in this study (available at <http://www.decode.com/addendum>) should serve as a valuable resource for genetics research for years to come.

METHODS SUMMARY

Subjects were 20,217 distinct individuals genotyped using various Illumina BeadChips and processed to determine parental origin. When searching for variants associated with the hot-spot phenotype, adding to SNPs used to determine recombinations, another 497,257 SNPs typed for a subset of the individuals were imputed into the others with methods used before. Variants at the *PRDM9* gene were similarly imputed. Imputations were not used for map construction. The LD-based maps were downloaded from <https://mathgen.stats.ox.ac.uk/impute>. The entire zinc-finger domain of the *PRDM9* gene was amplified with unique primers outside the repetitive region, avoiding homology to chromosome 16, for a total of 575 Icelanders, 30 CEU and YRI trios, and 74 Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) samples. The amplified product was run on agarose gel to determine the number of zinc-finger repeats. For further analysis, 55 bands of different repeat lengths from Icelanders and 21 from YRI samples were isolated from the agarose gel, cloned and fully sequenced. Statistical tests used were mainly regression based, for example paired and unpaired *t*-tests, correlation tests and regressions. Genomic control¹⁸ was used for the genome-wide association analysis with the hotspot phenotype. For map comparisons, to handle correlations among close-by bins, a procedure that permutes and flips chromosomes was used to calculate adjustment factors for the test statistics. Another randomization procedure that permutes individuals was used to estimate false discovery rates for sex-specific and new hotspots. See Supplementary Information for details.

Received 18 May; accepted 14 September 2010.

1. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
2. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
3. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genet.* **40**, 1124–1129 (2008).
4. Baudat, F. *et al.* *PRDM9* is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
5. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).

6. Parvanov, E. D., Petkov, P. M. & Paigen, K. *Prdm9* controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
7. O'Reilly, P. F., Birney, E. & Balding, D. J. Confounding between recombination and selection, and the Ped/Pop method for detecting selection. *Genome Res.* **18**, 1304–1313 (2008).
8. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
9. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genet.* **40**, 1068–1075 (2008).
10. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
11. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**, 1–38 (1977).
12. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
13. Lang, M. R., Patterson, L. B., Gordon, T. N., Johnson, S. L. & Parichy, D. M. *Basonuclin-2* requirements for zebrafish adult pigment pattern development and female fertility. *PLoS Genet.* **5**, e1000744 (2009).
14. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
15. Kong, A. *et al.* Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science* **319**, 1398–1401 (2008).
16. Akey, J. M. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* **19**, 711–722 (2009).
17. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nature Genet.* **37**, 129–137 (2005).
18. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Reich for discussion and suggestions.

Author Contributions A.K. and K.S. planned and directed the research. A.K. wrote the first draft of the paper and, with K.S., U.T. and A.H., wrote most of the final version. D.F.G. improved previous phasing procedures and, with G.M., made the recombination calls. G.T. created the maps and, with M.L.F., assisted A.K. in the analyses. U.T., Aslaug J., A.S., Adalbjorg J., K.T.K. and G.B.W. performed experiments providing information on sequences at the *PRDM9* gene. A.G. did the variant imputations. G.M. and S.A.G. determined the locations and intensities of genomic features. A.H. assisted in the study on selection.

Author Information The maps constructed in this study are available at <http://www.decode.com/addendum>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.K. (kong@decode.is) or K.S. (kari.stefansson@decode.is).