Detecting Haplotype Effects in Genomewide Association Studies

B.E. Huang,[1] C.I. Amos,[2] and D.Y. Lin[1]*

[1]*Department of Biostatistics, University of North Carolina, North Carolina, USA*
[2]*Department of Epidemiology, MD Anderson Cancer Center, Texas, USA*

The analysis of genomewide association studies requires methods that are both computationally feasible and statistically powerful. Given the large-scale collection of single nucleotide polymorphisms (SNPs), it is desirable to explore the information contained in their interrelationships. In particular, utilizing haplotypes rather than individual SNPs and accounting for correlations of polymorphisms in adjustment for multiple testing can lead to increased power. We present a statistically powerful and numerically efficient method based on sliding windows of adjacent SNPs to detect haplotype-disease association in genomewide studies. This method consists of an efficient algorithm to calculate a proper likelihood-ratio statistic for any given window of SNPs, along with an accurate and efficient Monte Carlo procedure to adjust for multiple testing. Simulation studies using the HapMap data showed that the proposed method performs well in realistic situations. We applied the new method to a case-control study on rheumatoid arthritis and identified several loci worthy of further investigations. *Genet. Epidemiol.* 2007.

Key words: case-control studies; complex diseases; HapMap; linkage disequilibrium; rheumatoid arthritis; SNPs; statistical significance

Contract grant sponsors: National Cancer Institute; National Institute of Environmental Sciences.
*Correspondence to: Danyu Lin, Ph.D., Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, CB #7420, Chapel Hill, North Carolina 27599-7420. E-mail: lin@bios.unc.edu
Received 7 February 2007; Accepted 16 April 2007
Published online in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20242

INTRODUCTION

Genomewide association studies were proposed a decade ago as a potentially powerful tool to unravel the genetic basis of complex diseases [Risch and Merikangas, 1996]. However, it is only now that they are becoming practical realities. Genotyping costs have decreased greatly in recent years, to the point where chips containing 100 K single nucleotide polymorphisms (SNPs), or even 250 K have already been used in various studies [Ozaki et al., 2002; Klein et al., 2005; Maragenore et al., 2005; Thomas et al., 2005], and investigations of larger numbers of SNPs are already underway. While genomewide association studies are currently in wide proliferation, the methodology to perform the analysis has not kept pace with the collection of data [Thomas et al., 2005].

To assess the association between SNPs and disease, one can either examine individual SNPs or consider the haplotypes of multiple SNPs. While the latter is potentially more powerful than the former [Collins et al., 1997; Akey et al., 2001; Morris and Kaplan, 2002; Schaid et al., 2002; Zaykin et al., 2002], it is computationally more intensive. The creation of HapMap [Gibbs et al., 2003; International HapMap Consortium 2005], however, has facilitated the use of haplotypes. As the HapMap project and current SNP platforms focus on cataloging common SNPs, the single-SNP analysis is not capable of detecting rare causative SNPs. In contrast, haplotype-based analysis may be able to do so if the rare SNP is captured by a haplotype [de Bakker et al., 2005].

A major analytical challenge is that haplotypes are not directly measured. Several methods are available to infer individual haplotypes from unphased genotype data (e.g., [Excoffier and Slatkin, 1995; Stephens et al., 2001; Niu et al., 2002]). Using the inferred individual haplotypes in the ensuing association analysis can result in biased estimates and reduced power [Lin and Huang, 2007]. A few methods have been proposed to properly account for phase uncertainty in the association analysis [Zhao et al., 2003; Stram et al., 2003; Epstein and Satten, 2003; Lin et al., 2005], all of which are focused on the analysis of a single candidate gene.

In this article, we provide a computationally efficient and statistically powerful method for detecting haplotype-disease association in genomewide studies. We consider sliding windows of adjacent SNPs (see Mathias et al. [2006] and the references therein). Within each window, we use an efficient and stable algorithm to calculate a likelihood-ratio test statistic that properly accounts for phase uncertainty and case-control sampling. The

windows may be overlapping or non-overlapping, and the window sizes may be fixed or variable. We allow exhaustive testing, which considers all possible windows up to a certain size and thus encompasses the single-SNP analysis.

The number of tests can be very large, particularly in the case of exhaustive testing. It is common to use the Bonferroni correction or permutation to adjust for multiple testing. The former is overly conservative, especially for overlapping windows and exhaustive testing, while the latter is computationally impractical for haplotype-based genomewide testing. Another limitation of these approaches is that they aim to control the family-wise error rate (FWER), i.e., the probability of even one false positive result, which is a very stringent criterion for massive-scale hypothesis testing.

In this article, we propose a computationally efficient method to properly adjust for multiple testing in large-scale association studies. This method can be used to control the probability of k (≥ 1) or more false positives, denoted by k-FWER [Lehmann and Romano, 2005]. The basic strategy is to ascertain the joint distribution of the test statistics among windows and to evaluate this joint distribution by an efficient Monte Carlo (MC) procedure. By properly accounting for the correlations of the test statistics, the proposed method avoids the conservativeness of the Bonferroni approach. Our MC procedure reduces the computational burden by orders of magnitude in comparison to permutation. Simulation studies with the phased haplotypes of the HapMap CEU population showed that the proposed method with various window choices provides accurate control of the traditional FWER as well as the more general k-FWER and is substantially more powerful than the Bonferroni correction and its k-FWER version of Lehmann and Romano [2005].

We applied the new method to a case-control study of association between rheumatoid arthritis (RA) and 2,300 SNPs in a region of interest on chromosome 18. Previous studies had shown mild evidence for linkage in this region [Merriman et al., 2001] as well as possible links of this region to a variety of other auto-immune diseases such as type I diabetes and multiple sclerosis. The single-SNP analysis did not show any significant results (after adjusting for multiple comparisons), and neither did the haplotype-based analysis with the Bonferroni correction. The use of the proposed method revealed several areas that merit further investigations.

METHODS

Our method for assessing haplotype-disease association within a given window is as follows. We first estimate the frequencies of all possible haplotypes for cases and controls separately by using the expectation-maximization (EM) algorithm (Excoffier and Slatkin, 1995). To improve stability and speed up computation, we remove the haplotypes with estimated frequencies $< c_f$ in the control group, where c_f is a very small number, say $1/n$ or $2/n$, and n is the total number of subjects in the study. The remaining number of haplotypes is denoted by K. The subjects whose genotypes are not compatible with the retained haplotypes are dropped from the data. For $k = 1, \ldots, K$, let h_k denote the kth haplotype and let π_k denote the frequency of h_k in the whole population. We fit a logistic regression model with additive haplotype effects that compares all haplotypes with estimated frequencies $> c_e$ in both cases and controls, where c_e is a small number, say $5/n$ or $10/n$. We use the most frequent haplotype as the reference group in the model. The haplotypes with estimated frequencies less than the threshold c_e are also included in the reference group. Estimating separate effects of such rare haplotypes would be numerically unstable. The number of haplotype effects in the model is denoted by r. We set c_e higher than c_f because it is more difficult to estimate the effect of a rare haplotype than to estimate its frequency.

The observed data consist of (Y_i, G_i), $i = 1, \ldots, n$, where Y_i and G_i denote the disease status and genotype for the ith subject. With H representing the pair of haplotypes for a subject, the logistic regression model takes the form:

$$\Pr(Y = 1 | H = (h_k, h_l)) = \frac{e^{\alpha + \beta^T Z(h_k, h_l)}}{1 + e^{\alpha + \beta^T Z(h_k, h_l)}},$$

where α pertains to the intercept, β represents log-odd ratios,

$$Z(h_k, h_l) = \begin{bmatrix} I(h_k = h_1) + I(h_l = h_1) \\ \vdots \\ I(h_k = h_r) + I(h_l = h_r) \end{bmatrix},$$

and $I(\cdot)$ is the indicator function. This formulation differs from that of previous work [Lin et al., 2005] in that we are comparing all the r haplotypes simultaneously rather than comparing a single halotype to all others. The likelihood should take into account the phase uncertainty in the genotype data as well as the biased sampling of the case-control design. Under the assumption of rare disease and Hardy-Weinberg equilibrium, the likelihood $\prod_{i=1}^{n} \Pr(G_i | Y_i)$ can be shown to be

$$\prod_{i=1}^{n} \frac{\Sigma_{k,l} I[(h_k, h_l) \in S(G_i)] e^{Y_i \beta^T Z(h_k, h_l)} \pi_k \pi_l}{\Sigma_{k,l} e^{Y_i \beta^T Z(h_k, h_l)} \pi_k \pi_l},$$

where $S(G)$ denotes the set of haplotypes compatible with genotype G, and the summation of (k, l) is taken from 1 to K.

To incorporate the constraints that $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k \geq 0$, $k=1,\ldots, K$, into the calculations, we reparametrize the model by defining $\pi_k^* = \pi_k/\pi_K$ and $\nu_k = \log \pi_k^*$, $k=1,\ldots, K$. Write $\nu = (\nu_1, \ldots, \nu_{K-1})$ and $\theta = (\beta, \nu)$. Then the log-likelihood can be written as

$$l(\theta) = \sum_{i=1}^{n} \log \left[\frac{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} e^{\theta^T W(Y_i, h_k, h_l)}} \right],$$

where

$$W(Y_i, h_k, h_l) = \begin{bmatrix} Y_i Z(h_k, h_l) \\ I(h_k = h_1) + I(h_l = h_1) \\ \vdots \\ I(h_k = h_{K-1}) + I(h_l = h_{K-1}) \end{bmatrix}.$$

The corresponding score function and information matrix are

$$\begin{aligned} &U(\theta) \\ &= \sum_{i=1}^{n} \left[\frac{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] W(Y_i, h_k, h_l) e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] e^{\theta^T W(Y_i, h_k, h_l)}} \right. \\ &\quad \left. - \frac{\sum_{k,l} W(Y_i, h_k, h_l) e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} e^{\theta^T W(Y_i, h_k, h_l)}} \right], \end{aligned}$$

and

$$\begin{aligned} \Sigma(\theta) = &\sum_{i=1}^{n} \left[\frac{\sum_{k,l} W(Y_i, h_k, h_l)^{\otimes 2} e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} e^{\theta^T W(Y_i, h_k, h_l)}} \right. \\ &\left. - \left\{ \frac{\sum_{k,l} W(Y_i, h_k, h_l) e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} e^{\theta^T W(Y_i, h_k, h_l)}} \right\}^{\otimes 2} \right] \\ &- \sum_{i=1}^{n} \left[\frac{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] W(Y_i, h_k, h_l)^{\otimes 2} e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] e^{\theta^T W(Y_i, h_k, h_l)}} \right. \\ &\left. - \left\{ \frac{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] W(Y_i, h_k, h_l) e^{\theta^T W(Y_i, h_k, h_l)}}{\sum_{k,l} I[(h_k, h_l) \in S(G_i)] e^{\theta^T W(Y_i, h_k, h_l)}} \right\}^{\otimes 2} \right], \end{aligned}$$

where $a^{\otimes 2} = aa^T$. To obtain the maximum likelihood estimate $\widehat{\theta}$, we solve the score equation $U(\theta) = 0$ by using the Newton-Raphson method. We set the initial value of θ to $\widetilde{\theta} = (0, \widetilde{\nu})$, where $\widetilde{\nu}$ is the maximum likelihood estimate of ν in the pooled sample obtained by the EM algorithm.

We can test the haplotype-disease association by using the likelihood ratio statistic $2[l(\widehat{\theta}) - l(\widetilde{\theta})]$, the score statistic, or the Wald statistic. All the three test statistics have approximately the χ^2 distribution with r degrees of freedom. In deriving the joint distribution of the test statistics over different windows, it is convenient to work with the score statistic. We partition the score function and information matrix to conform with the partition of β and ν in θ, i.e.,

$$U(\theta) = \begin{bmatrix} U_\beta(\theta) \\ U_\nu(\theta) \end{bmatrix},$$

and

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{\beta\beta}(\theta) & \Sigma_{\beta\nu}(\theta) \\ \Sigma_{\nu\beta}(\theta) & \Sigma_{\nu\nu}(\theta) \end{bmatrix}.$$

Also, let $U_{\beta,i}(\theta)$ and $U_{\nu,i}(\theta)$ denote the contributions from the ith subject to $U_\beta(\theta)$ and $U_\nu(\theta)$. The score statistic can then be written as

$$T = U_\beta(\widetilde{\theta})^T V^{-1} U_\beta(\widetilde{\theta}),$$

where $V = \sum_{i=1}^{n} U_i U_i^T$ and $U_i = U_{\beta,i}(\widetilde{\theta}) - \Sigma_{\beta\nu}(\widetilde{\theta}) \Sigma_{\nu\nu}^{-1}(\widetilde{\theta}) U_{\nu,i}(\widetilde{\theta})$.

We approximate the joint distribution of the test statistics over windows through a MC simulation procedure. Specifically, we construct $\widetilde{T} = \widetilde{U}^T V^{-1} \widetilde{U}$, where $\widetilde{U} = \sum_{i=1}^{n} U_i X_i$, and X_i, $i = 1, \ldots, n$, are independent standard normal random variables. Suppose that we have a total of m windows, which may or may not be overlapping and which cover the whole region one is scanning. Let T_j and $\widetilde{T}_j$ denote the values of T and $\widetilde{T}$ in the jth window. The same set of X_i, $i = 1, \ldots, n$, is used for all the m simulated statistics $\widetilde{T}_1, \ldots, \widetilde{T}_m$. By the arguments of Lin [2005], the joint distribution of $(T_1, \ldots, T_m)$ can be approximated by the joint distribution of $(\widetilde{T}_1, \ldots, \widetilde{T}_m)$. We obtain realizations from the latter distribution by generating the normal samples $(X_1, \ldots, X_n)$ while fixing the genotype and phenotype data at their observed values.

The above MC approximation is valid whether $T_1, \ldots, T_m$ are the likelihood ratio, score or Wald statistics. Our simulation studies revealed that the approximation tends to be more accurate for the likelihood-ratio statistics than the score and Wald statistics although the differences are generally very small. The numerical results reported in this article pertain to the likelihood ratio.

In the standard multiple-testing framework [Westfall and Young, 1993; Lin, 2005], the m test statistics have the same degrees of freedom. In our setting, the test statistics have different degrees of freedom because the numbers of haplotype effects tested vary among windows. Thus, we propose a step-down multiple-testing procedure which orders the P-values of the test statistics rather than the actual values of the test statistics. This is similar to Algorithm 2.8 in Westfall and Young [1993], which is based on resampling from the original data.

For $j = 1, \ldots, m$, let p_j be the (observed) P-value associated with the test statistic T_j, which is obtained from the χ^2 distribution with r_j degrees of freedom, where r_j is the number of haplotype effects tested in the jth window. Let $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$ be the ordered P-values, and let $H_{(1)}, \ldots, H_{(m)}$ be the

corresponding null hypotheses. In addition, let $\widetilde{T}_{(1)}, \ldots, \widetilde{T}_{(m)}$ be the simulated test statistics associated with $H_{(1)}, \ldots, H_{(m)}$, and let $\widetilde{p}_{(1)}, \ldots, \widetilde{p}_{(m)}$ be the corresponding simulated P-values. The adjusted P-value for testing $H_{(j)}$ (i.e., the smallest significance level at which $H_{(j)}$ would be rejected by the multiple testing procedure) is determined by

$$\Pr\left(\min_{j \le l \le m} \widetilde{p}_{(l)} \le p_{(j)}\right).$$

We estimate this probability with 5,000 realizations of $\widetilde{p}_{(1)}, \ldots, \widetilde{p}_{(m)}$, which are obtained by repeatedly generating the normal samples $(X_1, \ldots, X_n)$ while holding the observed data fixed. To control the traditional FWER at α, one would reject only those hypotheses whose adjusted P-values are less than α.

The traditional FWER may be too stringent in massive-scale hypothesis testing. Thus, we extend Lehmann and Romano's [2005] idea of controlling k-FWER, which is the probability of rejecting greater than or equal to k true hypotheses. To obtain the MC adjusted P-values on the basis of k-FWER, we simply replace the minimum P-value in the above formula by the kth smallest P-value. The adjusted P-values based on the Bonferroni correction, as suggested by Lehmann and Romano, are $p_{(j)}m/k$, $j = 1, \ldots, m$. Only $k = 1$ and 2 were used in our calculations in this article, although a larger value of k may be desirable for increasing quantities of markers.

RESULTS

SIMULATION STUDIES

We simulated data from the 120 phased haplotypes of Caucasians in the Phase I HapMap data. We considered two regions on chromosome 18: the ENCODE region, which consists of 796 SNPs, and the full set of 32,177 SNPs for the chromosome. We selected a pair of haplotypes randomly from the HapMap data for each subject and then added the two haplotypes to give the subject's genotype. We generated disease according to an additive-effect logistic model with an overall disease rate of 5%.

Figures 1 and 2 display the locations of the two HapMap regions and the linkage disequilibrium (LD) among the SNPs. The SNPs in the ENCODE region show much higher levels of LD than the full set of SNPs. This reflects the fact that the density of SNPs in the ENCODE region is higher than elsewhere.

We used the ENCODE data to assess the performance of the proposed MC method and Bonferroni correction for different window sizes, and overlapping versus non-overlapping windows. We set $c_f = 2/n$ and $c_e = 10/n$. We considered both the FWER and 2-FWER, denoted by Bon and Bon-2 for the Bonferroni correction and by MC and MC-2 for the proposed MC method. The results of these studies for windows of three and four SNPs are presented in Table I. For both window sizes, the causative haplotype began at SNP 601 and had frequency of .14. The type I error pertains to the probability of declaring any disease-causing SNPs when no effect exists, while the power is the same quantity when one haplotype is in fact causative. Both MC and MC-2 provide accurate control of the type I error in all cases, whereas both Bon and Bon-2 are severely conservative and thus much less powerful than MC and MC-2. As expected, MC-2 is considerably more powerful than MC. The power of

TABLE I. Type I error/power of haplotype tests at the .05 nominal significance level based on the ENCODE data

		Overlapping windows				Non-overlapping windows			
Odds ratio	Sample size	Bon	Bon-2	MC	MC-2	Bon	Bon-2	MC	MC-2
				Windows of three SNPS					
1.0	1,000	.014	.016	.041	.041	.015	.017	.054	.055
	2,000	.009	.015	.041	.049	.016	.010	.047	.039
1.5	1,000	.369	.455	.488	.557	.371	.447	.509	.611
	2,000	.754	.798	.880	.914	.798	.853	.876	.931
1.7	1,000	.693	.741	.843	.899	.725	.786	.829	.900
				Windows of four SNPS					
1.0	1,000	.007	.009	.038	.043	.023	.018	.046	.040
	2,000	.013	.013	.049	.053	.017	.009	.040	.036
1.5	1,000	.286	.356	.481	.572	.397	.480	.512	.631
	2,000	.735	.791	.876	.907	.813	.863	.879	.935
1.7	1,000	.676	.736	.821	.867	.743	.808	.845	.907

Note: Sample size pertains to the total number of study subjects, half of which are cases and half are controls. Bon and Bon-2 pertain to the FWER and 2-FWER based on the Bonferroni correction, and MC and MC-2 to the FWER and 2-FWER based on the Monte Carlo procedure. Each entry is based on 1,000 simulated datasets.
FWER, family-wise error rate.

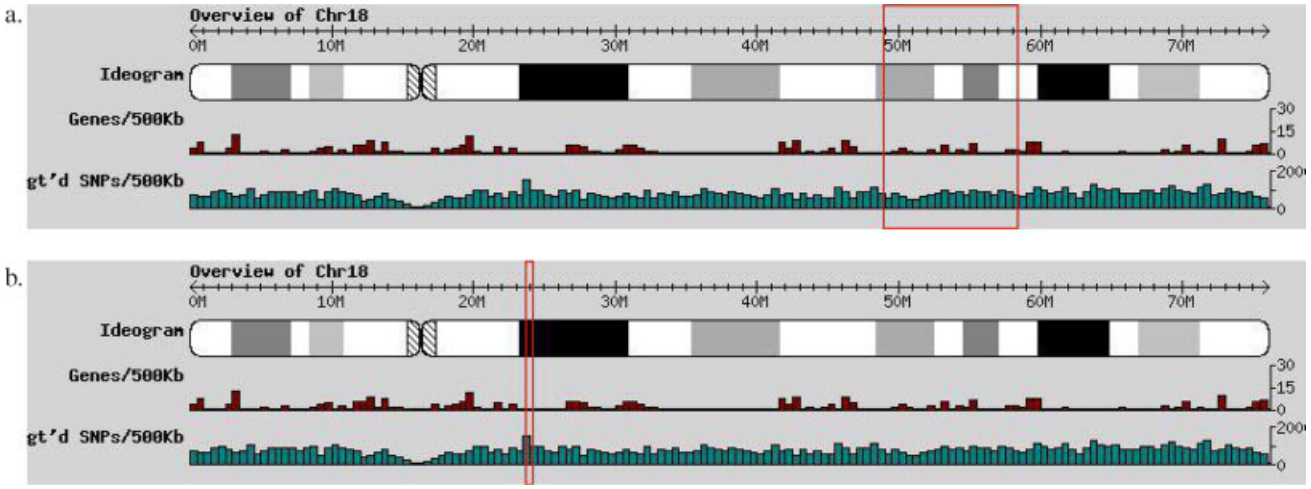

Fig. 1. Locations of single nucleotide polymorphisms (SNPs) in two regions of interest on chromosome 18. (a) 2,300 SNPs from the rheumatoid arthritis case-control study; (b) 796 SNPs from the HapMap ENCODE region.

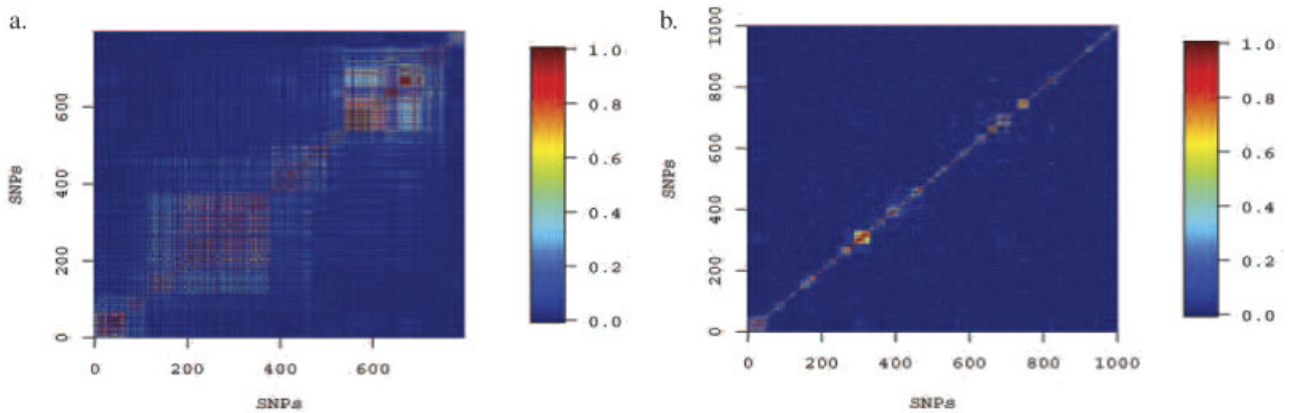

Fig. 2. Patterns of linkage disequilbrium (LD), as measured by the squared correlation coefficient r^2 between pairs of markers, in two HapMap regions on chromosome 18. (a) 796 single nucleotide polymorphisms (SNPs) in the ENCODE region; (b) first 1,000 SNPs in the full set of SNPs.

MC is similar to and often higher than that of Bon-2. Using the proposed method, a sample size of 2,000 subjects is sufficient to detect an odds ratio of 1.5 with high power, and even a sample size of 1,000 provides power $>.8$ for an odds ratio of 1.7. Non-overlapping windows appear to have higher power than overlapping windows.

In these studies, the single-SNP analysis reduced the power by about 3% for both window sizes of 3 and 4. The fairly moderate reduction in power is attributed to the simulation setup, which involves a common causal haplotype. We performed additional studies with a rare causative SNP that is captured by a rare haplotype but not actually measured in the data and observed power gain of up to 25% for the haplotype analysis over the single-SNP analysis (results not shown).

For the full set of 32,177 SNPs, we used non-overlapping windows of size 3, and the results are presented in Table II. The causative haplotype began at SNP 637, and had frequency of .18. As expected, the larger magnitude of data in this setting leads to lower all around power as compared with the ENCODE data. The decline in power, however, is not drastic in view of the fact that the number of tests is increased by a factor of 40.

In the above two sets of studies, causative haplotypes had the same length as the window size used for analysis, so the power would be higher than what might be expected in a real study, where the length of the disease-predisposing haplotype is unknown. Thus, we considered exhaustive testing of non-overlapping windows of 1–4 adjacent SNPs in the ENCODE data. The results are presented in

TABLE II. Type I error/power of haplotype tests with non-overlapping windows of three SNPs at the .05 nominal significance level based on the full set of SNPs on chromosome 18 of the HapMap data for studies with 500 cases and 500 controls

Odds ratio	Bon	Bon-2	MC	MC-2
1.0	.022	.015	.039	.046
1.5	.221	.296	.292	.477
1.7	.577	.649	.646	.761

Note: Bon and Bon-2 pertain to the FWER and 2-FWER based on the Bonferroni correction, and MC and MC-2 to the FWER and 2-FWER based on the Monte Carlo procedure. Each entry is based on 1,000 simulated datasets. FWER, family-wise error rate.
SNP, single nucleotide polymorphism.

TABLE III. Type I error/power of the exhaustive testing with non-overlapping windows of one–four SNPs based on the ENCODE data when the causative haplotype contains four SNPs

Odds ratio	Sample size	Bon	Bon-2	MC	MC-2
1.0	1,000	.016	.010	.051	.057
	2,000	.011	.016	.057	.060
1.5	1,000	.308	.388	.480	.567
	2,000	.755	.818	.877	.920
1.7	1,000	.677	.747	.822	.890

Note: Sample size pertains to the total number of study subjects, half of which are cases and half are controls. Bon and Bon-2 pertain to the FWER and 2-FWER based on the Bonferroni correction, and MC and MC-2 to the FWER and 2-FWER based on the Monte Carlo procedure. Each entry is based on 1,000 simulated datasets. FWER, family-wise error rate.
SNP, single nucleotide polymorphism.

Table III. The increase in the multiplicity of tests seems to cause only a slight loss of power in comparison to the non-overlapping windows of a fixed size.

We conducted another set of simulation studies to assess the sensitivity of our method to various assumptions. To increase genetic diversity, we generated data from the full set of SNPs on chromosome 18 according to the algorithm of Durrant et al. [2004]. The causative haplotype had a frequency of .18 and was located at the same window of SNPs as in the previous studies using the full set of SNPs, which started at the 637th SNP. We generated haplotypes under the following form of Hardy-Weinberg disequilibrium:

$$\pi_{kl} = \begin{cases} \pi_k^2 + \rho\pi_k(1-\pi_k), & k = l, \\ (1-\rho)\pi_k\pi_l, & k \neq l, \end{cases}$$

where $\rho = .02$ [Lin et al., 2005]. We increased the overall disease rate to 10% and decreased thresholds c_f and c_e to $1/n$ and $5/n$, respectively. We considered 10,000 windows of five SNPs, each overlapping by three SNPs. The results are presented in Table IV. The MC method continues to have correct type I error while the Bonferroni correction remains conservative. The relative power of Bon, Bon-2, MC and MC-2 has the same trend as in the previous studies.

TABLE IV. Type I error/power of haplotype tests with partially overlapping windows of five SNPs at the .05 nominal significance level under Hardy-Weinberg disequilibrium and common disease based on the full set of SNPs on chromosome 18 of the HapMap data for studies with 500 cases and 500 controls

Odds ratio	Bon	Bon-2	MC	MC-2
1.0	.021	.016	.035	.050
1.5	.108	.157	.137	.271
1.7	.434	.492	.473	.628

Note: Bon and Bon-2 pertain to the FWER and 2-FWER based on the Bonferroni correction, and MC and MC-2 to the FWER and 2-FWER based on the Monte Carlo procedure. Each entry is based on 1,000 simulated datasets. FWER, family-wise error rate.
SNP, single nucleotide polymorphism.

RHEUMATOID ARTHRITIS STUDY

Study subjects were taken from the North American Rheumatoid Arthritis Consortium (NARAC). Numerous studies [Jawaheer et al., 2004; Plenge et al., 2005] have used data from this source, and details of enrollment procedures have been published previously [Jawaheer et al., 2001]. Detailed clinical and marker data are available on the NARAC website (http://www.naracdata.org). Families in this consortium satisfied the following requirements: two or more siblings fulfilled the American College of Rheumatology (ACR) [1987] criteria for RA [Arnett et al., 1998]; at least one sibling had documented erosions on hand radiographs; and at least one sibling had disease onset between the ages of 18 and 60 years. Families with any other disease associated with similar articular symptoms, such as psoriasis or inflammatory bowel disease, were excluded.

A total of 460 cases were chosen from throughout the United States, and confirmation of RA diagnosis was obtained from patients' rheumatologists. Radiographs of the hands and wrists were also obtained to document the presence and extent of joint involvement. A total of 460 unrelated controls from Long Island were matched to the cases on the basis of age and sex. All subjects are non-Ashkenazi Caucasians. Informed consent was obtained from all subjects, and approval of the local institutional review

board was secured at every recruitment site before enrollment.

The SNPs were a custom set selected from ''double hit'' SNPs (dbSNP) on the basis of their distribution and favorable assay design characteristics. The 2,297 SNPs represent the SNPs successfully typed with minor allele frequency greater than 5% out of the 3,072 SNPs attempted in a region of chromosome 18; the region is shown in Figure 1(a). The assumption of Hardy-Weinberg equilibrium (HWE) was examined for single markers using the exact test implemented in Merlin. Several were identified with significant deviations from HWE, even though neighboring markers often showed good coincidence between observed and expected genotype frequencies. Because some significant deviations from HWE are expected by chance even when the assumption holds and departures from HWE may be caused by association between the marker alleles and disease susceptibility, we did not exclude any markers from the analysis.

We applied the proposed Monte Carlo method as well as the Bonferroni and permutation methods to this study, and considered both FWER and 2-FWER. We set $c_f = 1/n$ and $c_e = 10/n$. The results for non-overlapping windows of size 4 are summarized in Table V; only the windows with adjusted MC-2 *P*-values of less than .25 are shown. The last two windows, D and E, in the table merit special attention, as their MC-2 *P*-values are less than .1. As expected, the MC and MC-2 adjusted *P*-values are much smaller than their Bonferroni counterparts. Indeed, the Bonferroni-adjusted *P*-values are two–three fold of their MC counterparts. For this study, permutation was computationally feasible (although very slow) and yielded similar results to those of the MC method. Table VI identifies the SNPs and the most significant haplotypes in the five windows with MC-2 adjusted *P*-values of $<.25$.

TABLE V. The adjusted *P*-values for the five most significant non-overlapping windows of four SNPs in the rheumatoid arthritis study

Window	Bon	Bon-2	MC	MC-2	Perm	Perm-2
A	.694	.347	.334	.137	.341	.145
B	1.000	.580	.470	.225	.479	.242
C	1.000	.553	.455	.217	.465	.234
D	.467	.234	.248	.087	.262	.090
E	.289	.144	.163	.049	.162	.047

Note: Bon and Bon-2 pertain to the FWER and 2-FWER based on the Bonferroni correction, MC and MC-2 pertain to the FWER and 2-FWER based on the Monte Carlo procedure, and Perm and Perm-2 pertain to the FWER and 2-FWER based on permutation. SNP, single nucleotide polymorphism; FWER, family-wise error rate.

There were no significant SNPs in the single-SNP analysis, whether with the simple Bonferroni correction or the more powerful MC method. The lowest adjusted *P*-value for any single SNP was a MC-2 *P*-value of 0.16. The single-SNP analysis yielded unadjusted *P*-values of .621, .554, .077 and .151 for the four SNPs in window E, which has an unadjusted *P*-value of .0005 for the overall haplotype test and an unadjusted *P*-value of .0025 for the effect of haplotype 1111. Thus, the haplotype analysis provides much stronger evidence for genetic effects than the single-SNP analysis in this study.

We also performed the exhaustive testing of non-overlapping windows of sizes one–four, which did not produce any significant SNPs or windows. This is not surprising, as this procedure entails more than eight times as many tests as the analysis of non-overlapping windows of size 4, which had only mildly significant results. In this study, the gain in power from looking at different size windows did not compensate for the extra quantity of tests.

In summary, the proposed MC-2 method produced two adjusted *P*-values of $<.1$. This degree of significance was achieved because the analysis made use of haplotypes and 2-FWER. No adjusted *P*-value would be $<.1$ if the analysis was based on individual SNPs, traditional FWER, or Bonferroni correction.

TABLE VI. Estimated haplotype effects for the five most significant non-overlapping windows of four SNPs in the rheumatoid arthritis study

Window	SNPs	Haplotype	Frequency	Odds ratio	Unadjusted *P*-value
A	(377, 378, 379, 380)	0110	.052	.46	.00052
B	(685, 686, 687, 688)	0110	.032	1.94	.021
		1011	.147	.70	.0096
C	(1097, 1098, 1099, 1100)	0110	.280	1.43	.0006
D	(1101, 1102, 1103, 1104)	0100	.305	1.41	.00083
E	(1141, 1142, 1143, 1144)	1111	.030	2.54	.0025
		0001	.053	.67	.061

SNP, single nucleotide polymorphism.

DISCUSSION

The proposed method incorporates several new ideas: (1) a stable and efficient algorithm was constructed to calculate a proper statistic for testing haplotype-disease association for a given window of SNPs; (2) the joint distribution of such test statistics over different windows was derived; (3) a multiple-testing procedure for statistics with different degrees of freedom was introduced; (4) the concept of k-FWER was adopted; (5) an accurate MC procedure for multiple testing was developed. The concept of k-FWER is useful in genomewide association studies even if one is not interested in haplotype analysis. The MC multiple-testing procedure for statistics with different degrees of freedom also has other applications.

Like Epstein and Satten [2003] and Lin et al. [2005], our statistic for testing haplotype-disease association for a set of SNPs is based on the retrospective likelihood, which properly reflects the case-control sampling. The calculation of our test statistic makes use of a novel parameterization, which lends itself to a simple Newton-Raphson algorithm that is more efficient and more reliable than the EM algorithms used by the previous authors. More important, this article deals with haplotype analysis in association scans rather than candidate genes. The computer program implementing the new method will be posted on the website: http://www.bios.unc.edu/~lin.

Our analysis of the RA study suggested loci for further investigations. Our collaborators at the NARAC are currently genotyping an additional 667 cases and 662 controls in the regions shown in Table VI. Furthermore, an independent set of cases and controls from Europe will be used for confirmation.

The values of c_f and c_e determine which rare haplotypes are removed from the analysis and which are omitted from association testing. Lower values of c_f and c_e permit characterization of rarer haplotypes and their association with disease; however, overly low values will destabilize the algorithm. An alternative strategy is to adopt the haplotype clustering methods of Tzeng et al. [2006]. Because those methods are formulated in terms of the score test of Schaid et al. [2002] multiplied by an allocation matrix, they could be incorporated within the MC framework described here.

The selection of windows requires some thought. The number of windows to be used must be balanced against the degree of penalty for multiple testing. It may be more powerful to focus on non-overlapping windows than to consider every possible adjacent group of SNPs. One compromise is to use exhaustive testing with non-overlapping windows. The level of LD in the region of causative SNPs will also affect the power. A lower level of LD will create a greater number of common haplotypes, and thus will reduce the power to detect a true effect. In regions of high LD, non-overlapping windows will certainly have high power even if the causative haplotype happens to be out of phase with the windows. A longer causative haplotype will not be as well detected by windows of four or five SNPs as by a larger window. Testing for larger windows increases the computational intensity greatly, because of the increase in the numbers of haplotypes. The need for testing with large windows can be alleviated by using tag SNPs. If a few SNPs encode much of the variation in a region, then a small set of tag SNPs can capture the effect of a long haplotype.

We have selected windows without considering the actual LD patterns. An alternative approach is to select windows in such a way that the SNPs are in strong LD within windows and in low LD between windows. Specifically, we may select non-overlapping windows based on a definition of haplotype blocks, such as all SNPs within the block having pairwise correlation $>.8$. This allows for variable length blocks, although it entails an arbitrary definition of haplotype blocks. It would be worthwhile to investigate the performance of such strategies.

The choice of k for k-FWER will affect the interpretation of results. It is clear that using a larger k results in greater power; however, this increase in power is accompanied by an increased number of false positives. Chen and Storey [2006] discussed a similar measure, GWER-k, for linkage analysis, which is equivalent to the $(k+1)$-FWER. In their simulation studies, controlling the GWER-1 at the .05 level resulted in GWER-0 rates which ranged from .13 to .34. Considering different values of k may thus be more practical than attempting to achieve the same increase in power by increasing the α level for $k = 1$.

This article is focused on genetic effects. In some studies, investigators are interested in gene-environment interactions. By incorporating the profile likelihood approach of Lin et al. [2005], we can extend the proposed method to detect haplotype-environment interactions in genomewide association studies. In addition, we may accommodate Hardy-Weinberg disequilibrium as in Lin et al. [2005]. It would be difficult to use permutation if one is interested in testing gene-environment interactions.

The proposed MC procedure is substantially more powerful than the conventional Bonferroni correction while providing accurate control of the type I error. The MC procedure requires nearly a thousandth the computing time of the permutation procedure (with 1,000 permuted data sets) and thus can be used for studies involving large quantities of

SNPs. This is because calculating the simulated statistics only involves generation of normal random variables, whereas calculating the test statistics involves fitting the model through iterative numerical algorithms. For the RA study, it took about 320 sec on an IBM BladeCenter HS20 (Intel Xeon 2.8 GHz) machine to carry out the MC procedure for non-overlapping windows of four SNPs, as opposed to 39 h for permutation. Exhaustive testing for windows ranging from one SNP to four SNPs required 1,225 seconds.

Lin et al. [2004] considered exhaustive testing of haplotype-disease association over all possible windows of segments, and used a computationally efficient permutation procedure to assess the significance of the correlated tests. Their approach is based on a version of the transmission disequilibrium test and is applicable to family data only. Our approach can also be extended to family studies.

ACKNOWLEDGMENTS

This research was supported by the National Cancer Institute and the National Institute of Environmental Sciences. We wish to thank Drs. Peter Gregersen and Elaine Remmers, and NARAC for allowing us to use data from the rheumatoid arthritis study, which was supported by the NIH grants R01AR44422 and N01AR82232. We are also grateful to Dr. Donglin Zeng for helpful discussion.

REFERENCES

Akey J, Jin L, Xiong M. 2001. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 9: 291–300.

Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG. 1998. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31:315–324.

Chen L, Storey JD. 2006. Relaxed Significance Criteria for Linkage Analysis. Genetics 173:2371–2381.

Collins FS, Guyer MS, Chakravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. Science 278: 1580–1581.

De Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. 2005. Efficiency and power in genetic association studies. Nat Genet 37:1217–1223.

Durrant C, Zondervan KT, Cardon LR, Hunt S, Deloukas P, Morris AP. 2004. Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. Am J Hum Genet 75:35–43.

Epstein MP, Satten GA. 2003. Inference on haplotype effects in case-control studies using unphased genotype data. Am J Hum Genet 73:1316–1329.

Excoffier L, Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927.

Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'ang LY, Huang W, Liu B, Shen Y, et al. 2003. The International HapMap Project. Nature 426:789–796.

International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437:1299–1320.

Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Monteiro J, Kern M, Criswell LA, Albani S, Nelson JL, Clegg DO, Pope R, Schroeder HW, Bridges SL, Pisetsky DS, Ward R, Kastner DL, Wilder RL, Pincus T, Callahan LF, Flemming D, Wener MH, Gregersen PK. 2001. A genome-wide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. Am J Hum Genet 68:927–936.

Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. 2004. Clustering of disease features within 512 multicase rheumatoid arthritis families. Arthritis Rheum 50:736–741.

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389.

Lehmann EL, Romano JP. 2005. Generalizations of the familywise error rate. Ann Stat 33:1138–1154.

Lin DY. 2005. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 21: 781–787.

Lin DY, Zeng D, Millikan R. 2005. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet Epidemiol 29:299–312.

Lin DY, Huang BE. 2007. On the use of inferred haplotypes in downstream analyses. Am J Hum Genet 80:577–579.

Lin S, Chakravarti A, Cutler DJ. 2004. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. Nat Genet 36:1181–1188.

Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG. 2005. High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693.

Mathias RA, Gao P, Goldstein JL, Wilson AF, Pugh EW, Furbert-Harris P, Dunson GM, Malveaux FJ, Togias A, Barnes KC, Beaty TH, Huang S-K. 2006. A graphical assessment of *P*-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. BMC Genet 7:38.

Merriman TR, Cordell HJ, Eaves IA, Danoy PA, Coraddu F, Barber R, Cucca F, Broadley S, Sawcer S, Compston A, et al. 2001. Suggestive evidence for association of human chromosome 18q12-q21 and its orthologue on rat and mouse chromosome 18 with several autoimmune diseases. Diabetes 50:184–194.

Morris RW, Kaplan NL. 2002. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. Genet Epidemiol 23:221–233.

Niu T, Qin ZS, Xu X, Liu JS. 2002. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. Am J Hum Genet 70:157–169.

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Hori M, Nakamura Y, Tanaka T. 2002. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32: 650–654.

Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altshuler D, Gregersen PK, Klareskog L, Rioux JD. 2005. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with *PTPN22*, *CTLA4* and *PADI4*. Am J Hum Genet 77:1044–1060.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. Science 273:1616–1617.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. 2002. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70:425–434.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989.

Stram DO, Pearce CL, Bretsky P, Freedman M, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Thomas DC. 2003. Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. Hum Hered 55:179–190.

Thomas DC, Haile RW, Duggan D. 2005. Recent developments in genomewide association scans: a workshop summary and review. Am J Hum Genet 77:337–345.

Tzeng J-Y, Wang C-H, Kao J-T, Hsiao CK. 2006. Regression-Based Association Analysis with Clustered Haplotypes through Use of Genotypes. Am J Hum Genet 78:231–242.

Westfall PH, Young SS. 1993. Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment. New York: Wiley.

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. 2002. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 53:79–91.

Zhao LP, Li SS, Khalid N. 2003. A method for the assessment of disease associations with single-nucleotide polymorphism haplotypes and environmental variables in case-control studies. Am J Hum Genet 72:1231–1250.