Supplementary Notes Robustness of results to model misspecification

1. We can reject the hypothesis that all hotspots estimated in the chimpanzee sample are also present in the human sample:
This conclusion depends on our power to detect hotspots and hence on model assumptions. However several lines of argument suggest that the conclusion is robust. First, we saw no evidence for a difference in the distribution of hotspot intensity for central chimpanzees and African-American samples (results not shown). So it appears to not be the case, that, for instance, λ in humans is systematically lower than it is in chimpanzees and that in our simulations the power was thus inflated. Furthermore, for the presented results, we correct for apparent biases we saw in the performance of PHASE (see Methods). Correcting for these biases we found that the average in 100 simulations was 72% and the minimum was 59%. However, if we do not adjust for the bias in the parameter estimates, then the average in 100 simulations was 85% and the minimum was 74%. Thus, both sets of simulations suggest a much higher proportion than what is observed. Given the huge discrepancy between observed and expected (Fig. 2), our conclusion that hotspots are not identical in the two species is likely to be robust to departures from model assumptions.

2. We can reject the hypothesis that total recombination rates are perfectly correlated between the two species:
This finding depends on our ability to estimate recombination rates, which is model dependent. Thus, one concern might be that our results are dependent on the hotspot model, in particular, on the density of recombination hotspots. For the presented results, we assumed one hotspot per window. We ran additional simulations in which we relaxed this assumption and choose the number of hotspots per window independently in the two species from a truncated Poisson distribution. The mean correlation coefficient was very slightly higher in this set of simulations (0.73 vs. 0.69). The similarity under these two hotspot models suggests that the results are not overly sensitive to the specific hotspot model. Furthermore, we are comparing the total recombination rate per 50 kb window in the two species, a parameter that depends less on assumptions about hotspots (results not shown). Finally, since the discrepancy between what is expected and what is observed is large (Fig. 4), our conclusion is likely to be robust to misspecification of the hotspot model.

3. We can reject the hypothesis that background recombination rates are perfectly correlated between the two species:
One concern might be that since we assume one hotspot per window, undetected hotspots could inflate the background rate in a subset of windows. In particular, the correlation that we observe in background rates might reflect conservation in the number of hotspots per window (even though their exact locations are not conserved). The presence of multiple hotspots within a 50 kb window seems unlikely, however, given that hotspots in humans are thought to occur every 60kb-200kb[4,7]. Furthermore, windows with an inferred hotspot in chimpanzee data are no more likely to contain an inferred hotspot in the homologous window in African-American data (results not shown), in contrast to what might be expected if the density of hotspots were conserved. Thus, the correlation in background recombination rates is unlikely to be affected by our assumption of one hotspot per window. Finally, since the discrepancy between what is expected and what is observed is large, our conclusion is likely to be robust to misspecification of the hotspot model.

Supplementary Table 1 Observed frequency of hotspots. The 14Mb region was divided into windows of 70kb (with an overlap of 20kb). Presented here are the number of windows per population and what percentage of these windows contained a distinct hotspot (*i.e.* not overlapping with a hotspot in the adjacent window; see Methods). While these observations raise the possibility that hotspot frequencies vary among populations, simulations of the standard coalescent suggest that these differences can be explained by varying power among samples due to differences in effective population size (Supplementary Table A.2). Furthermore, these simulations suggest that few of the inferred hotspots are likely to be spurious (Supplementary Table A.2 and A.3). In light of these considerations, we are likely to have underestimated the true hotspot frequencies in both species – especially since we have assumed at most one hotspot per 70kb window, when there could be more.

	Number	Perc. Hot.
CenChimp	260	15%
AfAmer	234	58%
ChnAmer	226	50%
EurAmer	227	45%

Supplementary Table 2 Results from power simulations. Simulations were run for four values of λ: 1, which corresponds to no hotspot, and 50, 100, and 200 which corresponds to a hotspot. In addition, three sets of ρ values were run: low (25^{th} percentile value estimated for each population independently), intermediate (median value), and high (75^{th} percentile). For each combination of parameters and for each population, 100 simulations were run. See methods for other details of how these simulations were run. In each cell is the percentage of runs (out of 100) in which a hotspot was inferred. Hence, for $\lambda = 1$, this is an estimate of the false positive rate, and for $\lambda > 1$, of the power to detect a hotspot.

	$\lambda = 1$			$\lambda = 50$	$\lambda = 100$			$\lambda = 200$
	low	interm	high	interm	low	interm	high	interm
CenChimp	1%	2%	1%	25%	48%	44%	28%	55%
AfAmer	0%	3%	6%	59%	86%	86%	80%	93%
ChnAmer	0%	1%	8%	44%	42%	79%	83%	82%
EurAmer	2%	1%	7%	52%	43%	68%	91%	92%

Supplementary Table 3 Simulation results for other demographic assumptions. To examine the effect of more realistic demographic assumptions on the false-positive rate (*i.e.* the probability of detecting a hotspot when none exists) in the four populations, we ran simulations with a bottleneck model for the central chimpanzee[27], Chinese-American[29] and European-American samples[29-31], and a model of growth for the African-American sample[29,30]. Simulations were run for $\lambda = 1$ and intermediate values of ρ. For each population, 100 simulations were run. See Methods for other details.

	$\lambda = 1$
	interm
CenChimp bottleneck	1%
AfAmer growth	13%
Chn/EurAmer bottleneck	2%

Additional Literature Cited

29. Marth, G.T., Czabarka, E., Murvai, J., Sherry, S.T. The allele frequency spectrum in genome–wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* **166**, 351–372 (2004).
30. Akey, J.M., Eberle, M.A., Rieder, M.J., Carlson, C.S., Shriver, M.D., Nickerson, D.A., Kruglyak, L. Population history and natural selection shape patterns of genetic variation in 132 genes. *PloS Biol.* **2**, 1–9 (2004).
31. Marth, G., *et al.* Sequence variations in the public human genome data reflect a bottlenecked population history. *Proc. Natl. Acad. Sci. USA* **100**, 376–381 (2003).

Supplementary Figure 1 Comparison of the method and data used here with results from two sperm typing experiments for the MHC region. The data for the MHC were collected along with the other human data analyzed in this paper[12], so that the sampling scheme and individuals are identical. The results for all three populations are shown; note that the results are expected to be similar, given that the three share much of their evolutionary history. We find estimates of total recombination rates per window similar to those estimated by sperm-typing and the inferred locations of recombination hotspots are in accordance as well. The concordance of LD-based estimates of recombination with more direct measurements has been demonstrated before[7,10], including for these same hotspots[4,6], although not for recombination estimated at the scale of tens of kilobases as examined here. Thus, the indirect estimation procedure used in this paper appears to produce reliable results for the cases tested.

A-C: Ability to estimate relative recombination rates

The sperm typing data for these plots come from Cullen *et al.*; they typed sperm for 11 European-Americans and one African-American in a 3.3 Mb interval for windows of size ranging from 30kb to 200kb[8]. To obtain estimates of recombination rate (cM/Mb) for these 30 windows, we divided the observed number of recombinants in each window (from Figure 2, ref. 9) by the total number of informative meioses (20031, ref. 8) and then by length of the segment. We then multiplied this by 100 to convert from Morgans to cM. The length of the segment was obtained by computing the difference in the nucleotide position of the two flanking markers (from Table A, ref. 8). To convert from r to ρ, we used two estimates of N_e: 10,000 which is suitable for samples of European and Chinese ancestry[28,32] and 15,000 which is suitable for samples of African-American ancestry[28]. Different assumptions about N_e will shift the graph vertically by a constant. For the non-sperm typing data, we used the same analysis as in the Methods, except that the window sizes matched those in Cullen *et al.* [8]. The windows were generated by aligning the primers for each marker (Table A, ref. 8) against build 34 of the human genome using BLAST[22], and then finding the homologous regions in the Hinds *et al.* data[12]. The 5% (low) and 95% (high) estimate of the natural log of the total ρ for each window is then plotted for the three populations.

There is reasonable agreement in the relative rates of recombination: in particular, PHASE correctly infers the two large peaks in recombination rates and several of the smaller ones. Interestingly, many areas where ρ falls above the sperm typing estimates are regions thought to experience balancing selection[33] or linked regions. The particular genes affected are indicated on the plots below. Under a model of multi-allelic balancing selection, ρ is generally inflated due to a deeper genealogical history (C. Spencer, pers. comm.). Note that there is also error in the measurements of recombination rates by sperm typing, as well as uncertainties in estimates of N_e, which are not included here.

D-F: Ability to infer hotspots

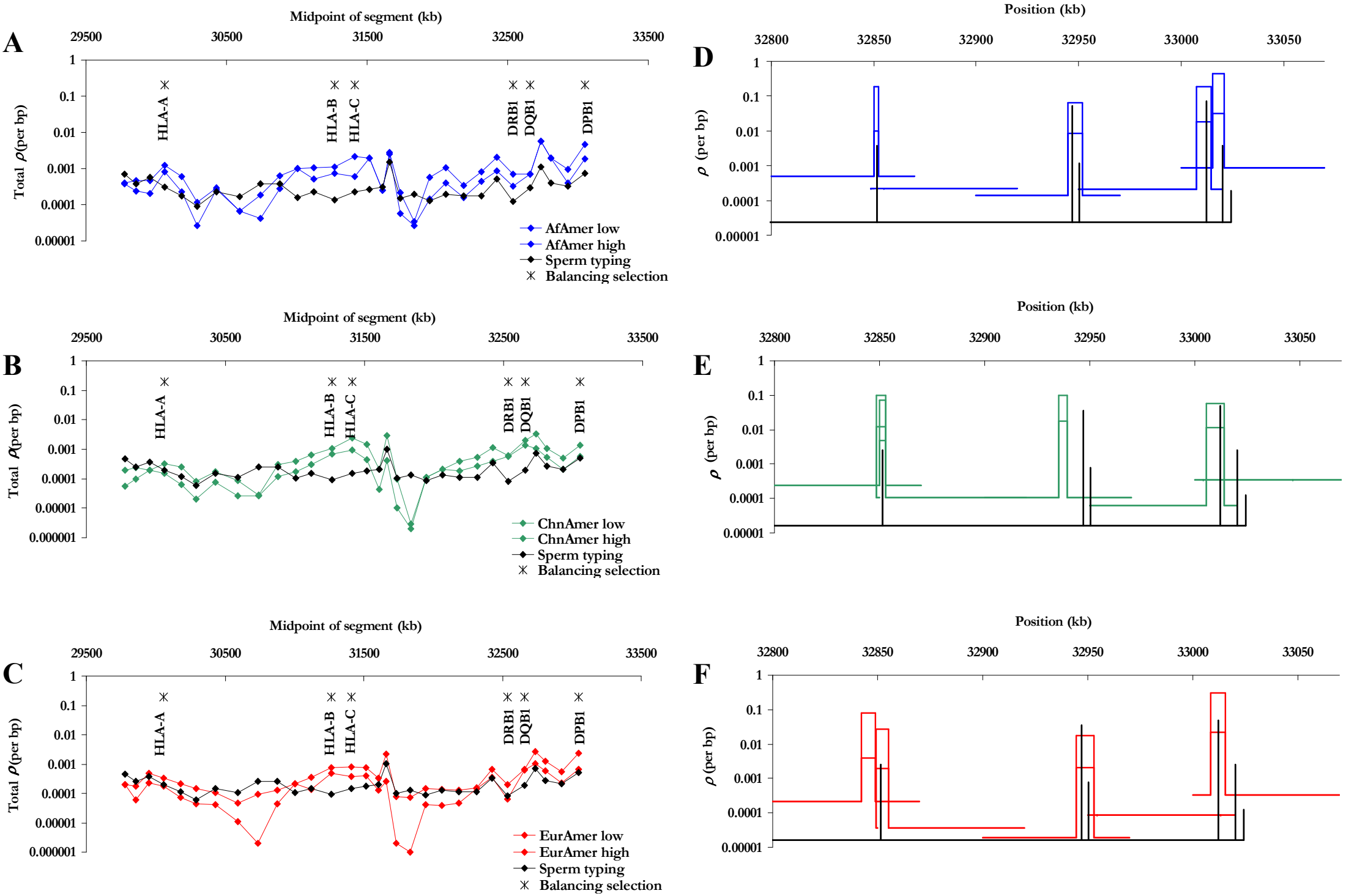
The sperm typing data on these plots come from Jeffreys *et al.,* in which sperm was typed for two to six UK males for a ~220 kb region[34,35]. They identified six hotspots. To determine the location of the hotspots on build 34, we used Blat (http://genome.ucsc.edu/cgi-bin/hgBlat) to align the center of the *TAP2* hotspot (http://www.le.ac.uk/ge/ajj/tap2/SEQdata.html) to build 34. We used the midpoint as the location of the hotspot, and then computed the relative midpoints of the other hotspots from Figure 5 in Jeffreys *et al.*[34]. Likewise the values of λ come from Figure 5.

Finally, we used the background recombination rate of 0.04 cM/Mb, and converted it to ρ using the same estimates of N_e as above. The non-sperm typing data for these graphs were broken up into five 70kb overlapping windows. Plotted is the natural log of the median background ρ, and if a hotspot was inferred (Bayes Factor > 50), the 5% and 95% estimate of lambda multiplied by the median background ρ for the 95% CI on location. The legends for these graphs are the same as for the graphs on the left.

There is reasonable agreement as to the presence and location of hotspots, although in a few windows, we fail to detect a hotspot found by sperm-typing. In these windows either the hotspot is on the edge or there are multiple hotspots. Simulations suggest these situations decrease the power to detect a hotspot (data not shown). Since the simple hotspot model is assumed (*i.e.* one hotspot per window), multiple hotspots within a cluster could not have been identified. The background rate of recombination measured by sperm typing is much lower than estimated here. However these hotspots are near genes that are putatively under balancing selection, as discussed above. Also, the relative intensities of the hotspots are not as concordant. While sperm typing experiments provide recombination rates in extant males, the method used here provides historical sex-averaged recombination rate estimates. Hence differences might reflect a difference in intensities between males and females, or a change in intensity over time. Finally, there are errors in the measurements of recombination rates by sperm typing and uncertainties in the estimates of N_e, which are not included here.

Additional Literature Cited

32. Frisse, L. *et al.* Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831-843 (2001).
33. Hughes, A.L. & Yeager, M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* **32**, 415-435 (1998).
34. Jeffreys, A.J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* **29**, 217-222 (2001).
35. Jeffreys, A.J., Ritchie, A. & Neumann, R. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum. Mol. Genet.* **9**, 725-733 (2000).


A
Midpoint of segment (kb)
29500
30500
31500
32500
33500
Total ρ (per bp)
1
0.1
0.01
0.001
0.0001
0.00001
HLA-A
HLA-B
HLA-C
DRB1
DQB1
DPB1
AfAmer low
AfAmer high
Sperm typing
Balancing selection
B
Midpoint of segment (kb)
29500
30500
31500
32500
33500
Total ρ (per bp)
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
HLA-A
HLA-B
HLA-C
DRB1
DQB1
DPB1
ChnAmer low
ChnAmer high
Sperm typing
Balancing selection
C
Midpoint of segment (kb)
29500
30500
31500
32500
33500
Total ρ (per bp)
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
HLA-A
HLA-B
HLA-C
DRB1
DQB1
DPB1
EurAmer low
EurAmer high
Sperm typing
Balancing selection
D
Position (kb)
32800
32850
32900
32950
33000
33050
ρ (per bp)
1
0.1
0.01
0.001
0.0001
0.00001
E
Position (kb)
32800
32850
32900
32950
33000
33050
ρ (per bp)
1
0.1
0.01
0.001
0.0001
0.00001
F
Position (kb)
32800
32850
32900
32950
33000
33050
ρ (per bp)
1
0.1
0.01
0.001
0.0001
0.00001

Supplementary Figure 2

Allele frequency distribution for the chimpanzee and human data. The observed distributions were created by determining the minor allele frequency for all SNPs that were called in all individuals. The observed distribution is then compared to what is expected under the standard neutral model for a given value of θ. This θ value was estimated by matching the total number of segregating sites used in the analysis, *i.e.* by solving the following equation for θ, $S_C = \theta \sum_{i=C+1}^{i=nchrom-(C+1)} \frac{1}{i}$,

where *nchrom* is the number of chromosomes. For the human data, C is 2 and S_c is the number of non-singleton and non-doubleton segregating sites (*i.e.* minor allele frequency > 5%). For the chimpanzee data, C is 1 and S_c is the number of non-singleton segregating sites.