Human Molecular Genetics, 2004, Vol. 13, No. 22 **2823–2828**
doi:10.1093/hmg/ddh310
Advance Access published on September 22, 2004

Evidence and characteristics of putative human α recombination hotspots

Jing Zhang[1,2], Fei Li[1], Jun Li[1], Michael Q. Zhang[1,3] and Xuegong Zhang[1,2,*]

[1]MOE Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing 100084, China, [2]Chinese National Human Genome Center, Beijing 100176, China and [3]Cold Spring Harbor Laboratory, New York, NY, USA

Received July 16, 2004; Revised September 6, 2004; Accepted September 17, 2004

Understanding recombination rate variation is very important for studying genome diversity and evolution, and for investigation of phenotypic association and genetic diseases. Recombination hotspots have been observed in many species and are well studied in yeast. Recent study demonstrated that recombination hotspots are also a ubiquitous feature of the human genome. But the nature of human hotspots remains largely unknown. We have developed and validated a novel computational method for testing the existence of hotspots as well as for localizing them with either unphased or phased genotyping data. To study the characteristics of hotspots within or close to genes, we scanned for unusually high levels of recombination using the European population samples in the SeattleSNPs database, and found evidence for the existence of human α hotspots similar to those of yeast. This type of hotspots, found at promoter regions, accounts for about half of the total detected and appears to depend on some specific transcription factor binding sites (such as CGCCCCCGC). These characteristics can explain the observed weak correlation between hotspots and GC-content, and their variation may contribute to the diversity of hotspot distribution among different individuals and species. These long-sought putative human α recombination hotspots should deserve further experimental investigations.

INTRODUCTION

Meiotic recombination has a profound influence on genome diversity and evolution, and understanding the non-random distribution of recombination events is important for explaining genome plasticity and for mapping disease-related loci. Recombination hotspots (defined as genomic regions with unusually high levels of recombination relative to the background) (1) in yeast have been investigated in detail (2), but the distribution and characteristics of hotspots for most species and human are largely unknown. In yeast, the recombination hotspots can be classified into three major categories: α hotspots, which require the binding of transcription factors and locate to transcriptional promoters, β hotspots, which correspond to regions of nuclease-sensitive chromatin and γ hotspots, which are associated with high GC-content (2). It has also been suggested that these three categories of hotspots might be mechanistically related (2). A recent study demonstrated that recombination hotspots are ubiquitous in the human genome, occurring on average every 200 kb or less, and preferentially outside genes (3). It is also recently reported that 47% of genes showed substantive evidence for a hotspot (4). But the nature and causes of their existence remain unclear, except a weak positive correlation between recombination rate and GC-content has been reported (5–8). In this paper, we focus on investigating statistical characteristics of putative human hotspots.

SNPs in the human genome are thought to be organized into blocks of high internal linkage disequilibrium (LD), separated by intermittent recombination hotspots. To investigate possible characteristics of human hotspots, we have developed and validated a novel computational approach (named BPBP) to identify recombination hotspots from patterns of genetic variation. In order to investigate the characteristics of hotspots within or close to genes, we applied our method to scan for putative hotspots in the gene loci corresponding to the European population samples in the SeattleSNPs database and studied their salient features.

*To whom correspondence should be addressed. Tel: +86 1062794919; Fax: +86 1062786911; Email: zhangxg@tsinghua.edu.cn

RESULTS

Validation of BPBP with experimental data

The method of BPBP is described in the Materials and Methods section. We first used experimentally known positive and negative data sets (see Materials and Methods) to validate the method.

Human leukocyte antigen hotspots data. The human leukocyte antigen (HLA) region is a well-studied region where recombination hotspots have been identified with sperm typing technology (9). Applied on the HLA SNP data set, our method detects the existence of hotspots in this region with *P*-value <0.001, and also clearly highlights the hotspot regions (Fig. 1). The peaks of the *b*-score curve (see Materials and Methods section) are fairly consistent with the regions containing hotspots as revealed by the sperm typing experiment (9).

TAP2 hotspot data. For the *TAP2* hotspot data (10), our approach also correctly predicts the existence of the hotspot ($P = 0.048$), and delineates the hotspot location (Fig. 2). According to Jeffreys *et al.* (9), there might be another hotspot in the nearby downstream region of *TAP2* (marked by '??' in Fig. 2), thus the *TAP2* hotspot might also be a member of a cluster just as those in *HLA-DOA* and *HLA-DMB* (9). The significant rising of *b*-score in this region also indicates that there could be another hotspot at the downstream region suggested by the previous experiment.

Human mitochondrial data. We then applied our approach to the human mitochondrial SNP data (11,12). There has been little evidence reported for widespread recombination in human mtDNA (13). Our approach does not find any significant hotspots in this negative data set (*P*-value >0.5), which is again consistent with the current knowledge.

Human hotspots prediction with BPBP

Seattle SNP data. The SeattleSNPs is a genotype data set of gene loci that are related to human diseases. We applied the BPBP method to the entire 127 gene loci (totally $\sim$3 Mb long) in this database to test whether these gene loci contain recombination hotspots. For the European American samples, hotspots were detected with significance (*P*-value <0.05) in 24 out of the 127 genes (Supplementary Material, Table 2). In these 24 genes, we identified 35 putative hotspot regions surrounded by 36 putative coldspot regions, which are segments ($\sim$2 kb or longer) covered by sets of consecutive SNPs whose *b*-scores are less than 0.2 of the maximum (see Supplementary Material). The average length of the detected putative hotspot regions is 4.5 kb. They may be regions containing one or more individual hotspots [average human hotspot width is $\sim$1–2 kb according to Jeffrey *et al.* (9)]. The total length of the putative coldspot regions is $\sim$212 kb, with average length of 5.9 kb each.

For the African American samples in the same database, only two of the 127 genes were detected to have putative hotspots. This may be caused by some noticeable differences in the LD block patterns between the two populations, noting that our approach is based on haplotype block partitioning (see Supplementary Material). In African American, LD blocks are much shorter, and detecting fine-scale recombination variations in the African American samples is more difficult than in the European American samples due to population bottlenecks that have probably occurred in the history of European populations but not in African populations (3,14). In European American, the background LD blocks are longer, which make the recombination hotspots more prominent and much easier to detect. Because of the limited number of significant loci and the possible complication in the African American samples, we only considered the European American samples for the following investigation and leave the African American data for future study.

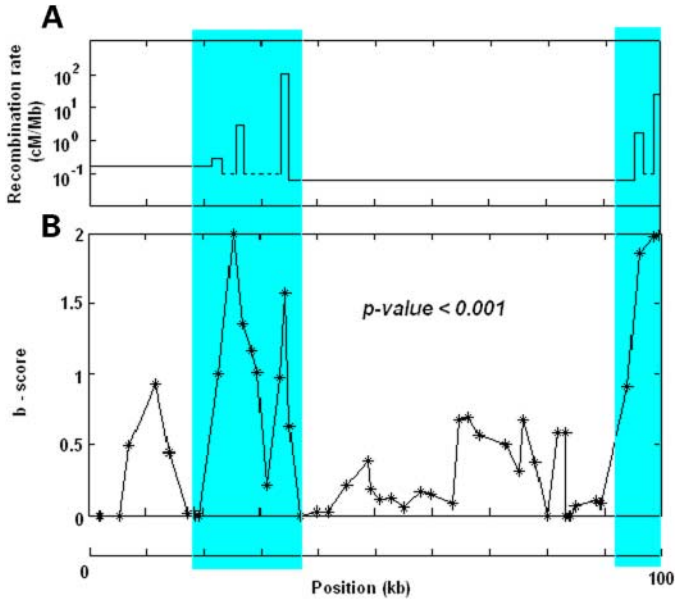

Figure 1. The BPBP result on HLA data. The statistical test detected hotspots in this data ($P < 0.001$), and the *b*-score highlights the regions containing hotspots. (**A**) The estimated intensities and positions of the hotspots, as well as the background recombination rate, according to the sperm typing result of Jeffreys *et al.* (9) are shown. (**B**) The *b*-score curve (each * represents the *b*-score at a SNP site) is shown. The BPBP result is consistent with the result of sperm typing experiments of Jeffreys *et al.* (9).

Relationship between human hotspots and transcription. In yeast, α hotspots requiring the binding of transcription factors are found at transcriptional promoters (2). In our study on human genes, among the 24 genes where hotspots were detected, the transcriptional promoter regions of 17 genes are included in the predicted putative hotspot regions and those of the rest seven are in putative coldspot regions. Nearly half of the putative hotspot regions cover a promoter region (17 out of 35). This indicates a significant association between putative hotspot regions and promoters ($P = 0.0055$, which is counted according to Poisson random process, i.e. the lengths of coldspot regions are drawn from an exponential distribution which approximately corresponds to the assumption that hotspots distribute randomly across the genome). More than expected promoter regions are in putative hotspot regions, indicating that there may exist

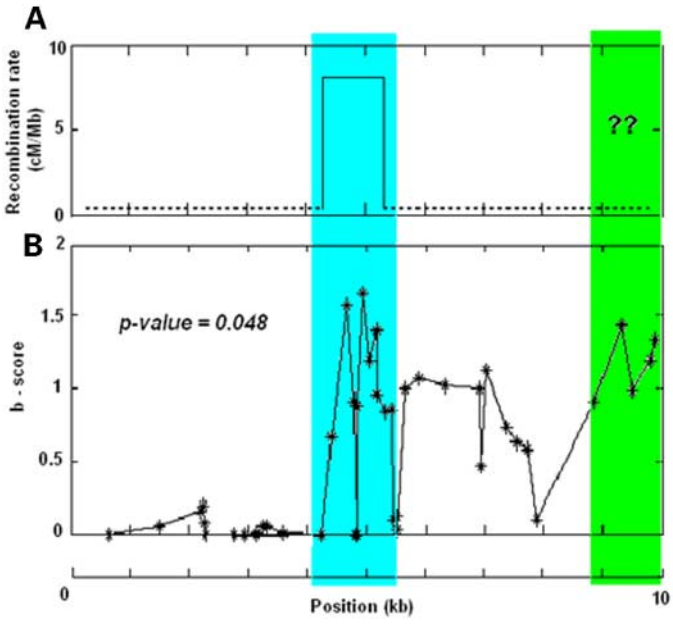

Figure 2. The BPBP result on *TAP2* data. The statistical test detected hotspots in this data ($P = 0.048$), and the *b*-score highlights the regions containing hotspots. (**A**) The approximate intensity, the position of the hotspot and the background recombination rate as estimated by Jeffreys *et al.* (10). It has been suggested that there might be another hotspot near the regions marked by '??' (9). The *b*-score in (**B**) highlights the known *TAP2* hotspot region correctly, and its rising in the 9–10 kb area also supports the suggestion that there is another hotspot in this region.

a similar category of human hotspots that require the binding of transcription factors just as the α hotspots do in yeast.

According to the studies of the yeast *HIS4* hotspot, the activity of this α hotspot requires the binding of transcriptional factors of *Bas1p*, *Bas2p* and *Rap1p*, but not the binding of *Gcn4p* (15). It is likely that some transcription factors are more efficient at stimulating recombination than others (16). If there are human α hotspots behaving in a similar way, one would expect to find conserved DNA sequences reflecting functional transcription factors binding sites within this subset of putative hotspots. We searched the 24 genes with TRANSFAC motifs (http://www.gene-regulation.com/pub/databases.html#transfac), and found four of them have experimentally verified transcription factor binding sites (totally nine distinct binding sequences). We searched for exact matches of these nine binding sequences in the 24 genes (35 putative hotspot regions and 36 putative coldspot regions). One of them, the palindromic sequence CGCCCCCGC (or the complementary sequence GCGGGGGCG), showed a significant enrichment in the putative hotspot regions than in the putative coldspot regions. Out of nine exact matches found, eight are located in the putative hotspot regions and one in the putative coldspot regions ($P = 0.0054$ by χ^2 analysis, which is still <0.05 after Bonferroni-correction, because we have tested nine motifs). This binding sequence (called early growth response, EGR, consensus sequence) is a recognition element common to the EGR family of Zn^{2+} finger transcriptional activators (17). It is also annotated as part of the binding sites for transcription factor *Sp1*, and Wilms tumor suppressor *WT1* (*WT1–KTS*, *WT1 I*, *WT1 I–KTS*, *WT1-del2*, *WT1 I-del2*) in both mouse and human. Moreover, seven of the nine exact matches are in promoter regions, and the other two (both in the putative hotspot regions) are in introns. Since the yeast α hotspot *HIS4* requires the binding of transcription factors but not a high level of transcription (18), hotspots with transcription binding sites but outside promoter regions might also be stimulated by the binding of transcription factors.

In human, there seems to be a similar underlying α hotspots mechanism of recombination as in yeast. Initiation of α hotspots might be associated with transcription factors and their binding sequences but not with transcriptional activity *per se*. The tendency of α hotspots to be at promoter regions might be due to the enrichment of binding sequences for those transcription factors required for recombination initiation (such as double strand break or DSBs). This enrichment is clear in our study: seven exact matches of EGR consensus sequences were found in promoters against two outside ($P = 1.68 \times 10^{-7}$ by χ^2 analysis).

It is likely that most human hotspots locate in intergenic regions. By confining on the hotspots within or close to genes on human autosomal chromosomes, we observe a significant (50%) tendency for them to locate at promoters. Noting that intergenic regions are not included in our study, these observations are in concord with the fact that out of the four (*TAP2* hotspot, *HLA-DOA1* hotspot, *HLA-DMB1* hotspot and β-globin hotspot) human hotspots localized by sperm typing that locate in gene regions, exactly half, *HLA-DOA1* and β-globin hotspots (19,20), do reside in the promoter regions. *HLA-DOA1* hotspot covers 1.9 kb region from ∼1 kb 5′ to intron 1 of *HLA-DOA* gene (9). For β-globin hotspot, crossovers in three families were experimentally mapped to a 1.5 kb region that stretches from ∼900 bp 5′ to intron 2 of β-globin gene (21).

Relationship between human hotspots and GC-content. In yeast, it has been observed that there is significant association between hotspots and regions of high-GC base composition (16). Studies also observed positive correlations between GC-content and the rate of crossover in other organisms. However, these correlations are generally weak, and for humans the correlation coefficient is in the range of 0.20–0.40 (8). We investigated the relationship between putative hotspot regions and regions of high GC-content. We labeled each of 71 putative regions (35 hotspot regions and 36 coldspot regions) by two characteristics: (1) hotspot or coldspot regions and (2) the GC-content of this putative region is higher or lower than the average of the corresponding gene locus. These two characteristics show a weak positive but significant correlation with the correlation coefficient $r = 0.2397$ ($P = 0.026$ by permutation), consistent with the previous reports.

Since most of human promoter regions are GC-rich, such a weak positive correlation can be explained by the existence of promoter enriched α hotspots. Also, some binding sequences for the required transcription factors can be GC-rich (as in the case of CGCCCCCGC and its reverse complement GCGGGGGCG). In order to examine the effect of promoter regions on the observed correlation, we studied all the 18 putative hotspot regions and 27 putative coldspot regions

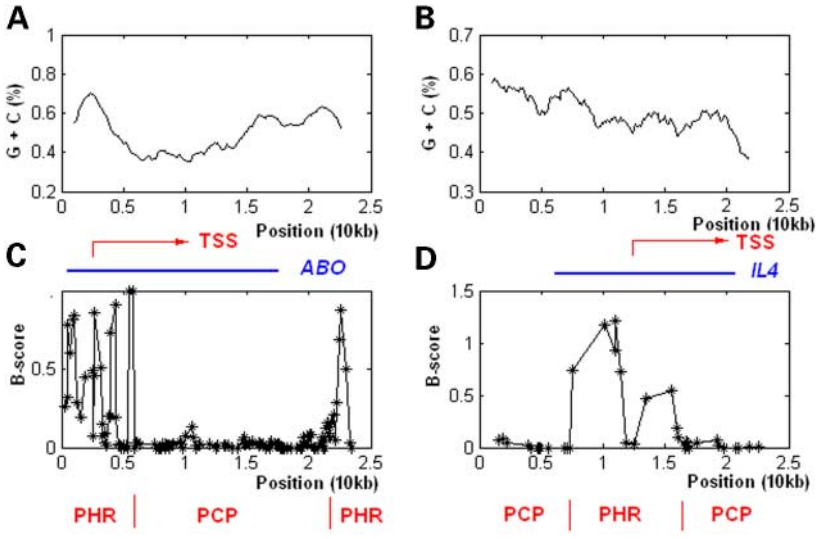

Figure 3. GC-content and BPBP result for *ABO* and *IL4* gene loci. (**A** and **C**): *ABO* and (**B** and **D**): *IL4*. (A and B) the GC-contents measured in a 2 kb sliding window shifted at 100 bp intervals. The vertical bars and the arrows indicate transcription start sites (TSS) and transcript direction. (C and D) The *b*-score curves. Detected putative hotspot regions are denoted as PHR, and putative coldspot region as PCR. The two PHRs in *ABO* locus correspond to GC-rich regions, whereas the one in *IL4* locus does not. Both the TSS are included in PHR.

that are not in promoter regions. For these data, the putative hotspot regions and high-GC regions showed no significant correlation: $r = 0.1516$ ($P = 0.167$ by permutation), which supports the explanation that existence of α hotspots may be the major cause of the weak positive correlation between hotspots and high GC-contents. As examples, Figure 3 shows the GC-content and BPBP result of *ABO* and *IL4* gene loci in SeattleSNPs.

A number of hypotheses have been proposed to explain the positive correlation between base composition and recombination rate, such as 'recombinogenic' GC-rich regions (2,16) and Biased Gene Conversion (8). Our study focused on the fine-scale relationship between high GC-content regions and recombination hotspots within or close to genes. The weak correlation we observed could be better explained by the existence of α hotspots in human. The fact that no such correlation exists outside promoter regions suggests that non-α hotspots might not be related with base composition while α hotspots are, and α hotspots and non-α hotspots could result from different recombination mechanisms.

Recent studies suggested that fine-scale recombination rate variation may change rapidly over evolutionary time (3,19,22). Supporting observations for this suggestion include that recombination rates vary among individuals (23), that nucleotide polymorphisms influence local recombination rate (22), and that there are differences in recombination hotspots between human and other primates (19). The rapid change and poor conservation across species cannot be fully explained with variations due to mutation. The existence of α hotspots requiring certain transcription factors gives a new explanation. Study of the yeast *HIS4* hotspot showed that maximal stimulation of meiotic recombination by transcription factor requires both the transcription activation domain and a DNA binding domain (24). Variations in these transcription factors as well as nucleotide polymorphisms in their binding sites may influence substantially the recombination efficiency in α hotspots, causing more chances for diversity in recombination rate patterns across different individuals or species.

DISCUSSION

Historically, a direct measurement of recombination rate at high resolution is experimentally very challenging. The resolution of traditional pedigree-based methods is limited by the marker density used and the number of informative meioses contained in pedigree data. The new remarkable sperm typing technology provides a promising method but is still tedious and costly, and says nothing about female recombination rate variation. A number of indirect statistical/computational methods to infer local recombination rates from population genetic data can offer valuable alternatives as a guide to the location of putative hotspots (25). Recent studies (3,26) suggest that the population genetic statistical analyses can reach fairly accurate estimates of local recombination rates within a factor of 2. The method of Li and Stephens (26) can serve as a statistical test for the existence of hotspot; nevertheless, it requires the unrealistic assumption of only a single hotspot with known putative position. Both the methods of Li and Stephens (26) and McVean *et al.* (3) first estimate recombination rate along the sequence, then test the existence of fine-scale recombination rate variation (hotspot). Both of them are based on certain statistical models. Our method is based on bootstrap samplings of haplotype block partition and does not assume specific statistical models. It directly tests the existence of hotspots without estimating recombination rate. Further improvement of our method may be achieved by using a proper statistical model just as Li and Stephens (26) and McVean *et al.* (3).

According to our extensive simulation results in Supplementary Material, on the level of P-value <0.05, our method

is rather conservative. It may miss a substantial proportion of real hotspots with median hotspot intensity. In addition, limited sample size may also influence the ability of our method to detect hotspots. However, the false-positive rate is not affected (P-value <0.05) in our study. Since no information about promoter location or base composition is used during our hotspot testing, the locations of those possibly missing hotspots are therefore erratic and random, and do not possibly bias our significant observations. Since human hotspots seem to cluster into groups (like those in the MHC region), given the limited data, our method may not reach the resolution to locate each individual hotspot within a cluster. Thus, putative hotspot regions we found in SeattleSNPs data are likely hotspot clusters rather than individual hotspots (like the case for HLA shown in Fig. 1), namely they may contain smaller cold patches. It is therefore possible that functional promoter elements may actually localize to these patches.

We noticed that recently Crawford *et al.* (4) also used SeattleSNPs data but a different algorithm to identify recombination hotspots, and found distinct results. There are three main reasons causing the different results. First, the method used by Crawford *et al.* (4) assumed one single hotspot in each gene, which probably made them fail to detect those genes that contain two (or more) hotspots, especially when the hotspots are far away from each other just as the case of *ABO* gene in Figure 3. Second, as shown in the simulation results, our method may miss some hotspots with median hotspot intensity. Third, although we both used Seattle SNP data, the genes studied are not exactly the same. Among our total 127 genes, only 65 genes are also included in Crawford *et al.* (4).

With the proposed computational method for detecting putative hotspots from genotype/haplotype data, we were able, for the first time, to observe a strong evidence for the existence of human α hotspots. This category appears to account for about half of the putative hotspots around genes ($\sim \pm 3$ kb flanking regions in average) in the human genome, and plays a major role in the observed weak correlation between hotspots and high GC-content. The differential regulatory mechanism of α hotspots can also be one reason that causes the observed diversity and variation of hotspots among individuals and different species. These findings are crucial for understanding the biological process of recombination and the mechanism of recombination hotspots in human and other related species.

MATERIALS AND METHODS

Data sets

HLA hotspots data. We use this data set (9) to validate our method. This data set contains 50 unrelated UK male individuals with 296 biallelic markers including 274 SNPs and 22 1–11 bp indels. Because the markers locating between *HLA-DMB* region and *TAP2* region were spacing too sparsely (approximately average marker spacing of 7 kb) and in *TAP2* region there is a large missing data bulk (11 markers for 20 individuals), only markers in the first 100 kb were considered. Only 177 biallelic markers passed our selection criterion (see later). Five known hotspots cluster in two groups in this region. Because in this data set the markers spread very unevenly (much more markers within hotspots than outside), we randomly sampled 52 markers from 177 markers so that this 100 kb region was approximately evenly covered at an average marker spacing of 2 kb.

TAP2 hotspot data. This data set (10) is also used to validate our method. The unphased genotype data of *TAP2* recombination hotspot within the class II region of the MHC were obtained from Alec J. Jeffreys' research group webpage (10), which contain 30 UK Caucasians genotypes with 48 biallelic markers in 9.7 kb *TAP2* region. Forty-six markers were selected.

Human mitochondrial data. The 86 complete human mtDNA sequences (11,12) were downloaded from Human Mitochondrial Genome Database (http://www.genpat.uu.se/mtDB/), finding 822 biallelic SNPs. Only 134 SNPs were selected. Since there is little evidence for widespread recombination in human mtDNA (13), it would be surprising to find recombination hotspots in mtDNA data. Therefore, mtDNA data served as a negative control for our method.

Seattle SNP data. The SeattleSNPs data were downloaded from the University of Washington Fred Hutchinson Cancer Research Center (UW-FHCRC) Variation Discovery Resource Web site (http://pga.gs.washington.edu/) in December, 2003. This data set contains SNPs of 127 genes located at human autosomal chromosomes (and four at human X chromosome which are not considered in our study) from a sample of 24 unrelated African Americans (12 females and 12 males) and 23 unrelated European Americans (11 females and 12 males) from the Coriell Cell Repository. The loci were completely resequenced except a few small gaps. No gender information is available for each individual. In the European American (African American) samples, the total of 4818 (6734) selected SNPs have an average marker spacing of 628 bp (450 bp) spread over 3 Mb (3 Mb) of sequence. There is no recombination hotspot reported in these gene loci. In this work, we used our method to test the underlying recombination distribution for each locus and predicted putative hotspot regions, and studied the characteristics of the predicted regions.

All the selected SNPs for this work have a minor-allele frequency not less than 0.05 and with no significant departure from Hardy–Weinberg equilibrium (P-value <0.001).

Methods

The method we propose is based on a bootstrap strategy of haplotype block partitioning, which gives a b-score as an indicator of the strength for a marker to be block boundary. With this score, we devise a statistic to test the null hypothesis of uniform recombination rate model against the alternative model of containing recombination hotspots. If the hotspot model is accepted, i.e. the existence of hotspots is detected, the cluster regions of high b-scores are regarded as putative hotspot regions. This method (named BPBP) can be applied to both phased and unphased genotype data. We used a series of coalescent-based simulation to validate the method

and its power (see Supplementary Material for details). According to the simulation result, the power of our method is affected mainly by the intensity of hotspots for any moderate sample sizes. Those hotspots with high recombination intensity are fairly easy to detect.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

Thanks to Dr Yan Shen and Chaolin Zhang for their discussions on this work. The work is partially supported by NSFC projects no. 60275007 and 60234020 of China, and the National Basic Research Program of China.

REFERENCES

1. Lichten, M. and Goldman, A.S.H. (1995) Meiotic recombination hotspots. *Annu. Rev. Genet.*, **29**, 423–444.
2. Petes, T.D. (2001) Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.*, **2**, 360–369.
3. McVean, G.A.T., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R. and Donnelly, P. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.
4. Crawford, D.C., Bhangale T., Li, N., Hellenthal, G., Rieder, M.J., Nickerson, D.A. and Stephens, M. (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.*, **36**, 700–706.
5. Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W. *et al.* (2001) Comparison of human genetic and sequence-based physical maps. *Nature*, **409**, 951–953.
6. Fullerton, S.M., Carvalho, A.B. and Clark, A.G. (2001) Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.*, **18**, 1139–1142.
7. Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G. *et al.* (2002) A high-resolution recombination map of the human genome. *Nat. Genet.*, **31**, 241–247.
8. Marais, G. (2003) Biased gene conversion: implications for genome and sex evolution. *Trends Genet.*, **19**, 330–338.
9. Jeffreys, A.J., Kauppl, L. and Neumann, R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.*, **29**, 217–222.
10. Jeffreys, A.J., Ritchie, A. and Neumann, R. (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum. Mol. Genet.*, **9**, 725–733.
11. Ingman, M., Kaessmann, H., Paabo, S. and Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**, 708–713.
12. Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C. and Cabrera, V.M. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.*, **2**, 13
13. Posada, D., Crandall, K.A. and Holmes, E.C. (2002) Recombination in evolutionary genetics. *Ann. Rev. Genet.*, **36**, 75–97.
14. Kauppi, L., Sajantila, A. and Jeffreys, A.J. (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum. Mol. Genet.*, **12**, 33–40.
15. White, M.A., Dominska, M. and Petes, T.D. (1993) Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **90**, 6621–6625
16. Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O. and Petes, T.D. (2000) Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **97**, 11383–11390.
17. Rauscher, F.J., Morris, J.F., Tournay, O.E., Cook, D.M. and Curran, T. (1990) Binding of the Wilms' tumor locus zinc finger protein to the EGR-1 consensus sequence. *Science*, **250**, 1259–1262.
18. White, M.A., Detloff, P., Strand, M. and Petes, T.D. (1992) A promoter deletion reduces the rate of mitotic, but not meiotic, recombination at the HIS4 locus in yeast. *Curr. Genet.*, **21**, 109–116.
19. Wall, J.D., Frisse, L.A., Hudson, R.R. and Rienzo, A.D. (2003) Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. *Am. J. Hum. Genet.*, **73**, 1330–1340.
20. Schneider, J.A., Peto, T.E., Boone, R.A., Boyce, A.J. and Clegg, J.B. (2002) Direct measurement of the male recombination fraction in the human ß-globin hot spot. *Hum. Mol. Genet.*, **11**, 207–215.
21. Smith, R.A., Ho, P.J., Clegg, J.B., Kidd, J.R. and Thein, S.L. (1998) Recombination breakpoints in the human beta-globin gene cluster. *Blood*, **92**, 4415–4421.
22. Jeffreys, A.J. and Neumann, R. (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.*, **31**, 267–271.
23. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. and Weber, J.L. (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.*, **63**, 861–869
24. Kirkpatrick, D.T., Fan, Q.Q. and Petes, T.D. (1999) Maximal stimulation of meiotic recombination by a yeast transcription factor requires the transcription activation domain and a DNA binding domain. *Genetics*, **152**, 101–115.
25. Kauppi, L., Jeffreys, A.J. and Keeney, S. (2004) Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.*, **5**, 413–424
26. Li, N. and Stephens, M. (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, **165**, 2213–2233.