1   **Fine human genetic map based on UK10K data set**

2

3   Ziqian Hao[1,4], Pengyuan Du[1,4], Yi-Hsuan Pan[2,5], Haipeng Li[1,3,5]

4

5   [1]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and

6   Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences,

7   Shanghai 200031, China.

8   [2]Key Laboratory of Brain Functional Genomics of Ministry of Education, School of

9   Life Science, East China Normal University, Shanghai 200062, China.

10   [3]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of

11   Sciences, Kunming, China.

12   [4]These authors contributed equally to this work.

13   [5]Corresponding authors:

14   E-mail: yxpan@sat.ecnu.edu.cn (YHP), lihaipeng@picb.ac.cn (HL)

15

16   **Abstract**

17   Recombination is a major force that shapes genetic diversity. Determination of

18   recombination rate is important and can theoretically be improved by increasing the

19   sample size. However, it is challenging to estimate recombination rates when the

20   sample size is extraordinarily large because of computational burden. In this study, we

21   used a refined artificial intelligence approach to estimate the recombination rate of the

22   human genome using the UK10K human genomic dataset with 7,562 genomic

23   sequences and its three subsets with 200, 400 and 2,000 genomic sequences under the

24   Out-of-Africa demography model. We not only obtained an accurate human genetic

25   map, but also found that the fluctuation of estimated recombination rate is reduced

26   along the human genome when the sample size is increased. UK10K recombination

27   activity is less concentrated than its subsets. Our results demonstrate how the sample

28   size affects the estimated recombination rate, and analyses of a larger number of

29   genomes result in a more precise estimation of recombination rate.

30

31

32   **Key words:** Big data; artificial intelligence; recombination rate; UK10K; FastEPRR

33

2

Introduction

Recombination is a foundation of evolution[1]. The study of recombination helps to decipher mechanisms of linkage disequilibrium[2-4], nucleotide diversity[5;6], population admixture and differentiation[7-10], natural selection[11-14], and diseases[15;16]. Therefore, much attention has been paid to the determination of the recombination rates of various regions and identification of recombination hotspots in the human genome. Since pedigree studies[17;18] and sperm-typing studies[19-21] are technically challenging, indirect approaches, such as those using population genetic methods[22-26], are more commonly used. Population genetic methods estimate recombination rates based on variations in genomic DNA sequences of a selected population. The rationale behind these indirect approaches is the assumption that coalescent tree length and the number of recombination events occurred in the sample during evolution can be (indirectly) inferred.

According to the coalescent theory, the expected coalescent tree length and number of recombination events in the sample increases when the number of genomes (sample size) analyzed is increased. It has been suggested that the inference accuracy of recombination rates can be improved by using a larger sample size[27-30]. However, it is technically challenging to estimate recombination rates when the sample size is extraordinarily large because of computational burden. In this study, we refined FastEPRR[27], an extremely fast artificial-intelligence-based software, to estimate the recombination rates of various regions in the human genome using the UK10K data set ($n = 7,562$ genomic sequences)[31] and its three subsets with 200, 400 and 2,000 genomic sequences under the Out-of-Africa demographic model[32]. We found the smaller fluctuation of local recombination rate along the human genome and less concentrated recombination activity as the sample size was increased. Furthermore, our findings are robust with the assumption of demographic history since the same conclusion was made under the constant population size model. Therefore, the large sample size improves the accuracy of estimated recombination rate, and a fine human genetic map is obtained.

64   **Materials and Methods**

65   **Fine-tuning FastEPRR**

66   We have recently developed the software "Fast Estimation of Population

67   Recombination Rates" (FastEPRR) using an artificial intelligence method to estimate

68   recombination rate (ρ) based on single nucleotide polymorphism (SNP) data and the

69   finite-site model[27]. FastEPRR first calculates compact folded site frequency spectrum

70   (SFS)[33] and four recombination-related statistics for sliding windows. A training data

71   set is then generated using a modified version of Hudson's *ms* simulator[34], which

72   generates samples conditional on the compact folded SFS under neutral models.

73   Subsequently, the gamboost package[35] is used to fit the training set and build the

74   regression model. Recombination rate (ρ) is then inferred by the summary statistics of

75   the real data. FastEPRR has been implemented in the R software package[36] and can

76   estimate recombination rate under different demographic models.

77   In this study, the FastEPRR software was fine tuned. To optimize speed, for-loops

78   were replaced by "apply" family of functions because R language is greatly effective

79   in carrying out vectorized operations. Input and output operations were minimized to

80   increase the speed. Another modification was to enable the use of different ρ values

81   to generate training samples, determination of the number of independent training

82   samples and estimation of confidence interval. All of these improvements and updates

83   have been incorporated in FastEPRR 2.0.

84

85   **Validation of FastEPRR 2.0 with a very large sample size**

86   The neutral data was simulated by using Hudson's *ms* simulator[34], a software

87   widely used in the studies of humans and other species. It generates simulated

88   polymorphic data set conditional on mutation and recombination rate under neutral

89   demographic scenarios, according to Kingman's coalescent process[37]. The following

90   parameters were used: θ (population mutation rate) $= 4N\mu = 30$; ρ (population

91   recombination rate) $= 4Nr = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130,$

92   140, and 150; and $n = 8{,}000$; where N is effective population size, μ is mutation

93   rate per chromosomal region per generation, and r is recombination rate per

4

94   chromosomal region per generation. For each ρ, the mean and the standard deviation

95   were determined using 100 simulated data with specific ρ and θ.

96   To determine the accuracy in the estimation of recombination rates on potential

97   recombination hotspot regions, the simulated data were established using a set of large

98   ρ values (i.e., 500, 1000, 2000, 3000, 4000, and 5000). The mean and the standard

99   deviation were determined using 10 simulated data for each ρ value.

100

101   **Building the UK10K genetic map**

102   A total of 7,562 genomes (from 3,781 unrelated individuals) in the UK10K data

103   set were analyzed. To build genetic maps, the window size was set as 10 kb, and the

104   step length was 5 kb. Indels were removed. Windows were excluded if they

105   overlapped with known sequencing gaps in the human reference genome (hg19) or the

106   number of non-singleton polymorphic sites was smaller than 10. The default value set

107   of ρ (0, 0.5, 1, 2, 5, 10, 20, 40, 70, 110, 170, 180, 190, 200, 220, 250, 300, and 350)

108   was used to simulate the training sets.

109   To estimate recombination rates, the Out-of-Africa model was used[32]. To evaluate

110   the effect of demographic history, the constant size model was also considered. The

111   linear extrapolation was employed if the estimated value is beyond the training set

112   boundary. To ensure the accuracy of results, the default training set values of ρ were

113   adjusted, and FastEPRR was re-ran until the estimated values fell into the range of the

114   given set of ρ. The previous method in FastEPRR was used to evaluate the effects for

115   variable recombination rates within windows[28]. Three data sets with 200

116   (UK10K-sub200), 400 (UK10K-sub400) and 2,000 (UK10K-sub2000) genomes

117   randomly chosen from the UK10K data set were also analyzed.

118

119   **Results**

120   FastEPRR has been shown to perform well with different sample sizes ($n = 50$,

121   100, 200 and 1,000 chromosomes), and the standard deviation becomes smaller when

122   the sample size is increased[27]. However, its effectiveness on the analysis of a large

123   sample size (e.g., $n = 7,562$ chromosomes) has not been determined. Therefore, the

124 accuracy of recombination rate estimated by FastEPRR was examined when the

125 sample size is extremely large (Figure 1A and 1B). The FastEPRR estimate is

126 unbiased, and its standard deviation becomes much smaller. Thus the inference

127 accuracy of recombination rate is dramatically improved when a greatly large sample

128 size is available.

129 We previously examined the cases $20 \leq \rho \leq 150$ and found that the standard

130 deviation of estimated $\hat{\rho}_{FastEPRR}$ was smaller than that of $\hat{\rho}_{LDhat}$[27], suggesting that

131 FastEPRR estimated the recombination rate more precisely than LDhat when

132 recombination rate was not extremely low. However, it has been shown that certain

133 chromosomal regions, such as recombination hotspots, have higher recombination

134 rates than other regions. To assess the effectiveness of FastEPRR on these regions, the

135 accuracy to infer high recombination rate was re-examined. Analysis of chromosomal

136 regions with higher recombination rates (ρ ranging from 500 to 5,000) revealed that

137 FastEPRR were precise and unbiased on potential recombination hotpot regions

138 (Figure 1C).

139 FastEPRR provides the ability to infer recombination rate under different

140 demographic models[27], and it has been reported that demography affects estimated

141 recombination rate[38-40]. Thus in this study we estimated the recombination rates of

142 genomes in UK10K conditional on the Out-of-Africa model[32]. For the i-th

143 chromosomal region, we need to estimate N_0 first to calculate r_i from ρ_i ($=$

144 $4N_0 r_i$). Therefore, the average ρ ($= 4N_0 r$) per Mb of 22 autosomes estimated by

145 FastEPRR was 12,098.2, where N_0 is the effective population size at present. As the

146 average r of the same chromosomal regions in the 2019 deCODE family-based

147 genetic map has been determined to be 1.2806 cM/Mb[18], N_0 was estimated as

148 235,186 for the UK10K data set under the Out-of-Africa model, in the range of

149 estimated value in previous studies[32;41-43]. This value is 20.55 larger than the value

150 previously estimated for the CEU population based on 1000 genomes in the OMNI

151 CEU data set[44] by LDhat[25]. This difference is reasonable because the current analysis

152 was based on the Out-of-Africa model, whereas the previous one was based on the

153 constant size model.

6

154   To compare the newly established genetic map with others, the Pairwise Pearson

155   correlation coefficients between the UK10K map and 2019 deCODE family-based

156   genetic map and the map of 1000 Genomes in the OMNI CEU data set were

157   calculated at the 5-Mb and 1-Mb scales (Figure 2 and 3). The correlation coefficients

158   were determined to be 0.8413 and 0.7256 at the 5-Mb scale, and 0.7308 and 0.5267 at

159   the 1-Mb scale, indicating that the UK10K map is highly similar to the other two

160   genetic maps.

161   Genetic maps were also constructed under the Out-of-Africa model for the three

162   subsets (UK10K-sub200, UK10K-sub400 and UK10K-sub2000) with 200, 400 and

163   2,000 genomes randomly selected from UK10K. The Pearson correlation coefficients

164   between the UK10K and UK10K-sub200 maps were determined to be 0.9266 and

165   0.8216 at the 5- and 1-Mb scales, respectively, and those between the UK10K and

166   UK10K-sub400 and UK10K-sub2000 maps were 0.9393 and 0.9658 at the 5-Mb scale,

167   and 0.8501 and 0.9097 at the 1-Mb scale, respectively (Figure 2 and 3). These results

168   indicate that the UK10K map and the maps of the three subsets are highly correlated

169   with each other.

170   Recombination rates over different physical scales show that smaller scale has

171   higher resolution (Figure 4), consistent with the previous finding[1]. To compare the

172   distribution of recombination activities in different genetic maps, the recombination

173   rate at the 5-kb scale was used. The recombination rates in chromosome 1 are shown

174   as an example in Figure 5. Two regions (chr1:15,425,001-15,430,000 and

175   chr1:236,680,001-236,685,000) were found to have higher recombination rates with

176   seven PRDM9 (PR domain zinc finger protein 9) binding peaks[45], indicative of the

177   ability of FastEPRR to detect potential recombination hotspots. Since the human

178   leukocyte antigen (HLA) region is one of the most polymorphic and recombined

179   regions in human genome[46-49], it was analyzed and multiple recombination hotspot

180   candidates, hotter than the previous studies, were detected (Figure 6).

181   Although recombination hotspots candidates are revealed in the UK10K genetic

182   map, the distribution of recombination events is generally less zigzag than other

183   human genetic maps (Figure 5). The variance of recombination rate was then

184   determined and found to be 3.4477 cM^2/Mb^2 in UK10K, 4.7224 cM^2/Mb^2 in

185   UK10K-sub2000, 7.3097 cM^2/Mb^2 in UK10K-sub400, 7.9453 cM^2/Mb^2 in

186   UK10K-sub200, 17.9188 cM^2/Mb^2 in OMNI CEU LDhat, and 25.7734 cM^2/Mb^2 in

187   deCODE (calculated at the 5-kb scale). The variance of recombination rate in the

188   UK10K genomes is the smallest. The proportion of total recombination in various

189   percentages of the human genome was plotted and analyzed (Figure 7A). Overall,

190   recombination activity in the UK10K genomes is much less concentrated than those

191   of the genomes in other data sets. When the sample size was smaller than or equal to

192   400 chromosomes, 54.88-82.59% of crossover events were observed in 10% of the

193   human genome. This observation is consistent with previous findings[18;27;50]. Only

194   33.82% of crossover events were detected in 10% of the human genome in UK10K.

195   This percentage is strikingly lower than that of previous studies[18;27;50]. This difference

196   is likely due to the more precise estimation of recombination rates with an

197   extraordinarily large sample size.

198   To valid our conclusions, the recombination rates were also estimated conditional

199   on the constant size model. The Pearson correlation coefficient between the estimated

200   recombination rate under the Out-of-Africa model and that under the constant model

201   was only 0.7893 at the 5-kb scale, indicating that demography affects the estimation

202   of recombination rates[27;38-40]. However, the newly estimated recombination rates are

203   still less variable with larger sample size (Figure 7B), indicating that our conclusion is

204   robust with the assumption of demography.

205   The reason why large sample yielded more accurate results was investigated. As

206   more recombination events and mutations (equivalent to SNPs) occur on a tree with a

207   longer tree length, both factors may affect the determination of recombination rates.

208   To determine the effect of SNP numbers, two cases were investigated. In the first case

209   (Case01), 200 genomes were selected from the samples in UK10K-sub400. In the

210   second case (Case02), all 400 genomes in UK10K-sub400 were considered, however,

211   only the SNPs in the Case01 were analyzed. In other words, these two cases had the

212   same number of SNPs but had different sample sizes as well as different

213   recombination events. Results showed that the estimated recombination rate of

214   Case02 is slightly more similar to that of UK10K-sub400 than that of the Case01

215   (Correlation coefficient 0.9770 *vs* 0.9609 at the 5-kb scale). The variance of

216   recombination rate of Case02 is also slightly smaller than that of Case01 (7.2850

217   cM^2/Mb^2 *vs* 7.5130 cM^2/Mb^2, calculated at the 5-kb scale). Similar results were

218   obtained with the constant size model (Correlation coefficient 0.9692 *vs* 0.9562, and

219   variance 10.5054 cM^2/Mb^2 *vs* 14.9622 cM^2/Mb^2, calculated at 5-kb scale). Therefore,

220   more recombination events observed in data set with large sample size, rather than

221   more SNPs, result in a more precise estimate of recombination rate.

222

Discussion

224   In this study, we employed the software FastEPRR to estimate the recombination

225   rate of various regions of the human genome using datasets with different sample

226   sizes. These datasets included UK10K, UK10K-sub2000, UK10K-sub400 and

227   UK10K-sub200 with 7,562, 2,000, 400, and 200 genomes, respectively. The mean

228   recombination rate in these four datasets is determined to be the same but their

229   standard deviation increases when the sample size is reduced. This result showed that

230   the larger the sample size for the analysis, the smaller fluctuation of estimated

231   recombination rate, and the less concentrated recombination activities of the human

232   genome.

233   These findings are robust with the assumption of demographic models. Under the

234   constant population size and the Out-of-Africa models, the same conclusions were

235   made. These findings are also consistent with the expectation of coalescent theory,

236   which indicates that the coalescent tree length of a larger sample is longer than that of

237   a smaller sample. The genomes of the former thus contain more recombination events

238   than those of the latter, leading a more precise estimate of recombination rate. In our

239   study, the sample size is not very small comparing with the long-term human effective

240   population size (~10,000). Although Hudson's *ms* simulator requires that the sample

241   size is much smaller than the effective population size[37], our findings are robust with

242   this assumption. For the exact coalescent[51] and the Moran model[52], given a true tree

243   with the sample size $(n-1)$, the tree with the sample n can be obtained by adding

244 a tip with a non-zero length, and thus the tree length increases. The same conclusion

245 can be made under different demographic histories. Since FastEPRR is currently the

246 only method which is fast enough to handle a super-large dataset, we cannot confirm

247 the findings by using alternative methods. However, it has been shown that when the

248 sample size was increased from 50 to 200 [27;28], the variance of recombination rate

249 estimated by LDhat[25], summary statistic likelihood (SSL)[53] and product of

250 approximate conditionals (PAC)[54] was reduced. As our results generated by FastEPRR

251 resemble those of these alternative methods, it is conceivable our findings represent a

252 general phenomenon in humans and other sexually reproduced species.

253 This study is the first attempt to estimate recombination rate using an

254 extraordinarily large population genetic data set. By processing genomes with a

255 sample size of four orders of magnitude, we demonstrated that artificial intelligence

256 approaches[28;55-57] are effective for population genetic analysis. The results suggest that

257 the fluctuation of recombination rate of human genome may be overestimated in

258 previous studies.

259

Data and Code Availability

261 The datasets used in this study are available at the UK10K Project Consortium[31].

262 The genetic maps of OMNI data set built by LDhat[25] were downloaded from the 1000

263 Genomes Project. FastEPRR 2.0 is written in R and integrated on the eGPS cloud[58]

264 (http://www.egps-software.net). The desktop version and the genetic maps established

265 in this study are freely available on the institute website

266 (https://www.picb.ac.cn/evolgen/). The genetic maps are also available as

267 supplemental materials.

268

Web Resources

270 FastEPRR 2.0 and UK10K genetic map, https://www.picb.ac.cn/evolgen/softwares/

271 UK10K Project Consortium, https://www.uk10k.org/

272 Genetic map of OMNI data set,

273 ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombi

274   nation_rates

275   eGPS software, http://www.egps-software.net

276

Acknowledgments

283

References

1. Coop, G., and Przeworski, M. (2007). An evolutionary view of human recombination. Nat Rev Genet 8, 23-34.

2. Hill, W.G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38, 226-231.

3. Ohta, T., and Kimura, M. (1971). Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. Genetics 68, 571-580.

4. Ardlie, K.G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3, 299-309.

5. Francioli, L.C., Polak, P.P., Koren, A., Menelaou, A., Chun, S., Renkens, I., van Duijn, C.M., Swertz, M., Wijmenga, C., van Ommen, G., et al. (2015). Genome-wide patterns and properties of de novo mutations in humans. Nat Genet 47, 822-826.

6. Kong, A., Thorleifsson, G., Frigge, M.L., Masson, G., Gudbjartsson, D.F., Villemoes, R., Magnusdottir, E., Olafsdottir, S.B., Thorsteinsdottir, U., and Stefansson, K. (2014). Common and low-frequency variants associated with genome-wide recombination rate. Nat Genet 46, 11-16.

7. Payseur, B.A., and Rieseberg, L.H. (2016). A genomic perspective on hybridization and speciation. Mol Ecol 25, 2337-2360.

8. Keinan, A., and Reich, D. (2010). Human population differentiation is strongly correlated with local recombination rate. Plos Genet 6, e1000886.

9. Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., and Myers, S. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. Plos Genet 5, e1000519.

10. Wegmann, D., Kessner, D.E., Veeramah, K.R., Mathias, R.A., Nicolae, D.L., Yanek, L.R., Sun, Y.V., Torgerson, D.G., Rafaels, N., Mosley, T., et al. (2011). Recombination rates in admixed individuals identified by ancestry-based inference. Nat Genet 43, 847-853.

11. Schumer, M., Xu, C.L., Powell, D.L., Durvasula, A., Skov, L., Holland, C., Blazier, J.C., Sankararaman, S., Andolfatto, P., Rosenthal, G.G., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science 360, 656-659.

12. Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M. (2017). Recombination: the good, the bad and the variable. Philos T R Soc B 372, 20170279.

13. Hernandez, R.D., Kelley, J.L., Elyashiv, E., Melton, S.C., Auton, A., McVean, G., Sella, G., Przeworski, M., and Project, G. (2011). Classic selective sweeps were rare in recent human evolution. Science 331, 920-924.

14. Stevison, L.S., Woerner, A.E., Kidd, J.M., Kelley, J.L., Veeramah, K.R., McManus, K.F., Bustamante, C.D., Hammer, M.F., Wall, J.D., and Project, G.A.G. (2016). The time scale of recombination rate evolution in Great Apes. Molecular Biology and Evolution 33, 928-945.

328 15. Hussin, J.G., Hodgkinson, A., Idaghdour, Y., Grenier, J.C., Goulet, J.P., Gbeha, E.,
329   Hip-Ki, E., and Awadalla, P. (2015). Recombination affects accumulation of
330   damaging and disease-associated mutations in human populations. Nat Genet
331   47, 400-404.
332 16. Weiss, K.M., and Clark, A.G. (2002). Linkage disequilibrium and the mapping of
333   complex human traits. Trends Genet 18, 19-24.
334 17. Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A.,
335   Jonasdottir, A., Walters, G.B., Jonasdottir, A., Gylfason, A., Kristinsson, K.T.,
336   et al. (2010). Fine-scale recombination rate differences between sexes,
337   populations and individuals. Nature 467, 1099-1103.
338 18. Halldorsson, B.V., Palsson, G., Stefansson, O.A., Jonsson, H., Hardarson, M.T.,
339   Eggertsson, H.P., Gunnarsson, B., Oddsson, A., Halldorsson, G.H., Zink, F., et
340   al. (2019). Characterizing mutagenic effects of recombination through a
341   sequence-level genetic map. Science 363, eaau1043.
342 19. Webb, A.J., Berg, I.L., and Jeffreys, A. (2008). Sperm cross-over activity in
343   regions of the human genome showing extreme breakdown of marker
344   association. P Natl Acad Sci USA 105, 10471-10476.
345 20. Arbeithuber, B., Betancourt, A.J., Ebner, T., and Tiemann-Boege, I. (2015).
346   Crossovers are associated with mutation and biased gene conversion at
347   recombination hotspots. P Natl Acad Sci USA 112, 2109-2114.
348 21. Bell, A.D., Mello, C.J., Nemesh, J., Brumbaugh, S.A., Wysoker, A., and
349   McCarroll, S.A. (2020). Insights into variation in meiosis from 31,228 human
350   sperm genomes. Nature 583, 259-264.
351 22. Hudson, R.R. (2001). Two-locus sampling distributions and their application.
352   Genetics 159, 1805-1817.
353 23. Fearnhead, P., and Donnelly, P. (2001). Estimating recombination rates from
354   population genetic data. Genetics 159, 1299-1318.
355 24. McVean, G., Awadalla, P., and Fearnhead, P. (2002). A coalescent-based method
356   for detecting and estimating recombination from gene sequences. Genetics 160,
357   1231-1241.
358 25. Auton, A., and McVean, G. (2007). Recombination rate estimation in the presence
359   of hotspots. Genome Res 17, 1219-1227.
360 26. McVean, G.A.T., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly,
361   P. (2004). The fine-scale structure of recombination rate variation in the
362   human genome. Science 304, 581-584.
363 27. Gao, F., Ming, C., Hu, W.J., and Li, H.P. (2016). New software for the fast
364   estimation of population recombination rates (FastEPRR) in the genomic era.
365   G3-Genes Genomes Genet. 6, 1563-1571.
366 28. Lin, K., Futschik, A., and Li, H.P. (2013). A fast estimate for the population
367   recombination rate based on regression. Genetics 194, 473-484.
368 29. Sall, T., and Nilsson, N.O. (1994). The robustness of recombination frequency
369   estimates in intercrosses with dominant markers. Genetics 137, 589-596.
370 30. Gartner, K., and Futschik, A. (2016). Improved versions of common estimators of
371   the recombination rate. Journal of Computational Biology 23, 756-768.

31. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., Lawson, D., et al. (2015). The UK10K project identifies rare variants in health and disease. Nature 526, 82-89.

32. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F.L., Gibbs, R.A., Bustamante, C.D., and Project, G. (2011). Demographic history and rare allele sharing among human populations. P Natl Acad Sci USA 108, 11983-11988.

33. Li, H.P., and Stephan, W. (2005). Maximum-likelihood methods for detecting recent positive selection and localizing the selected site in the genome. Genetics 171, 377-384.

34. Hudson, R.R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18, 337-338.

35. Hothorn, T., Buehlmann, P., Kneib, T., Schmid, M., and Hofner, B. (2018). mboost: Model-Based Boosting, R package version 2.9-1, https://CRAN.R-project.org/package=mboost.

36. R Core Team. (2019). R: A language and environment for statistical computing.

37. Kingman, J.F.C. (1982). On the genealogy of large populations. J Appl Probab 19, 27-43.

38. Spence, J.P., and Song, Y.S. (2019). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. Science Advances 5, eaaw9206.

39. Kamm, J.A., Spence, J.P., Chan, J., and Song, Y.S. (2016). Two-locus likelihoods under variable population size and fine-scale recombination rate estimation. Genetics 203, 1381-1399.

40. Dapper, A.L., and Payseur, B.A. (2018). Effects of Demographic History on the Detection of Recombination Hotspots from Linkage Disequilibrium. Molecular Biology and Evolution 35, 335-353.

41. Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet 49, 303-309.

42. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat Genet 46, 919-925.

43. Hu, W.J., Hao, Z.Q., Du, P.Y., Di Vincenzo, F., Manzi, G., Pan, Y.H., and Li, H.P. (2021). Genomic inference of a human super bottleneck in Mid-Pleistocene transition. bioRxiv 444351.

44. Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.

45. Altemose, N., Noor, N., Bitoun, E., Tumian, A., Imbeault, M., Chapman, J.R., Aricescu, A.R., and Myers, S.R. (2017). A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. eLife 6, e28383.

46. Wu, R.G., Li, H.X., Peng, D., Li, R., Zhang, Y.M., Hao, B., Huang, E.W., Zheng,

C.H., and Sun, H.Y. (2019). Revisiting the potential power of human leukocyte antigen (HLA) genes on relationship testing by massively parallel sequencing-based HLA typing in an extended family. J Hum Genet 64, 29-38.

47. Jeffreys, A.J., Kauppi, L., and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29, 217-222.

48. Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G., and Carrington, M. (2002). High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am J Hum Genet 71, 759-776.

49. Miretti, M.M., Walsh, E.C., Ke, X.Y., Delgado, M., Griffiths, M., Hunt, S., Morrison, J., Whittaker, P., Lander, E.S., Cardon, L.R., et al. (2005). A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. Am J Hum Genet 76, 634-646.

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

51. Fu, Y.X. (2006). Exact coalescent for the Wright-Fisher model. Theor Popul Biol 69, 385-394.

52. Wirtz, J., and Wiehe, T. (2019). The evolving Moran genealogy. Theor Popul Biol 130, 94-105.

53. Wall, J.D. (2000). A comparison of estimators of the population recombination rate. Molecular Biology and Evolution 17, 156-163.

54. Li, N., and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213-2233.

55. Lin, K., Li, H.P., Schlotterer, C., and Futschik, A. (2011). Distinguishing positive selection from neutral evolution: Boosting the performance of summary statistics. Genetics 187, 229-244.

56. Flagel, L., Brandvain, Y., and Schrider, D.R. (2019). The unreasonable effectiveness of convolutional neural networks in population genetic inference. Molecular Biology and Evolution 36, 220-238.

57. Pavlidis, P., Jensen, J.D., and Stephan, W. (2010). Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. Genetics 185, 907-922.

58. Yu, D.L., Dong, L.L., Yan, F.Q., Mu, H.L., Tang, B.X., Yang, X., Zeng, T., Zhou, Q., Gao, F., Wang, Z.H., et al. (2019). eGPS 1.0: Comprehensive software for multi-omic and evolutionary analyses. National Science Review 6, 867-869.
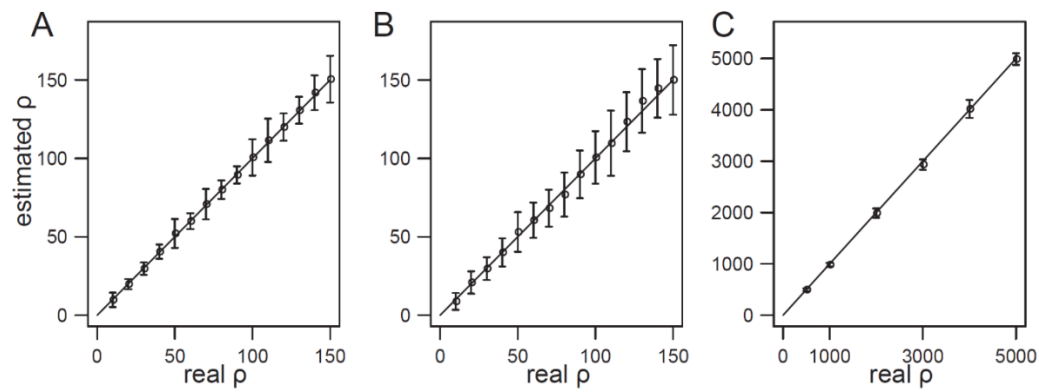
456

457   **Figure 1. Comparison between estimated** ρ **($\rho_{FastEPRR}$) and real** ρ **under**

458   **different conditions.**

459   (A) $\rho \leq 150$, $n = 8{,}000$.

460   (B) $\rho \leq 150$, $n = 200$.

461   (C) $\rho \geq 500$, $n = 8{,}000$.

462   The mean and the standard deviation of $\hat{\rho}$ in (A) and (B) were calculated from 100

463   simulated data sets, and those in (C) were calculated from 10 simulated data sets.

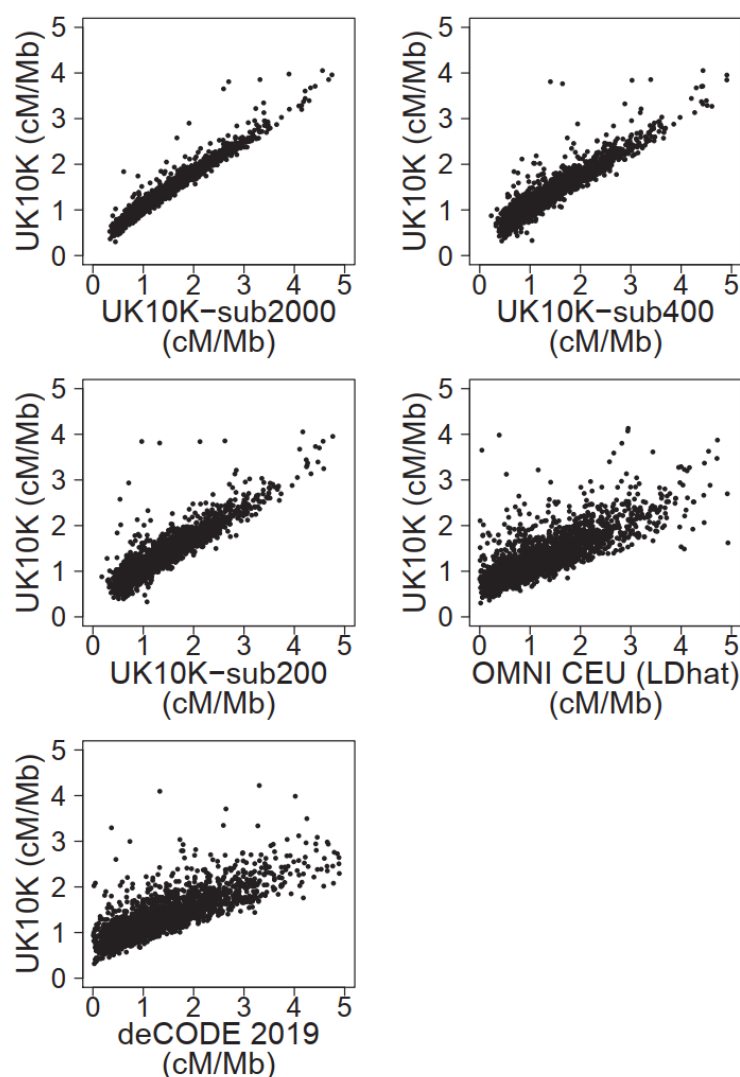464   Error bars represent the standard deviation of $\hat{\rho}$.

465

16

466
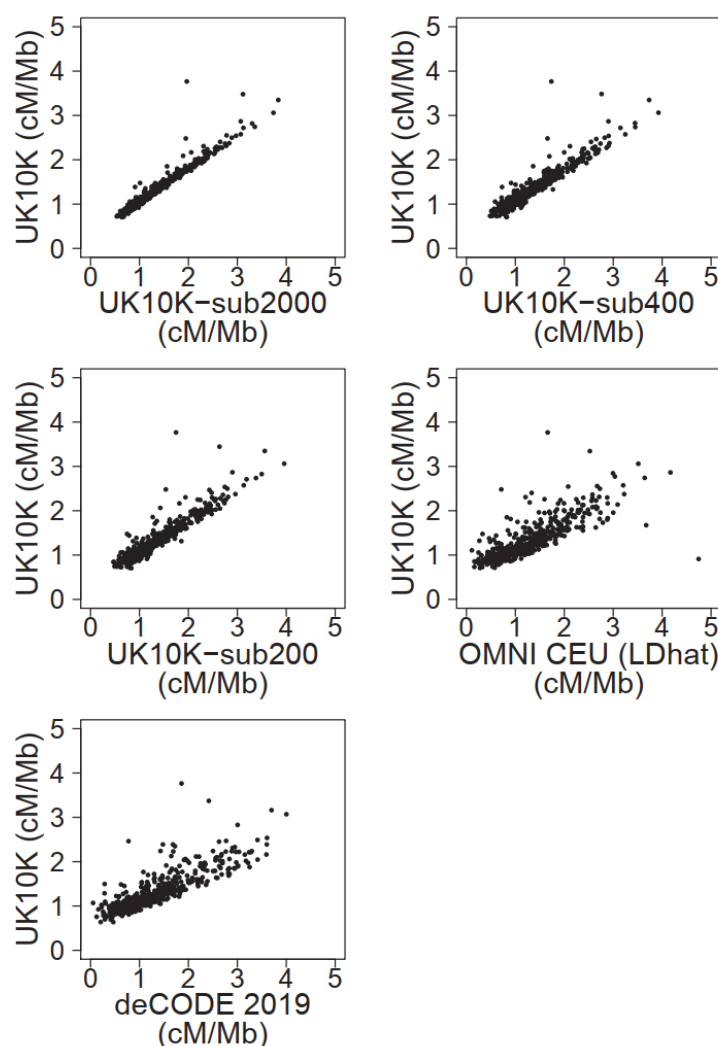
Figure 2. Correlation between the UK10K genetic map and five other maps at the 5-Mb scale.

The recombination rates of UK10K, UK10K-sub2000, UK10K-sub400 and UK10K-sub200 were estimated by FastEPRR under the Out-of-Africa model, while those of OMNI CEU genomes were estimated previously by LDhat under the constant size model[44]. Each black dot represents two recombination rates of a window estimated by two different methods.
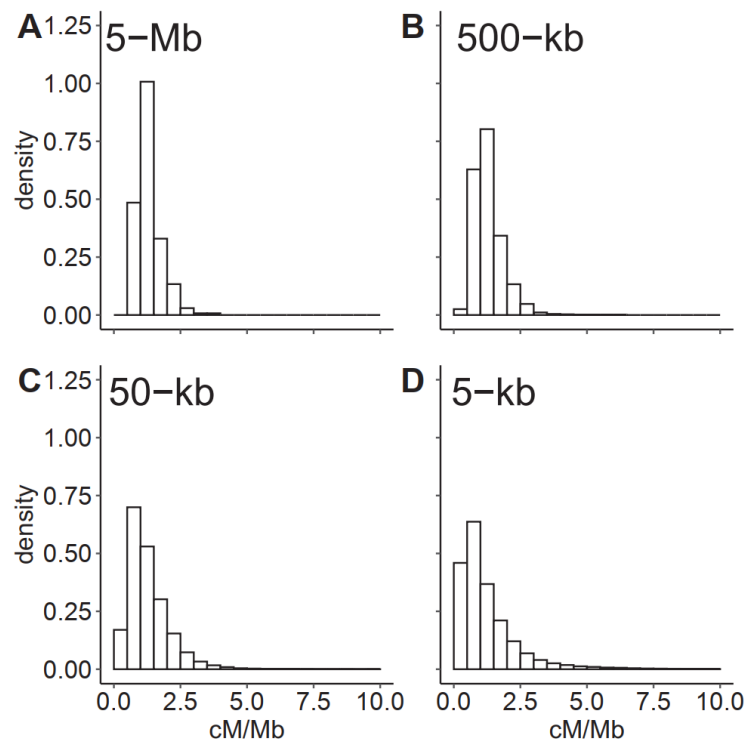
474

475

Figure 3. Correlation between the UK10K genetic map and five other maps at the 1-Mb scale.

The recombination rates of UK10K, UK10K-sub2000, UK10K-sub400 and UK10K-sub200 were estimated by FastEPRR under the Out-of-Africa model, while those of OMNI CEU genomes were estimated previously by LDhat under the constant size model[44]. Each black dot represents two recombination rates of a window estimated by two different methods.

483

Figure 4. Recombination rates of the human genome at various physical scales.

The skewed distribution indicates that the most of chromosomal regions have low recombination rate.
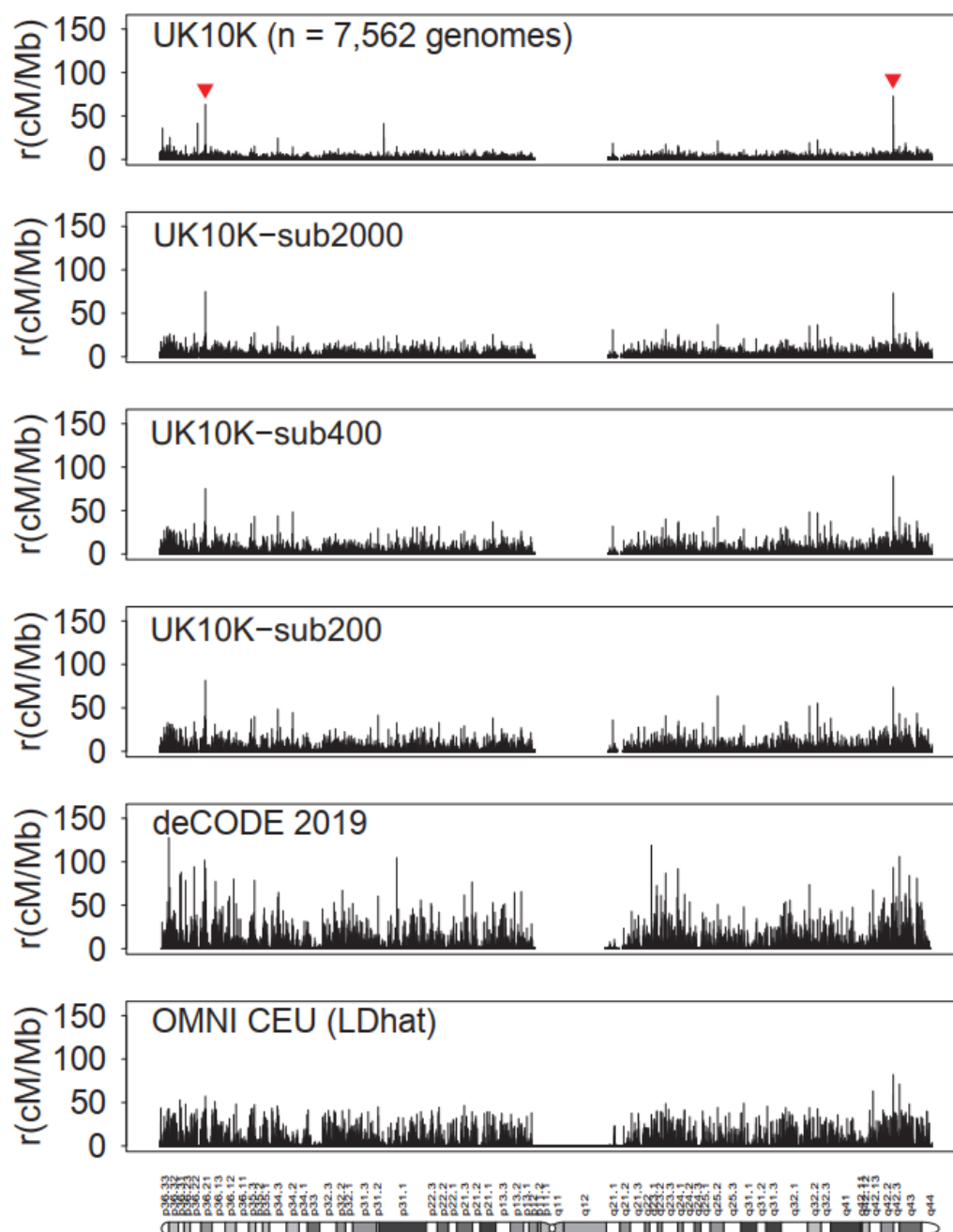
19

490

Figure 5. Recombination rates of various regions of human chromosome 1 obtained from six different genetic maps at the 5-kb scale.

The height of each spike on the maps represents the average recombination rate of a 5-kb region in chromosome 1. The recombination rates of various regions in chromosome 1 shown in the maps of UK10K, UK10K-sub2000, UK10K-sub400 and UK10K-sub200 were estimated by FastEPRR with the Out-of-Africa model. The recombination rates of chromosome 1 in deCODE 2019 and the OMNI CEU data set estimated by LDhat[44] with the constant size model are shown for comparison.
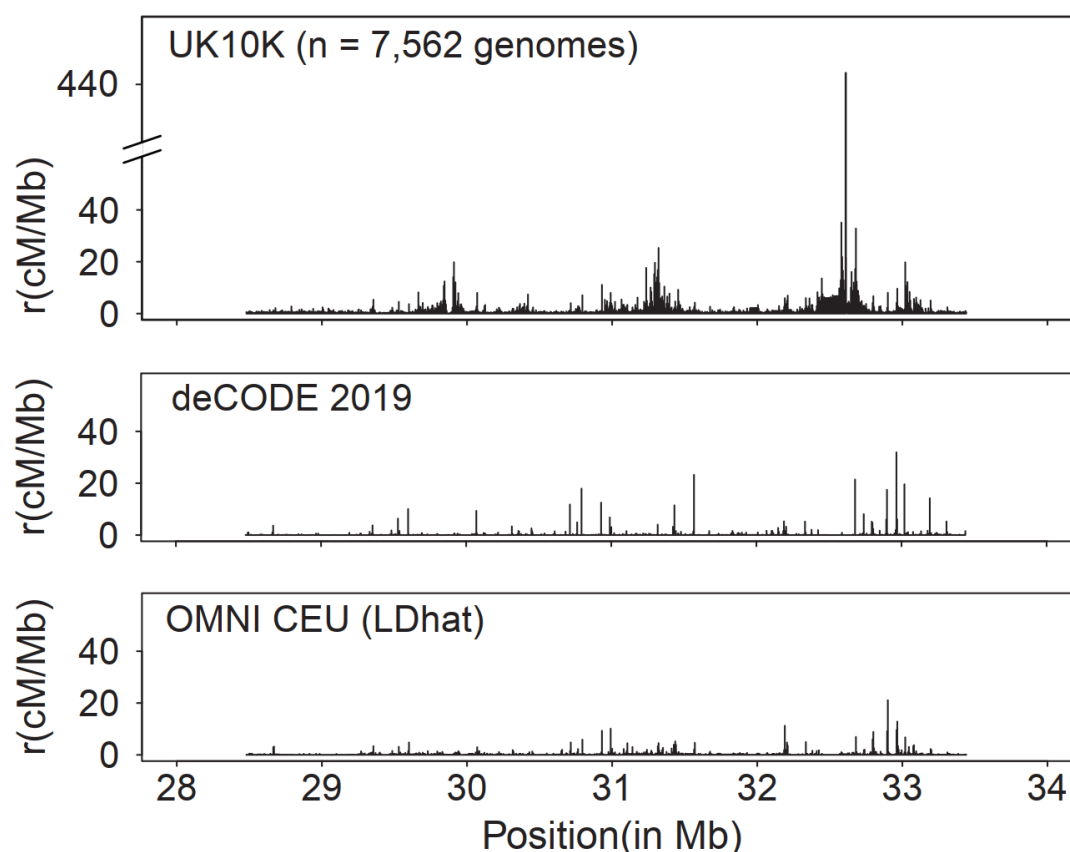
499

Figure 6. Recombination rates of various regions of the human HLA locus region estimated from three genetic maps at 5-kb scale.

The height of each spike on the maps represents the average recombination rate of a 5-kb region in the HLA locus. The recombination rates of various regions shown in the maps of UK10K were estimated by FastEPRR with the Out-of-Africa model. The recombination rates of deCODE 2019 and the OMNI CEU data set estimated by LDhat[44] with the constant size model are shown for comparison.
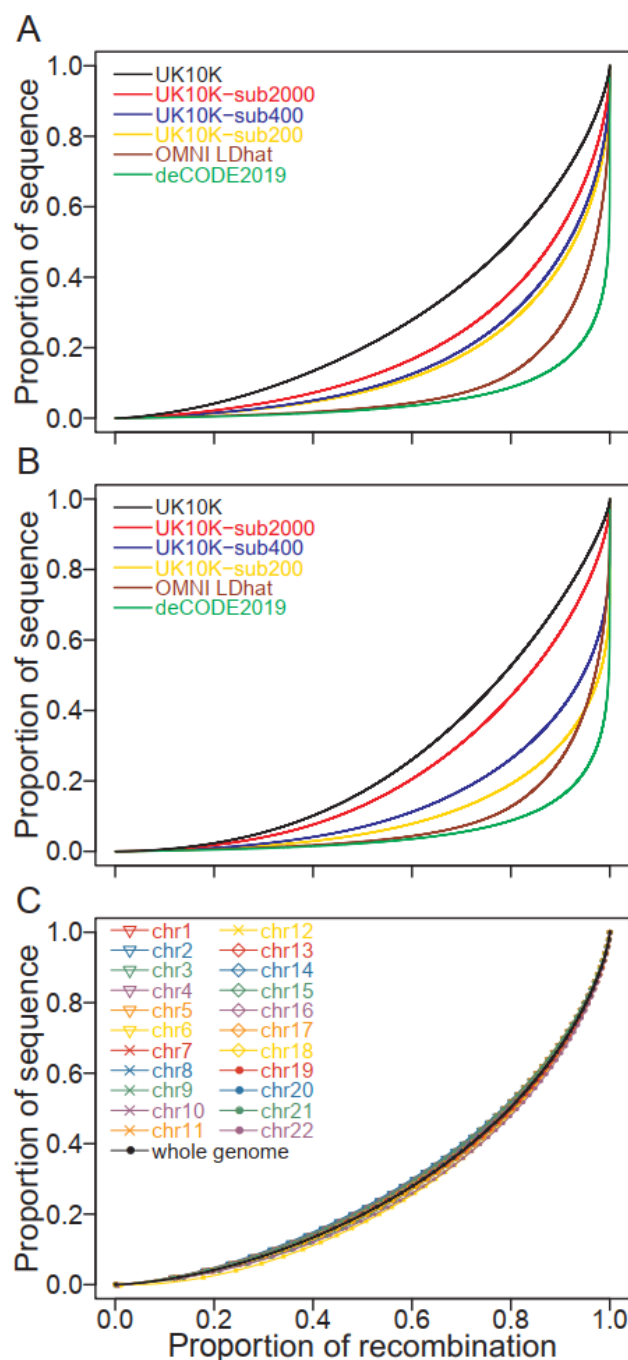
507

508

509

Figure 7. Proportion of recombination in different fractions of the sequence.

(A) Recombination rates of genomes in various datasets estimated using the Out-of-Africa model.

(B) Recombination rates of genomes in various datasets estimated using the constant size model.

(C) Proportion of recombination in different fractions of the sequence of 22 autosomes compared to that of the whole genome. The recombination rates of various chromosomes were estimated using the Out-of-Africa model.