Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/10/eaaw9206/DC1

Supplementary Materials for

Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations

Jeffrey P. Spence* and Yun S. Song*

*Corresponding author. Email: spence.jeffrey@berkeley.edu (J.P.S.); yss@berkeley.edu (Y.S.S.)

Published 23 October 2019, *Sci. Adv.* **5**, eaaw9206 (2019)
DOI: 10.1126/sciadv.aaw9206

This PDF file includes:

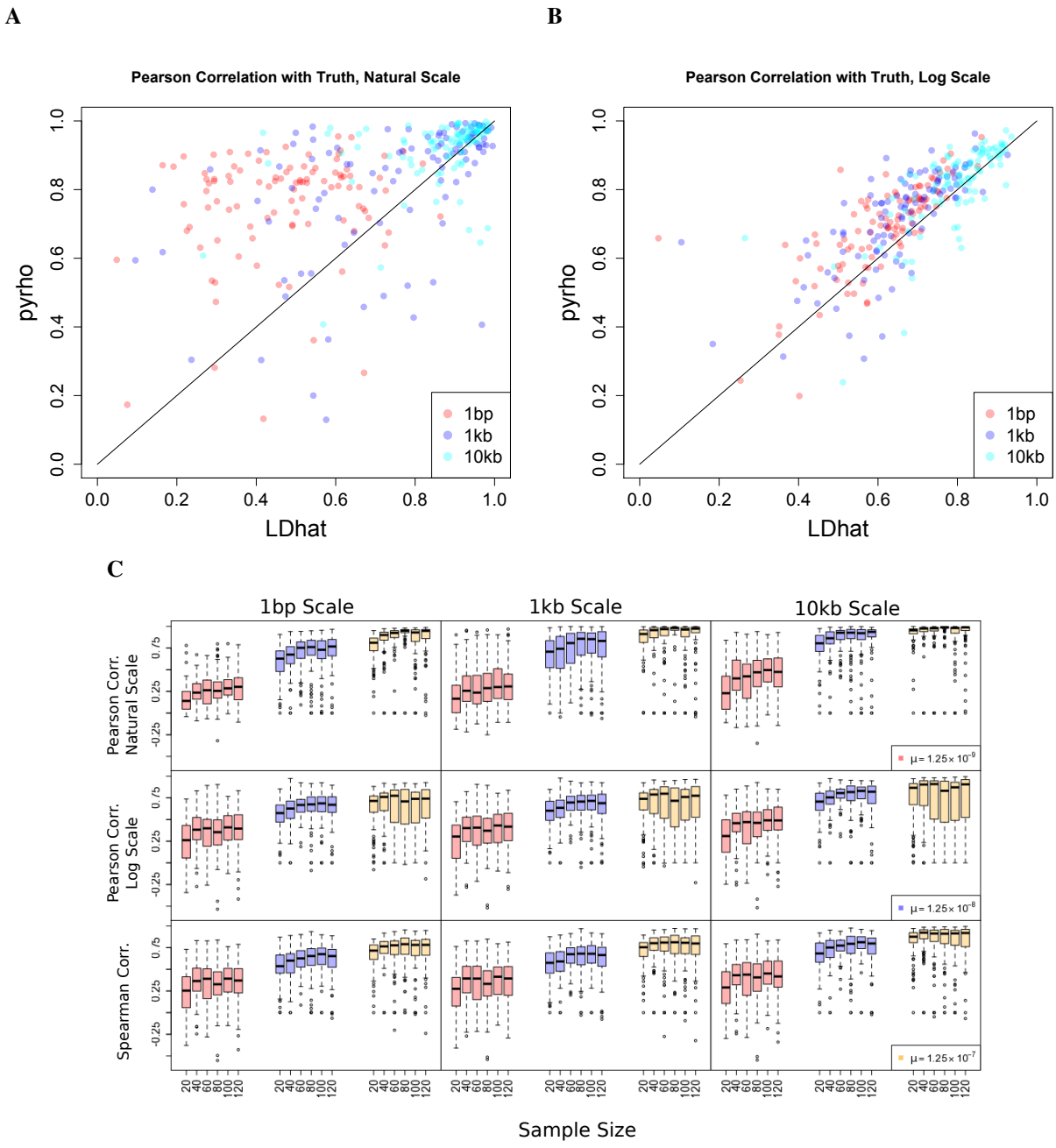

Fig. S1. Additional measures of accuracy on simulated data.

(**A**) Pearson correlation with the truth in natural scale and (**B**) Pearson correlation with the truth in log scale. (**C**) Results for different sample sizes and mutation rates. Each box is the distribution of correlations of a given type at a particular spatial scale across 100 simulations each of 1Mb in length.

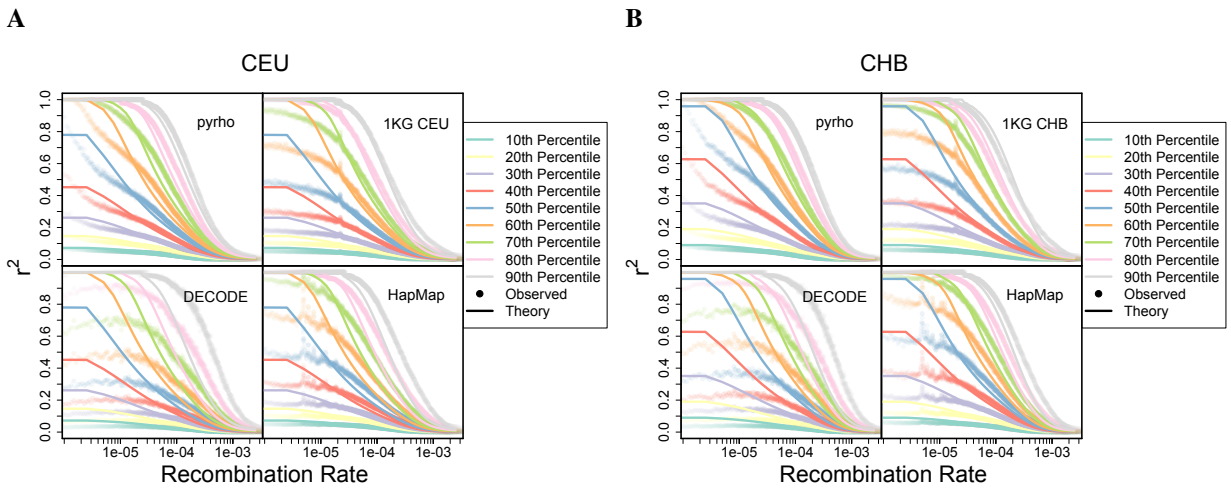

Fig. S2. Goodness-of-fit of inferred recombination maps.

Comparison of theoretical and observed r^2 for the inferred recombination maps. For a pair of SNPs, r^2 is a random quantity and depends on the rate of recombination between the SNPs. Solid lines show theoretical deciles of this distribution for pairs of sites separated by different recombination distances with minor allele frequency > 0.1 at both sites as calculated under the population size for YRI in Fig. 2A. Shaded points are the deciles of the empirical distribution obtained by considering pairs of sites with minor allele frequency > 0.1 binned by the recombination rate separating them according to the different recombination maps. 1KG YRI is the population-specific recombination map for YRI in (*15*); DECODE is the sex-averaged recombination map in (*28*); and HapMap is the recombination map in (*34*)

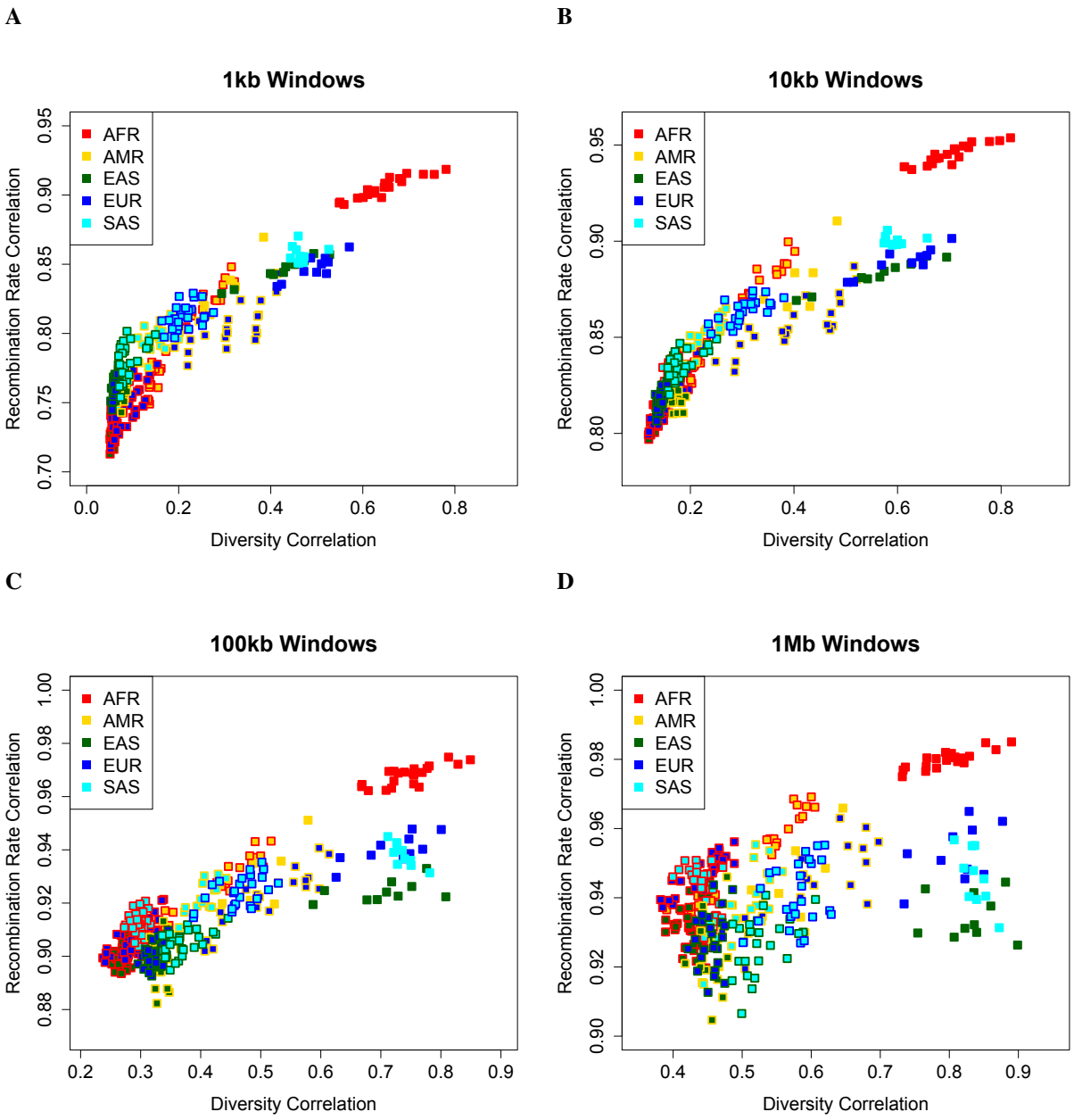

Fig. S3. Recombination rates are more similar across populations than a measure of diversity.

Comparison of the Spearman correlation in a measure of diversity, π, to the Spearman correlation of fine-scale recombination rates across populations at different scales. Points with a single color are comparisons of two populations within the same continental group. Points with two colors are comparisons of populations from different continental groups. In general, the recombination rates are more correlated than the measure of diversity, but populations with more correlated measures of diversity tend to also have more correlated recombination maps. The lower correlation in recombination rates between pairs of non-African populations from the same continental group may be due to smaller effective population sizes, making inference more difficult, resulting in regression attenuation.

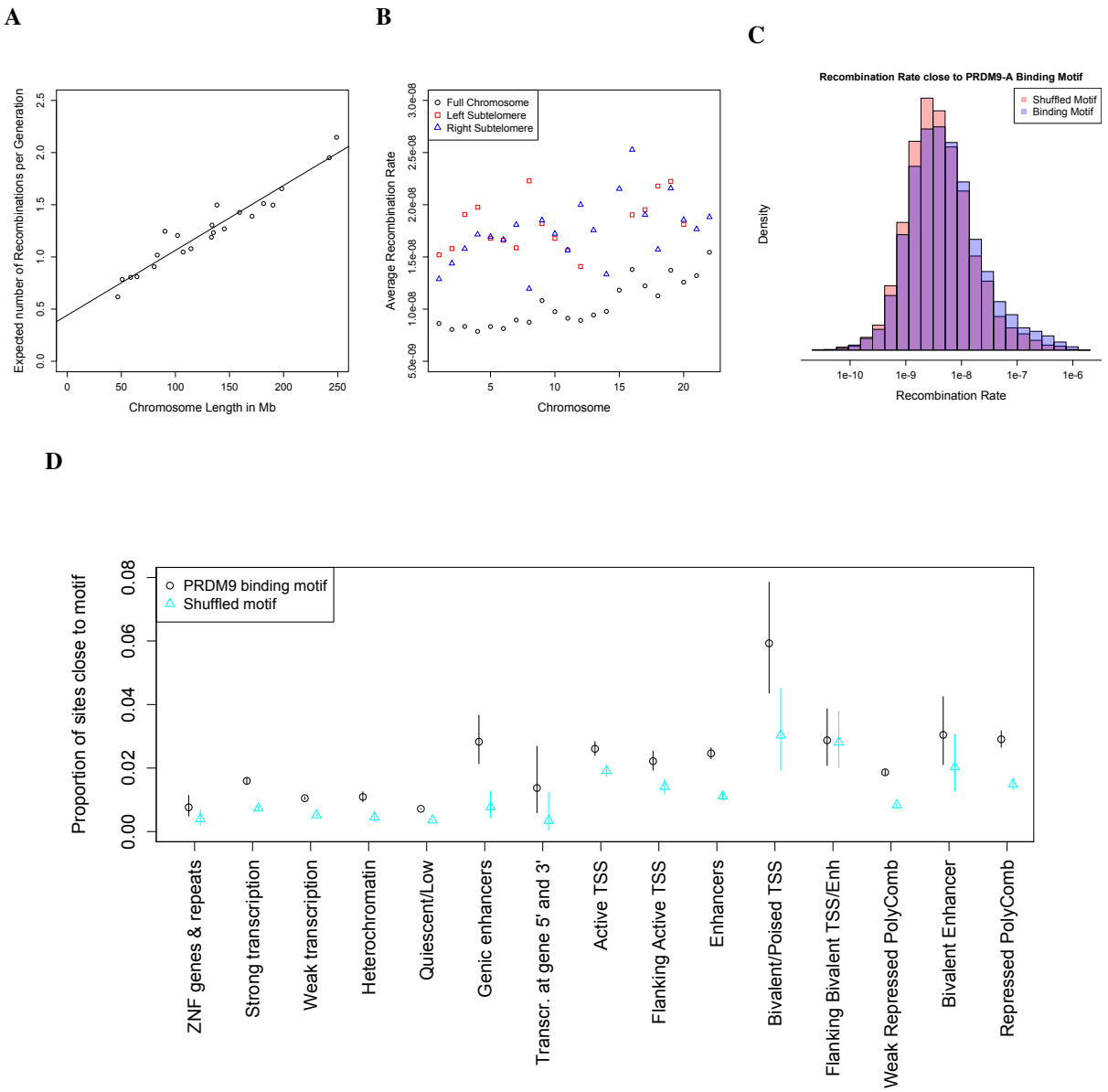

Fig. S4. Modulators of fine-scale recombination rates.

(**A**) Expected number of recombinations by chromosome length. The non-zero intercept suggests that there is a minimum number of crossovers required per meiosis, but the positive slope suggests that longer chromosomes can have more than this minimal number. (**B**) Subtelomeres show elevated rates of recombination on all chromosomes. (**C**) Regions within 100bp of a PRDM9-A binding motif have higher recombination rates on average, but the effect explains only a small amount of the variation in recombination rate. (**D**) Sites in some chromatin states are far more likely to be within 100bp of a PRDM9-A binding motif.

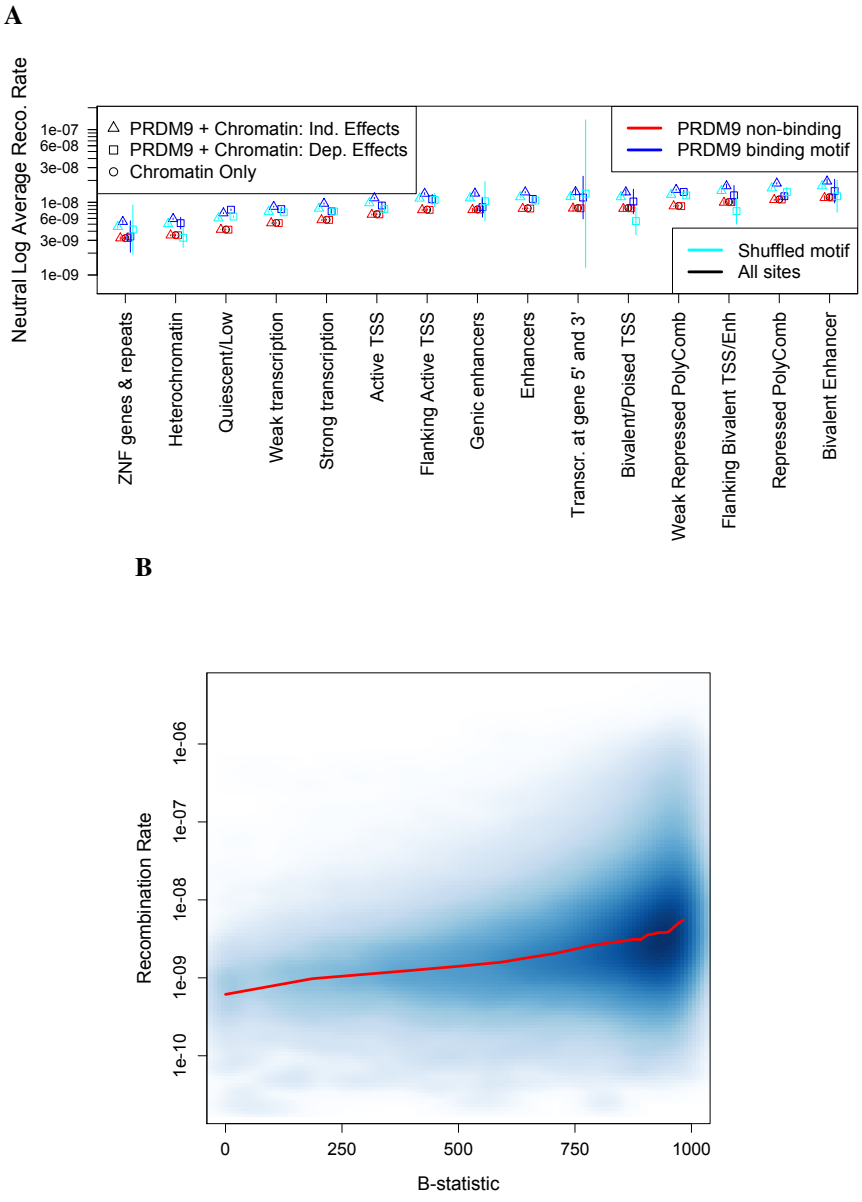

Fig. S5. Interplay of background selection and inferred recombination rates.

(**A**) Inferred regression coefficients after controlling for B-statistics. The results are comparable to Figure 4 in terms of relative ordering and relative effect size. ZNF: Zinc Finger Genes, TSS: Transcription Start Site. (**B**) Inferred recombination rates as a function of background selection, presented as a smoothed scatter plot with a local average.

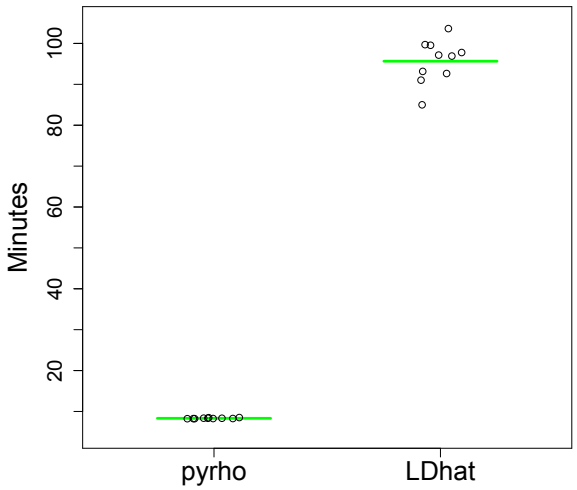

Fig. S6. Comparison of runtime for `pyrho` and `LDhat`.

Time in minutes to infer a recombination map for a simulated chromosome 1 with 196 haploids, using 32 cores for 10 replicate simulations.

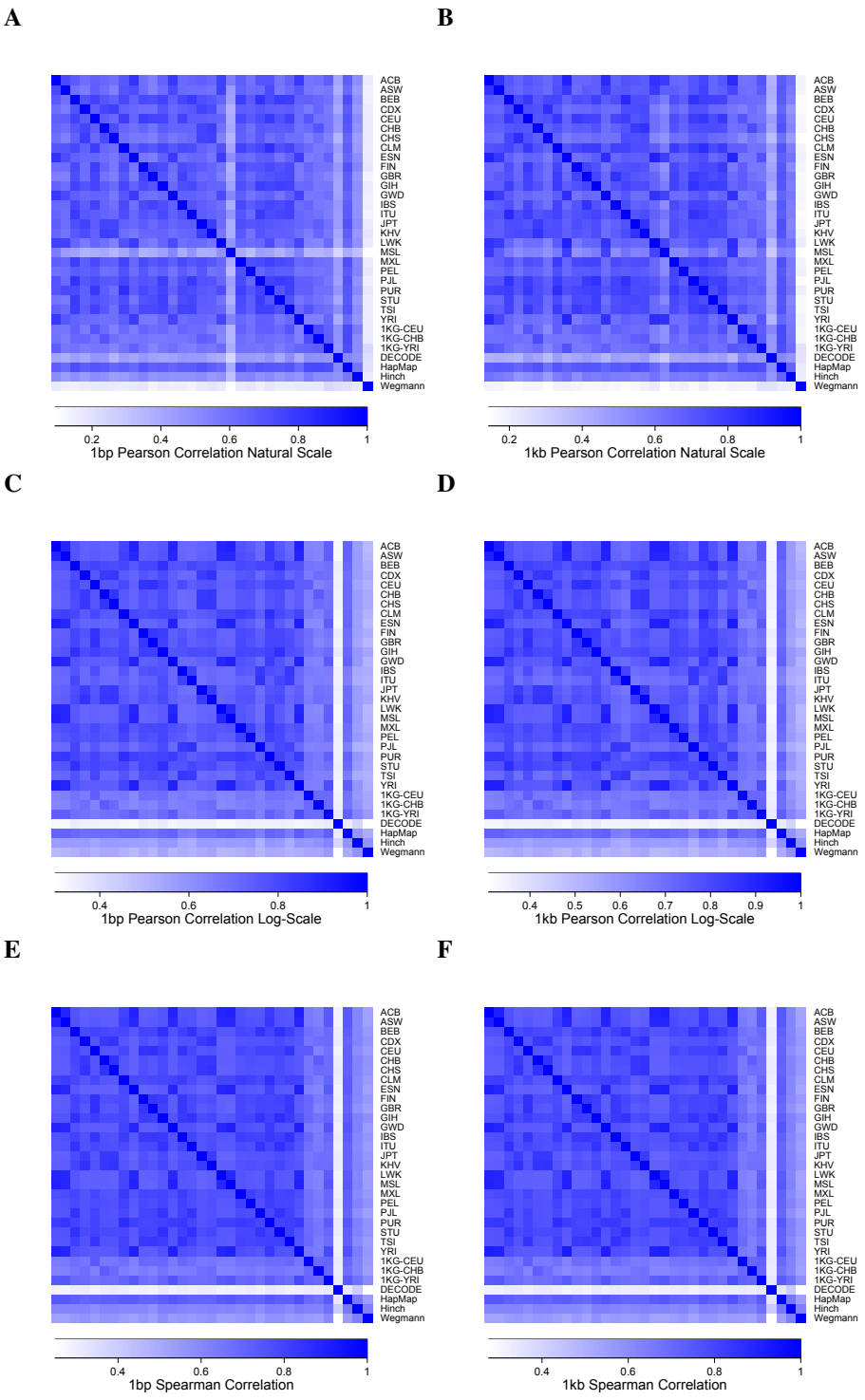

Fig. S7. Correlation between maps inferred by `pyrho` and maps inferred by previous methods.

Results inferred by `pyrho` for each population are indicated by their three letter population code (see Table 1). Population specific 1KG maps are described in (*15*); DECODE is the sex-averaged DECODE recombination map (*28*); HapMap is described in (*34*); Hinch in (*26*); Wegmann in (*27*). Each panel is a different measure of correlation (Pearson in natural and log scales and Spearman) at a different scale (1 bp or 1 kb).

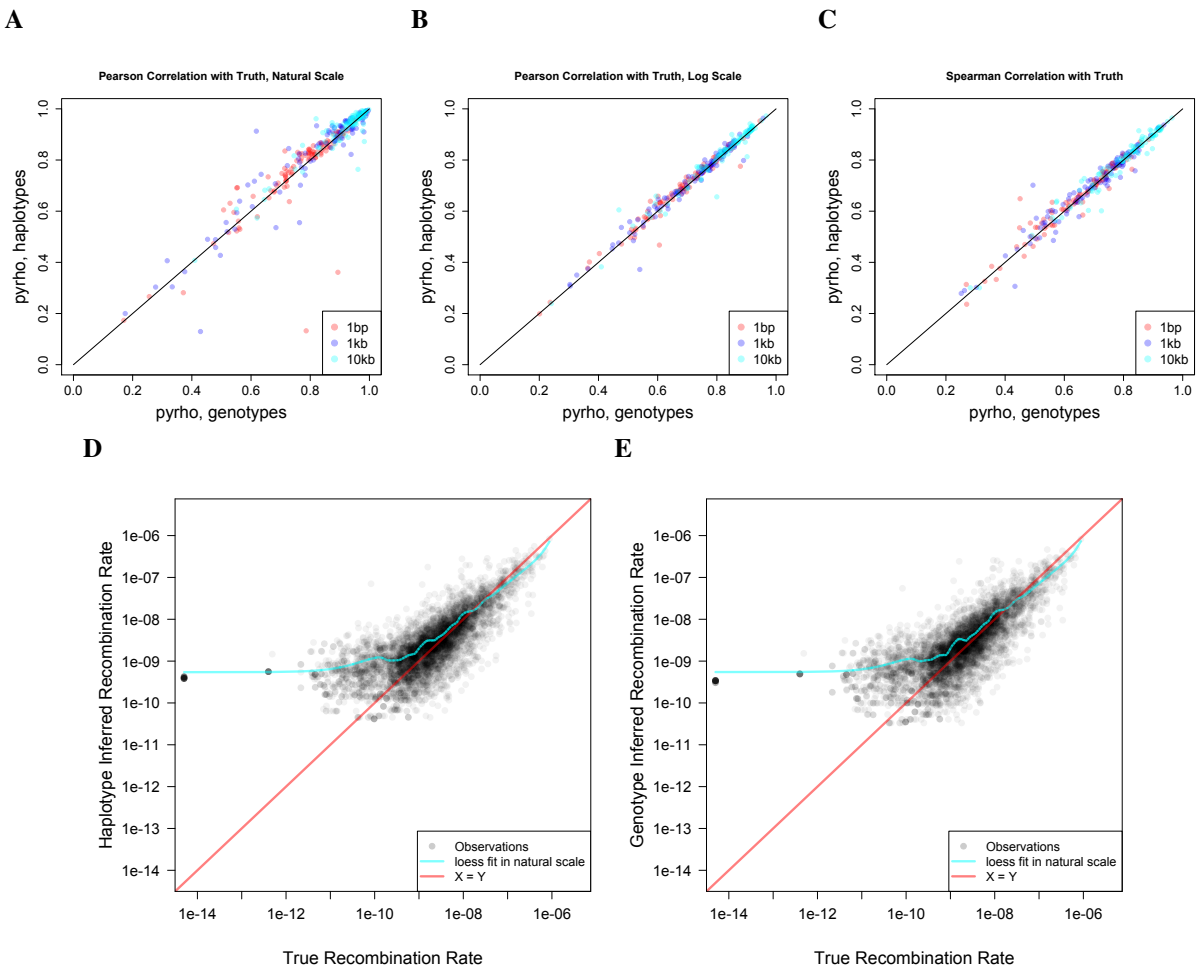

Fig. S8. Accuracy on simulations of `pyrho` using phased and unphased data.

Accuracy is measured in terms of (**A**) Pearson correlation with the truth in natural scale, (**B**) Pearson correlation with the truth in log scale, and (**C**) Spearman correlation with the truth. (**D**) Scatter plot of the inferred recombination rates using phased data at a 1bp resolution. (**E**) Scatter plot of the inferred recombination rates using unphased data at a 1bp resolution. There is little bias except at the smallest recombination rates when using either phased or unphased data.

Table S1. Correlation on simulated data at different spatial resolutions.

The mean correlation across 100 simulations ($\pm$ 2 standard errors) is reported, and the best performing method for each measure of accuracy is presented in boldface. We present results for our method, `pyrho`, as well as for `LDhat` using a demography-aware lookup table ($\texttt{LDhat}_{\text{demo}}$) or assuming a constant demography ($\texttt{LDhat}_{\text{const}}$). Overall, the methods that take demography into account outperform $\texttt{LDhat}_{\text{const}}$, and `pyrho` substantially outperforms $\texttt{LDhat}_{\text{demo}}$ at fine-scales, and performs comparably at broader scales.

Corr.	`pyrho`	$\texttt{LDhat}_{\text{demo}}$	$\texttt{LDhat}_{\text{const}}$
Pear., 1bp	$\mathbf{0.756 \pm 0.032}$	0.536 ± 0.033	0.471 ± 0.034
Pear., 1kb	$\mathbf{0.816 \pm 0.039}$	0.778 ± 0.039	0.746 ± 0.043
Pear., 10kb	0.916 ± 0.020	$\mathbf{0.921 \pm 0.018}$	0.864 ± 0.026
Pear., log-scale, 1bp	$\mathbf{0.687 \pm 0.013}$	0.627 ± 0.012	0.610 ± 0.013
Pear., log-scale, 1kb	$\mathbf{0.713 \pm 0.014}$	0.669 ± 0.124	0.648 ± 0.013
Pear., log-scale, 10kb	$\mathbf{0.811 \pm 0.012}$	0.808 ± 0.010	0.791 ± 0.011
Spear., 1bp	$\mathbf{0.659 \pm 0.014}$	0.605 ± 0.014	0.597 ± 0.014
Spear, 1kb	$\mathbf{0.689 \pm 0.014}$	0.653 ± 0.014	0.640 ± 0.014
Spear, 10kb	0.794 ± 0.012	$\mathbf{0.807 \pm 0.012}$	0.799 ± 0.012

Table S2. Effect of genome build.

Correlation between maps inferred on hg38 and those inferred on hg19 and lifted over to hg38. Pearson correlation is denoted by r with the subscript denoting whether it is in log-scale or natural scale. Spearman correlation is denoted by ρ. The amount of smoothing performed is denoted by the superscript. Population codes are listed in Table 1

Population	$r^{1bp}_{\text{nat. scale}}$	$r^{1kb}_{\text{nat. scale}}$	$r^{1bp}_{\text{log scale}}$	$r^{1kb}_{\text{log scale}}$	ρ^{1bp}	ρ^{1kb}
ACB	0.814	0.818	0.997	0.997	0.997	0.997
ASW	0.877	0.859	0.997	0.997	0.997	0.997
BEB	0.855	0.867	0.986	0.987	0.987	0.987
CDX	0.958	0.967	0.988	0.988	0.989	0.989
CEU	0.828	0.826	0.991	0.991	0.992	0.992
CHB	0.863	0.843	0.984	0.985	0.986	0.986
CHS	0.916	0.925	0.985	0.985	0.986	0.986
CLM	0.840	0.825	0.991	0.990	0.992	0.991
ESN	0.742	0.748	0.998	0.998	0.998	0.998
FIN	0.986	0.987	0.990	0.990	0.990	0.990
GBR	0.980	0.982	0.990	0.990	0.990	0.990
GIH	0.916	0.913	0.989	0.990	0.990	0.990
GWD	0.492	0.499	0.994	0.994	0.994	0.995
IBS	0.920	0.904	0.988	0.988	0.990	0.990
ITU	0.787	0.799	0.985	0.986	0.986	0.986
JPT	0.904	0.900	0.987	0.987	0.988	0.988
KHV	0.800	0.807	0.989	0.989	0.989	0.989
LWK	0.945	0.944	0.996	0.996	0.996	0.997
MSL	0.693	0.585	0.996	0.997	0.997	0.997
MXL	0.984	0.942	0.991	0.990	0.991	0.991
PEL	0.961	0.885	0.986	0.985	0.987	0.986
PJL	0.928	0.914	0.992	0.984	0.990	0.991
PUR	0.836	0.837	0.994	0.994	0.995	0.995
STU	0.474	0.475	0.984	0.984	0.986	0.986
TSI	0.925	0.895	0.987	0.986	0.987	0.987
YRI	0.920	0.927	0.994	0.994	0.995	0.995