Original Paper

Human Heredity

Hum Hered 2002;53:79–91

Received: September 24, 2001
Revision received: January 11, 2002
Accepted: January 23, 2002

Testing Association of Statistically Inferred Haplotypes with Discrete and Continuous Traits in Samples of Unrelated Individuals

Dmitri V. Zaykin Peter H. Westfall[b] S. Stanley Young[c] Maha A. Karnoub[a]
Michael J. Wagner[a] Margaret G. Ehm[a]

GlaxoSmithKline Inc., [a]Discovery Genetics and [b]Research Statistics Unit, Research Triangle Park, N.C. and
[c]Texas Tech University, Lubbock, Tex., USA

Key Words
Missing phase · Haplotypes · Multiple regression

Abstract
There have been increasing efforts to relate drug efficacy and disease predisposition with genetic polymorphisms. We present statistical tests for association of haplotype frequencies with discrete and continuous traits in samples of unrelated individuals. Haplotype frequencies are estimated through the expectation-maximization algorithm, and each individual in the sample is expanded into all possible haplotype configurations with corresponding probabilities, conditional on their genotype. A regression-based approach is then used to relate inferred haplotype probabilities to the response. The relationship of this technique to commonly used approaches developed for case-control data is discussed. We confirm the proper size of the test under H_0 and find an increase in power under the alternative by comparing test results using inferred haplotypes with single-marker tests using simulated data. More importantly, analysis of real data comprised of a dense map of single nucleotide polymorphisms spaced along a 12-cM chromosomal region allows us to confirm the utility of the haplotype approach as well as the validity and usefulness of the proposed statistical technique. The method appears to be successful in relating data from multiple, correlated markers to response.

Introduction

Haplotypes may be directly responsible for the observed variation in the trait of interest through the combined effects of multiple sequence variants on promoter activity or protein structure and function [8, 10, 19]. Even when a single, presumably unobserved polymorphism accounts for the trait variation, nearby markers may form haplotypes that are in much higher linkage disequilibrium (LD) with that functional polymorphism than are the individual markers because the disequilibrium between a single site and whole haplotypes includes all pairwise as well as higher order disequilibria terms [3]. Consider the following simplified example of a population with equally frequent three locus haplotypes: $A_1B_1C_1$, $A_2B_1C_2$, $A_1B_2C_2$, $A_2B_2C_1$. Here, the LD between loci pairs A, B and A, C is zero. Still, locus A is in complete LD with haplotypes formed by loci B and C (the BC pair here becomes a 'super-locus' with four alleles). The issue of haplotype utility is nevertheless controversial [6]. Analytical results and pow-

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0001–5652/02/0532–0079$18.50/0

Accessible online at:
www.karger.com/journals/hhe

Dmitri V. Zaykin
GlaxoSmithKline Inc., Department of Population Genetics
Five Moore Drive, PO Box 13398
Research Triangle Park, NC 27709 (USA)
Tel. +1 919 483 9391, Fax +1 919 315 4170, E-Mail dvz90620@gsk.com

er computations of Akey et al. [1] suggest that haplotypes can significantly improve power of association mapping. In contrast, simulation studies by Long and Langley [24] and Kaplan and Morris [20] found that single marker tests provide at least as much power. Fallin et al. [13] used statistically reconstructed haplotype frequencies for relating Alzheimer's disease with multiple single nucleotide polymorphism (SNP) markers on chromosome 19. They found examples of haplotype/disease associations that were not identified using single markers. Their results provide an example where haplotypes are more informative than a single-point analysis, even if the phase information is recovered by statistical techniques.

In this paper we focus entirely on data from unrelated individuals that do not provide full information about the gametic phase of alleles at multiple markers. There have been many descriptions of haplotype frequency inference when single-locus genotypes are scored. The expectation-maximization (E-M) algorithm, formalized by Dempster et al. [7], is a popular iterative technique for obtaining maximum likelihood estimates of sample haplotype frequencies. Little and Rubin [23] provided a general description of the E-M algorithm for multinomial data. Hill [18], MacLean and Morton [26], Hawley and Kidd [17] and Chiano and Clayton [4] discussed different variations of the algorithm. Weir and Cockerham [36] suggested that in the case of two loci it is feasible to avoid iterations and obtained an explicit equation for the maximum likelihood estimate of gametic frequencies. They pointed out situations when no real-valued solutions exist, and therefore no iterative technique may provide valid maximum likelihood estimates of haplotype frequencies. Therefore, the E-M algorithm should be used cautiously. Nevertheless, Fallin and Schork [14] performed extensive simulations showing the accuracy of the E-M algorithm for inferring haplotype frequencies on average, even when the random union of gametes assumption (also called Hardy-Weinberg equilibrium, or HWE) does not hold. Long et al. [25] discussed tests for LD and higher order interactions using E-M-inferred haplotype frequencies, giving elaborate details for implementing E-M computations in the three-locus case. Excoffier and Slatkin [11] and Slatkin and Excoffier [31] studied the importance of E-M assumptions and behavior of tests for disequilibrium using estimated frequencies.

Several authors have proposed a test that relates sample haplotype frequencies derived using the E-M algorithm with a discrete response [15, 40, 41]. This commonly used method is a two-stage procedure. First, haplotype frequencies are estimated using the E-M algorithm and sample likelihoods are evaluated in cases, controls, as well as in the pooled sample (L_1, L_2, L_3). Then the likelihood ratio (LR) test statistic, $\chi^2 = -2 \ln[L_3/(L_1\ L_2)]$, is calculated. This statistic has an asymptotic χ^2 distribution with degrees of freedom given by the number of haplotypes compatible with the sample minus one [40, 41]. The variance of this statistic may be inflated due to E-M estimation. Instead of utilizing the χ^2 approximation, the null distribution of χ^2 can be obtained by randomly shuffling the affection status among individuals and recalculating haplotype frequencies with corresponding likelihoods in cases and controls [15, 41]. Significance levels are given by the proportion of times when shuffled data produce as extreme or more extreme values of the χ^2 statistic than the original data set.

In this article we propose an association test for the situation when the response is either continuous, or discrete as in previously published research, and discuss statistical relationships to previously proposed LR approach for case-control samples. The method, briefly stated, is to relate the inferred haplotype frequencies to the observed response using a regression model. The test for association then uses an F test for a specialized additive model. The significance level can be evaluated using the F distribution or a permutation test.

The regression approach that relates probabilistically assigned predictors with trait values is not entirely new [16]. Here we present theoretical justification of the regression model by establishing its asymptotic superiority over a test based on the comparison of mean response values across haplotypes as well as the equivalence of both approaches under the null hypothesis (Appendix 1). Sasieni [29] made similar observations for the case of two alleles and binary phenotype and concluded that Armitage's trend test should be preferred over a 2×2 contingency table test based on allele counts. Thus, we extend Sasieni's results to the multiallelic case and continuous phenotype.

We study both the null hypothesis (H_0) error rates and power of the method using both simulated and real data. We confirm proper size of the test under H_0 and find an increase in power under the alternative by comparing test results using inferred haplotypes with single-marker tests using simulated data. More importantly, analysis of a large real data sample using a dense map of SNP markers distributed across a 12-cM chromosomal region allows us to confirm the utility of the haplotype approach as well as the validity and usefulness of the proposed statistical technique.

While we concentrate on haplotype frequencies obtained by the E-M algorithm [see 4 and 11, for a review of the general multilocus case], the method we propose can easily accommodate frequencies obtained by different statistical techniques [32]. We are currently working on estimation techniques that will incorporate situations when the assumption of random union of gametes does not hold. One limitation of the the E-M-based approach is that it does not allow the examination of effects due specifically to pairs of haplotypes, or 'diplotypes', which is a consequence of the HWE assumption. The diplotype frequencies can only be reconstructed from the products of estimated frequencies of haplotypes. Nevetherless, biologically realistic models (e.g. dominant or recessive models) that are not specifically haplotype driven, often result in induced (i.e. marginal) haplotype effects. From a statistical point of view, haplotype analysis can be considered as a way of reducing degrees of freedom. With full diplotype analysis, h haplotype categories result in $h(h + 1)/2$ diplotype combinations, making the large number of parameters a concern.

Methods

Statistical Techniques

The rationale for the regression test developed here stems from its asymptotic equivalence with previously proposed approaches. Haplotypes inferred in the case-control design discussed above can be arranged in a $2 \times L$ table. The two rows in the table represent the affection status and L columns correspond to the haplotypes compatible with observed genotypes in a sample of size N. Entries in the table are E-M-inferred haplotype counts that add up to $2N$. If identification of haplotypes were certain, the χ^2 test would have been analogous to the usual test of heterogeneity of allele frequencies between two groups at a multiallelic marker [39]. The test can be carried out using the LR test or the asymptotically equivalent test based on Pearson's goodness-of-fit statistics [5, 30]. Note that the individual 0-1 response that indicates whether a subject belongs to the case or to the control category is doubled, that is, each individual in a sample is counted twice. The approach is nevertheless valid if population HWE holds, because the multinomial distribution of allele counts is maintained [see 29, for details].

It is straightforward to show that the χ^2 goodness-of-fit test statistic is a simple linear function of the F statistic obtained from the analysis of variance (ANOVA) where the response is binary [34]. In fact, p values obtained from both tests are almost identical except for very small sample sizes. Therefore, in the case of continuous response it is tempting to apply the equivalent approach, the ANOVA with doubled sample size. We demonstrate, however (see Appendix 1), that while this approach is statistically valid, a more powerful test is available through our proposed haplotype trend regression (HTR). In addition to superior power, our HTR approach naturally extends to the case when haplotype frequencies are not directly observed.

To explain the HTR method simply (more details are given in Appendix 1), consider a simple example of three individuals with response values denoted as Y_1, Y_2, and Y_3 that have unambiguously resolved haplotype pairs h_1/h_1, h_2/h_3, and h_1/h_3. The regression equation $E(\mathbf{Y}) = \mathbf{D}\beta$ is then

$$E\begin{Bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{Bmatrix} = \begin{matrix} \mu & h_1 & h_2 & h_3 \end{matrix} \begin{Bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1/2 & 1/2 \\ 1 & 1/2 & 0 & 1/2 \end{Bmatrix} \begin{Bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{Bmatrix}$$

In the case where haplotypes are statistically inferred, the entries in the matrix **D** will be the inferred conditional probabilities of haplotypes given the genotype. Specifically, for haplotypes h_2 and h_3, the conditional probability of the pair (h_2, h_3) for the ith individual with genotype G_i is

$$\Pr(h_2, h_3 | G_i) = \frac{\Pr(G_i \mid h_2, h_3) p_{h_2} p_{h_3}}{\sum_{u, v} \Pr(G_i \mid h_u, h_v)\, p_{h_u} p_{h_v}}$$

where $p_{h_u} p_{h_v}$ denote haplotype frequencies. These probabilities correspond to columns 2 through 4 in the matrix **D** above. The probability $\Pr(G_i \mid h_k, h_j)$ is either 0 or 1, and so the **D** values for haplotypes incompatible with the ith subject genotype are equal to 0, and they are equal to ½ or to 1 otherwise for the case when haplotype identification is certain. In the case when haplotype frequencies are statistically inferred, the entries in **D** corresponding to haplotypes are no longer 0, ½ and 1, except for homozygous and single heterozygous subjects, reflecting the phase ambiguity. We test for the association of haplotypes with trait using the overall null hypothesis H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ and the F test or its permutational analogue. We test the individual jth haplotype effect using the same approach using the null hypothesis H_0: $\beta_j = 0$. The permutational version of the HTR requires the F statistic value for the original data set as well as F values for multiple data sets created by shuffling the response vector **Y**. The significance level is the proportion of times the F value for shuffled data sets is at least as large as the original. This computation can be done very efficiently, since most of the steps in obtaining F values for shuffled data can be pre-computed. For example, for calculating parameter estimates, we need the matrix $\mathbf{P} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ to get $\hat{\beta} = \mathbf{P}'\mathbf{Y}$. This matrix is invariant under shuffling and only needs to be computed one time.

Data Description and Laboratory Methods

137 individuals were available for analyses. There were 45 unrelated individuals, consisting of founders from families that were collected in Utah, from the CEPH collection (http://landru.cephb.fr/). Additionally, we included 92 unrelated Caucasian volunteers collected at GlaxoSmithKline, Research Triangle Park, N.C. 551 ordered SNPs in a 12-cM region on chromosome 12 were genotyped using a variety of methods, including single base chain extension (SBCE), TaqMan, oligonucleotide ligation, ABI sequencing, and polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Marker order was determined for the majority of SNPs by PCR-based scoring for the presence or absence of individual SNPs on BAC clones comprising an overlapping clone contig covering the 12-cM region with an average 4-fold redundancy of cover-

age and a mapping resolution of 40 kb pairs. Approximately one third of the SNPs locations were further refined to single base pair resolution on DNA sequence contigs ranging in size from 100 to 800 kb. The minor allele frequency range was 0.0125–0.5. The allele frequency distribution did not differ significantly from the uniform (0,1) distribution. 48% of the marker pairs separated by 50–100 kb and 68% of the markers separated by 30–50 kb were in high LD as measured by statistically significant correlation between alleles of at least 0.3 [34]. For this data set this value corresponds to the average value of Lewontin' D' of 0.75 [21]. The overall distribution of marker positions was consistent with a uniform distribution.

Type I Error Rates Using Samples from Simulated Population Frequencies of Haplotypes

We evaluated type I error rates for several scenarios which included haplotype frequencies derived from markers in linkage disequilibrium, from markers in linkage equilibrium, and from markers in linkage equilibrium but not in HWE in addition to several response distributions. Population haplotype frequencies were derived from the symmetric Dirichlet distribution, Dir($\gamma, \gamma, ..., \gamma$). This approach implicitly assumes a drift-mutation equilibrium [38]. As γ increases, the expected population haplotype diversity increases and the expected population LD decreases [see 14 for a similar approach]. The dimension of the Dirichlet distribution corresponds to the number of haplotypes. A single sample from this distribution generates a vector (**p**) of population haplotype frequencies with the property that $\Sigma p_i = 1$. Each component of the vector is then associated with a string that describes individual alleles at each SNP. For example, for two biallelic markers **p** has four frequency components with associated strings ('11', '12', '21', '22'). Once **p** is generated, a random sample of actual haplotypes can be obtained from it by multinomial sampling. Assuming HWE, sampled haplotypes are randomly paired to form individuals. We used the same simulation setup to assess power loss due to gametic phase ambiguity (Appendix 2). A complete linkage equilibrium (LE) situation was modeled by sampling individual population allele frequencies from a uniform distribution. Further, the population genotype frequencies under LE are formed from products of frequencies of alleles ± Hardy-Weinberg disequilibrium deviations. The deviations were set to either zero (HWE) or to the value that could be detected by the exact test 60% of the time in samples of 100 individuals [see 34, for discussion of the power calculations]. Each time the population frequencies were generated, we sampled genotype data for 25–100 individuals and replicated this 5,000–10,000 times. Response data was sampled from the binary, normal (0,1), mixture of normals, (0,1) and (5, 25), and normal (0,25) truncated at –5 and 5 distributions. For each replicate, we calculated the overall F test and corresponding asymptotic p value. We used 10 random E-M restarts to decrease chances of convergence to a local, nonglobal, maximum. The estimated type I error rate was the proportion of times that the null hypothesis was rejected at the 0.05 level.

Statistical Power Assuming an Explicit Population-Genetic Model ('Population Isolate')

Less extensive, but more realistic simulations assumed a drift with admixture population-genetic model that we used to form the isolated population described in [2, 33]. In the model used in the present study, we considered two starting populations (100 and 10,000 individuals for the isolate and the general populations) with no ancestral LD. Thus, the LD in our model was a consequence of the interplay of genetic drift (20 generations), migration from the larger (general) population into the isolate (first 8 generations, with a number of migrants equal to 5% of the isolate size), and recombination during the history of the population isolate. The initial values of population allele frequencies were generated from the uniform distribution. The final generation of the isolate consisted of 10,000 individuals. We performed 300 separate evolutions.

The genetic map was the same for all 300 evolutions. The first 50 of 200 SNPs were uniformly and randomly placed in the 100-kb 'candidate' region. 10 of 50 markers (markers 15–24) contributed to the value of the trait. These 10 markers defined multilocus genotypes with population means of response sampled from the normal (0, 50) distribution. Each individual in the population was given a response value sampled from the normal distribution with the population mean determined by the person's 10-SNP genotype and $\sigma^2 = 10$. The remaining 150 markers were uniformly and randomly placed on another 300-kb region and assumed to be unlinked to the first segment, thus providing means for evaluating the type I error rate. At the final generation, we sampled 100 random individuals from the isolate population for the HTR analysis. Markers directly affecting the response were removed from the analysis (i.e. assumed to be unobserved). Significant LD was observed throughout both regions dropping from the average correlation of 0.34 for neighboring markers to 0.1 for markers separated by 100 markers. In terms of D' these values were 0.80 and 0.49, correspondingly. We used a 'sliding window' approach, with 1–7 neighboring markers forming haplotypes. Empirical power was obtained by calculating the proportion of times in the 300 experiments when at least one of the sliding window p values starting from marker 43 (last 7 markers) in the response-affecting region was <0.001 for each window size. We chose these 7 distal markers to avoid high LD with the functional markers (markers 15–24). We also plotted the –log p value versus the marker map for various window sizes.

Type I Error Rates and Statistical Power Using Real Data

We examined the relative merits of haplotype-based tests using the data set of 551 markers in a 12-cM region typed on 137 individuals to provide a sampling of chromosomes from a real population. We used the marker genotypes to assign individuals to case and control categories. Our goal was to compare the performance of the HTR versus the likelihood ratio test (LRT) mentioned above. The binary phenotype was formed as follows. For a given marker, we assumed two penetrance parameters, γ for the genotype '11' with frequency P_{11}, and η for the genotypes '12' and '22', or vice versa. To imitate the case-control design with equal numbers of cases and controls, we set $\eta = 0.1$, and obtain the probability of an '11' individual being a 'case' as

$$\gamma = \frac{1/2 - \eta(1 - P_{11})}{P_{11}}$$

from $1/2 = \gamma P_{11} + \eta(P_{11} + P_{22})$, provided $P_{11} \geq 4/9$. If $P_{11} < 4/9$, we switch to the model $1/2 = \eta P_{11} + \gamma(P_{12} + P_{22})$ to satisfy restrictions that $0 \leq \eta, \gamma \leq 1$. An individual was labeled as a 'case' if a random draw from the uniform (0,1) distribution is smaller than the probability for this individual to be a 'case' given their genotype at a particular marker. Each marker in the data set was converted into a binary response gene using this procedure. We test for association of this 'response gene' with nearby markers or haplotypes using the LRT with permutation-based test statistic and both, asymptotic and permutation-based HTR using 3,200 shufflings [27].

Empirical power and error rates were obtained using the following algorithm for $i = 1, ..., 551$:

(1) Convert the ith SNP into a 'response' gene as described above.

(2) Estimate frequencies of haplotypes using 20 markers on either side of the response gene using nonoverlapping windows of sizes of 1–8 markers. Note that a window size of 1 corresponds to an allelic test.

(3) Run asymptotic and permutation versions of HTR as well as the LRT using the response variable and the estimated haplotype frequencies for the neighboring markers, and save the obtained p values (p_A) for each analysis.

(4) Randomly permute alleles at the 'response' marker, run both asymptotic and permutation versions of HTR as well as the LRT again, and save the resulting p values (p_0).

The set of p_As from the 551 experiments was used to estimate the empirical power, and the corresponding set of p_0s was used to estimate the empirical type I error rate for each window size. For a given significance level, α, the power is the proportion of $p_A \leq \alpha$ and the type I error rate is the proportion of $p_0 \leq \alpha$ among all tests for a given window size.

Results

The simulations performed to evaluate type I error rates (scenarios including haplotype frequencies derived from markers in LD, from markers in LE, and from markers in LE but not in HWE with several response distributions) confirm that the type I error rates of the method do not exceed the nominal 5% level. Asymptotic properties of the test are expected to be adversely affected as the amount of LD decreases, because it is accompanied by the increased diversity of sample haplotypes [14]. Tables 1–3 present worst case situations of complete LE under HWE as well as under deficit and excess of heterozygosity. In most cases, the type I error rate is near the nominal rate (5%). One exception is the case of a low (10%) frequency of one of the categories of the binary phenotype with a low sample size (25). Despite this, the method maintains proper size even when the response variable is not normally distributed.

Table 1. Type I error under HWE and five marker haplotypes

Response	Sample size			
	25	50	100	500
0/1, Pr(0) = 0.25	0.060	0.049	0.037	0.050
0/1, Pr(0) = 0.50	0.030	0.023	0.025	0.035
Normal	0.044	0.051	0.049	0.050
Normal mixture	0.045	0.038	0.041	0.045
Truncated normal	0.042	0.039	0.048	0.049

Table 2. Type I error under excess of heterozygotes and five marker haplotypes

Response	Sample size			
	25	50	100	500
0/1, Pr(0) = 0.25	0.056	0.049	0.044	0.049
0/1, Pr(0) = 0.50	0.033	0.033	0.036	0.039
Normal	0.049	0.056	0.053	0.051
Normal mixture	0.046	0.045	0.047	0.048
Truncated normal	0.045	0.043	0.044	0.047

Table 3. Type I error under deficit of heterozygotes and five marker haplotypes

Response	Sample size			
	25	50	100	500
0/1, Pr(0) = 0.25	0.070	0.049	0.045	0.049
0/1, Pr(0) = 0.50	0.048	0.033	0.038	0.040
Normal	0.048	0.050	0.050	0.049
Normal mixture	0.046	0.043	0.044	0.047
Truncated normal	0.041	0.047	0.038	0.049

We used the population isolate simulations to assess power. We used a 'sliding window' approach, forming haplotypes with 1–7 neighboring markers, and plotting –log(p value) on the y axis and the genetic map along the x axis. Figures 1–4 show the results obtained from analysis of 1 of the 300 replicates for window sizes of 1, 3, 5 and 7 using p values from the asymptotic F test.

We obtained empirical power values for these experiments by calculating the proportion of times when at least one of the p values in the distal region (markers 43–49) exceeded the stringent 0.001 significance level. These values were 0.221, 0.452, 0.585, 0.749, and 0.579 considering haplotypes formed by 1, 2, 3, 5 and 7 markers in the region defined by markers 43–49. When all markers are considered, the power values increase (0.726, 0.860, 0.953, 0.971, 0.933), however the relative ranking remains the same. Since none of these markers were directly affecting the response, the increase in power is attributable to higher LD between the marker haplotypes and the unobserved haplotypes related to the response. In this study, five-marker haplotypes exhibited the highest power

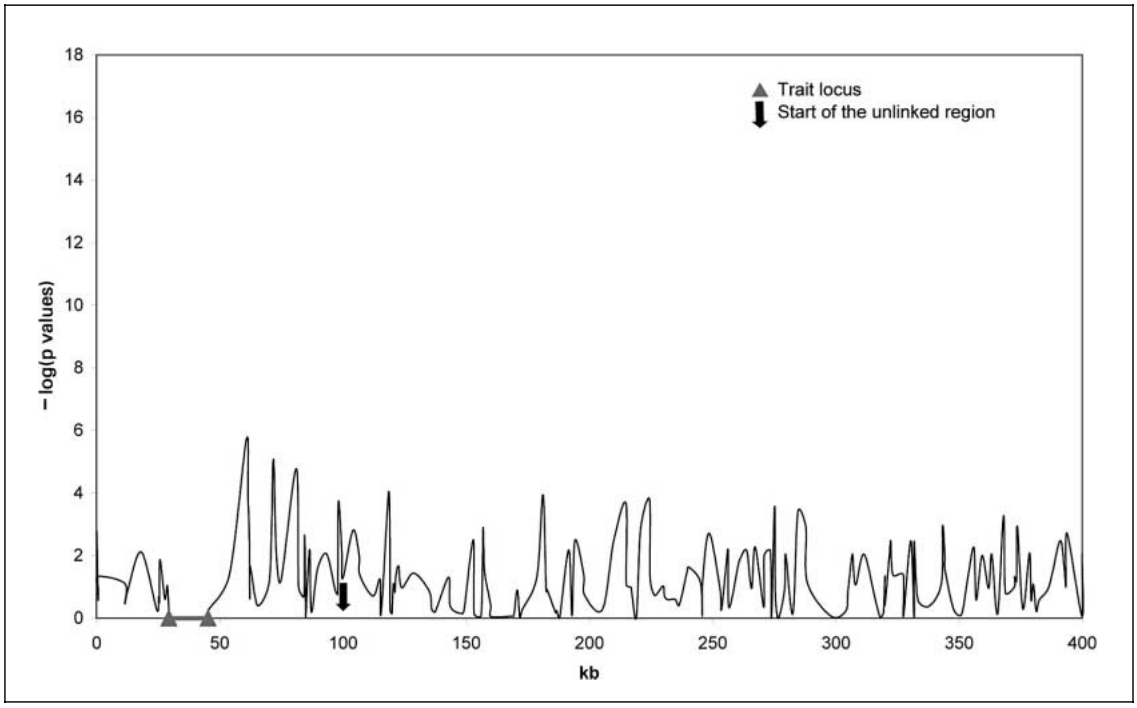

Fig. 1. Sample –log(p values) against the marker map plots for window size of 1 using p values from the asymptotic *F* test.

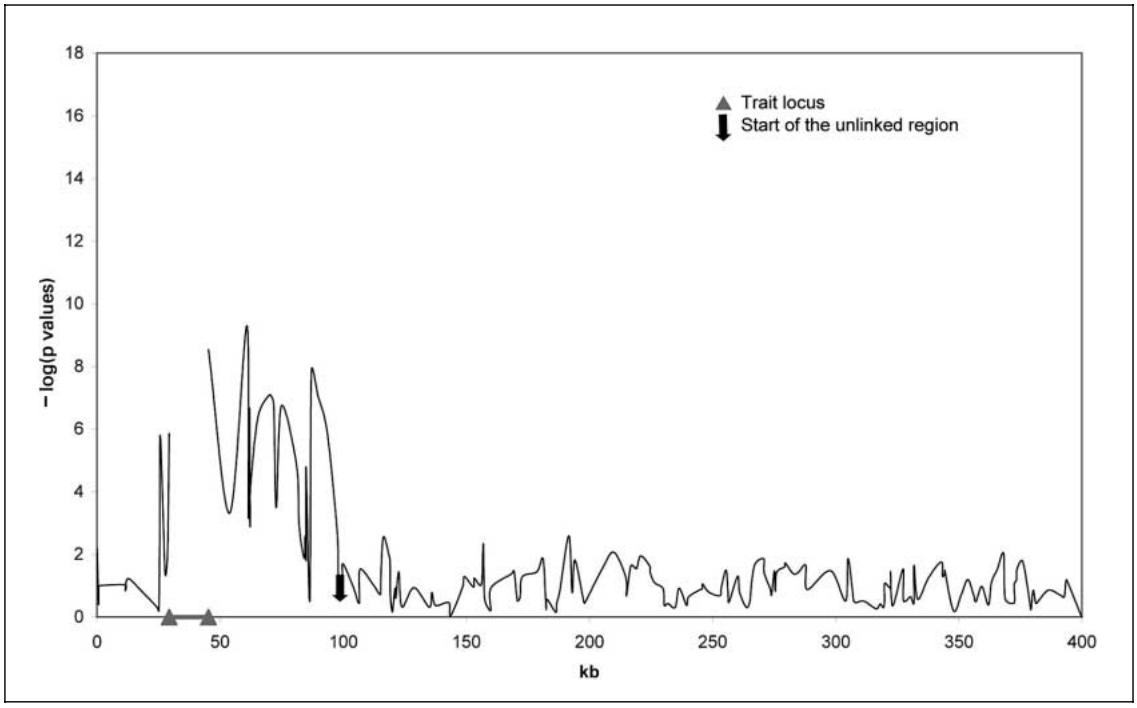

Fig. 2. Sample –log(p values) against the marker map plots for window size of 3 using p values from the asymptotic *F* test.

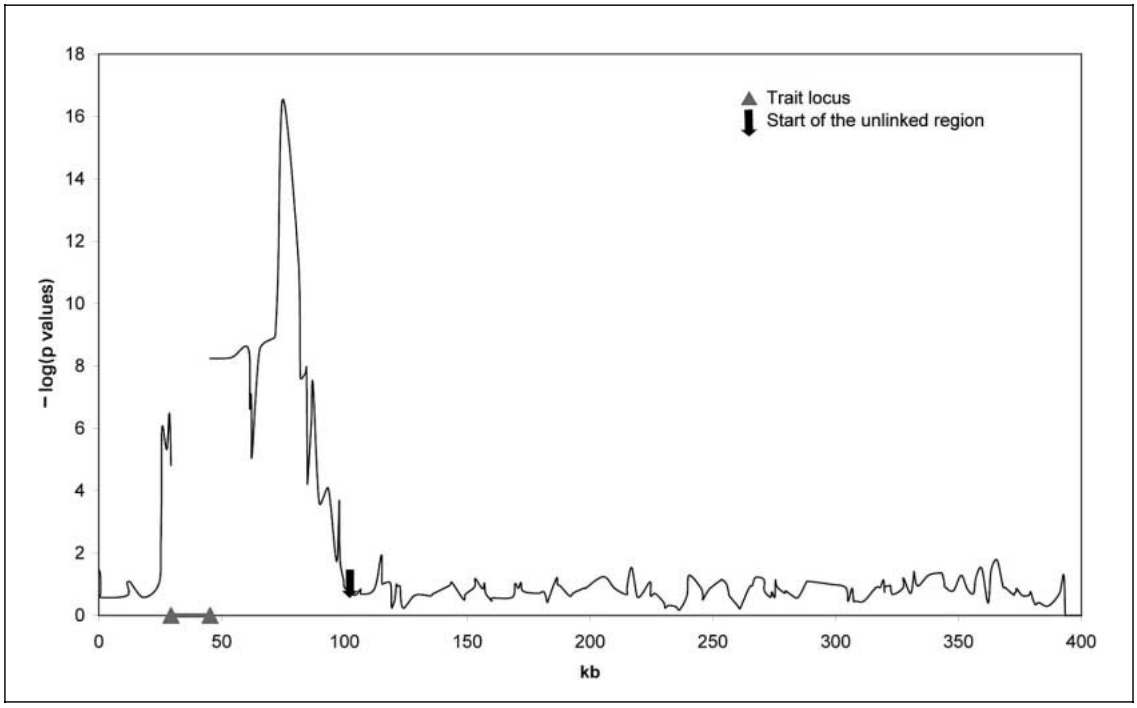

Fig. 3. Sample –log(p values) against the marker map plots for window sizes of 5 using p values from the asymptotic *F* test.

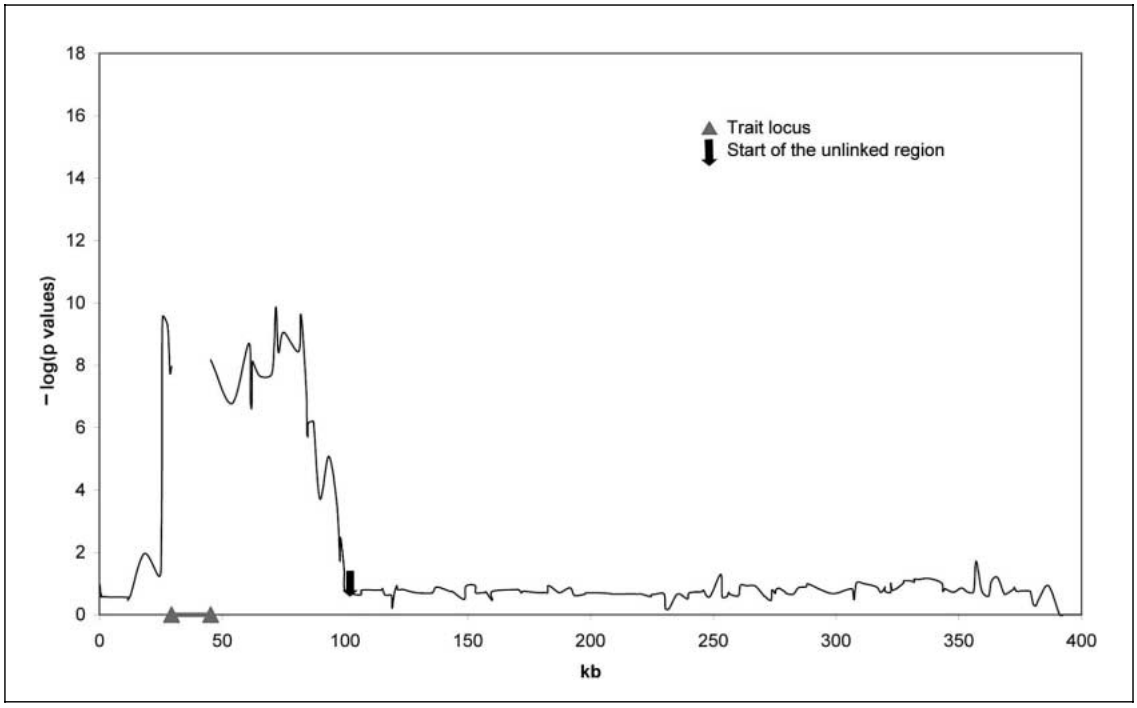

Fig. 4. Sample –log(p values) against the marker map plots for window sizes of 7 using p values from the asymptotic *F* test.

Table 4. Power values (H_A) and error rates (H_0) for the real data set

	Haplotype size					
	1	2	3	4	5	6
HTR-A (H_0)	0.056	0.034	0.033	0.029	0.022	0.027
HTR-P (H_0)	0.048	0.051	0.050	0.051	0.048	0.049
HTR-A (H_A)	0.321	0.352	0.373	0.412	0.408	0.427
HTR-P (H_A)	0.315	0.365	0.396	0.449	0.448	0.491
LRT-P (H_A)	0.310	0.357	0.388	0.420	0.444	0.436

A = Asymptotic test; P = permutational test.

reflecting the trade-off between the level of LD with functional haplotypes and the increase in degrees of freedom associated with longer haplotypes. We did not attempt to correct for the multiple testing, but simply recorded the minimum p value in the region. Still, haplotype-based tests show an increase in power. We note that the increase in power is more dramatic when only the marker 43 is used for single-marker tests, markers 43 and 44 for two-marker tests and so on.

The proportion of rejections in the unlinked region (150 markers to the right from marker 50) showed a slight increase of the type I error measured as the number of rejections over the number of tests across all markers and simulations. The increase was up to about 0.0017 for the 0.001 level and up to about 0.056 for the 0.05 level for all window sizes. This small increase is possibly due to to nonequilibrium conditions at the final generation (as a result of a relatively small number of generations and small population size during the evolution) as well as to increased probability of shared genealogy for the individuals of a similar phenotype. Devlin and Roeder [8] argue that the inflation of the type I error due to shared genealogy ('cryptic relatedness' in their terminology) can be rather substantial.

The chromosome 12 data set provides us with a unique opportunity to compare statistical methods and the utility of the haplotype approach using experimentally determined genotype data to estimate haplotypes on chromosomes drawn from a real human population, without having to resort to assumptions of simulation models. Table 4 summarizes the type I error and power estimates for the analysis of the chromosome 12 data set. We found the asymptotic LR test overly conservative for the haplotypes formed by three or more markers, because of the large degrees of freedom associated with rare haplotype classes (data not shown). We also found that a simple adjustment to the degrees of freedom by subtraction of the number of haplotypes with very low frequency leads to an anticonservative test. Fallin [12] reported similar observations. Therefore, we based our comparisons on the permutation-based version of the LR test. Note that the type I error estimates for the asymptotic or permutational HTR do not exceed 0.05. The asymptotic error rates appear to be conservative because of the increase in degrees of freedom contributed by rare haplotypes, as it is the case for the LRT [12]. The type I error rates for the permutation LRT have been studied extensively [12] and are not reported here. The power estimates for the LRT and HTR are similar with power increasing for an increased window size. While all values in table 4 were obtained using nonoverlapping windows to decrease the correlation between tests, we observed very similar results using sliding windows that shift one marker at a time (data not shown).

Discussion

The first simulation setup (Dirichlet-derived population frequencies) allowed us to investigate critical assumptions of the method without modeling population genetic evolution explicitly. We verified that the type I error is maintained for various levels of LD as well as for worst situations such as LE and nonnormal response. The asymptotic version of HTR appears to be quite robust, performing well under small sample sizes and various response models, even for binary data. The use of the binary response is motivated by the asymptotic equivalence of the F test for the one-way ANOVA and the χ^2 test for contingency tables, as discussed above. Li [22] presented related asymptotic theory results justifying F tests for dichotomous populations and confidence intervals for the coancestry coefficient [34].

The population isolate simulations were designed to compare the power and type I error rates for single marker and haplotypes of 2–7 markers. This setup gives us the ability to investigate type I error rates and power using statistically independent realizations of the same evolutionary process. This way, both statistical and actual genetic sampling [34] are taken into account. Although it is a more realistic model, it is also computationally demanding. Consequently, we only studied the case of a normally distributed phenotype. The phenotypic variation was modeled through the complex action of a single gene. The phenotypic variation is influenced by multimarker genotypes rather than haplotypes, therefore the model is not specifically in favor of haplotype tests. The type I error rates from the unlinked region obtained as averages over independent evolutions confirm the validity of the power comparisons for the haplotypes of different sizes. By looking at the minimum p value for markers within the region including markers 43–49, we are biasing our results toward haplotypes formed by smaller numbers of markers. Because the sliding window approach results in seven single-marker tests, six two-marker haplotype tests, etc. down to a single seven-marker haplotype test, only the seven-marker haplotype test power values are not inflated. An alternative approach is to consider only the marker 43 (which is the closest marker to the functional sites) when doing single-marker tests, markers 43 and 44 for two-marker tests, and so on. However, because of the variability of the LD, it is not guaranteed that the closest marker provides the most powerful test. Moreover, it is possible that the increase in power for the haplotype-based tests could be attributable to one of the markers forming the haplotype [20]. Therefore, such an approach would not address our aim to illustrate that we are indeed finding increased information with the increase of the haplotype size.

The third experiment (chromosome 12 data) was designed to investigate the power and type I error rates of the HTR using data with actual allele frequencies, intermarker linkage and higher order disequilibria. While the population genetic model simulation discussed above utilized a continuous response, this study used a binary response mimicking a case-control design. This allows us to directly compare the HTR to the LRT. Furthermore, this study allows us to determine the extent of the association of haplotypes of different sizes for this particular data set. The optimal window size for this data set appears to be about five or six, corresponding to the average size of haplotypes of about 100 kb. The power values drop to 46 and 44% for seven-marker and eight-marker haplotype tests (these values are obtained only for the permutational HTR). This 'fixed number of markers' approach is sufficient for the purpose of comparison of methods and demonstration of utility of the HTR. An alternative approach is to construct haplotypes of a variable number of markers in a window of fixed genetic distance.

The dominant/recessive model of response with a penetrance parameter and a phenocopy rate was chosen to model a complex trait using a situation that again is not specifically favoring haplotype-based tests. Note that the p values calculated from these experiments are not independent. The first 'replicate' includes the first marker as the 'response gene' along with neighboring markers. The second replicate has the second marker as the response gene with neighboring markers. Nevertheless, the same data sets are used for the HTR and LRT, and the results are comparable. Despite the nonindependence of replicates, the asymptotic type I error rate of 0.025 shown for the binary response in table 1 (sample size of 100) is similar to that shown in table 4 (0.022).

Power studies based on both simulated and real data sets indicate an increase in power provided by haplotype-based tests. The increase in power is less dramatic for the real data. This is due to sparser map and smaller LD between markers, as well as to limitations in genetic map resolution, which may result in an incorrect order of some of the markers. Results based on the chromosome 12 data set (table 4) indicate that the permutational LRT and the HTR have a similar power, with the permutational version of HTR performing somewhat better than LRT. Nevertheless, the HTR has at least two practical advantages: (1) it is readily applied to continuous and binary phenotypes providing unified approach for testing haplotype-phenotype associations and (2) it is much faster as it does not require haplotype frequency estimation at every shuffling step. The second feature is particularly important as the number of studies involving hundreds and thousands of markers is increasing. Note that it is possible to tailor HTR specifically to the binary phenotype by estimating haplotype frequencies separately in cases and controls as well as by using logistic regression instead of the ordinary least squares technique. However, significance levels would have to be obtained using the permutation test that involves E-M estimation at each shuffling, as it is in the case of the LRT. Future studies are needed to see if this will result in a more powerful test. Another potential improvement of the method would be to allow grouping of 'similar' haplotype categories to reduce degrees of freedom. Such grouping can be based on phylogenetic similarity between haplotypes and requires an evolutionary

model. Seltman et al. [37] explored this approach in the transmission/disequilibrium test framework.

Computer programs (written in C++) for the algorithms described here are available from the first author (DVZ) upon request. The programs are capable of handling variable numbers of bi- or multiple-allelic markers forming haplotypes. They also implement the overlapping sliding window of markers approach for automatic processing of large numbers of markers. Overall association tests, tests for individual haplotypes, and mean response values for haplotype classes are available with significance levels obtained through either asymptotic tests or tests based on permutation distributions.

Acknowledgements

The authors are thankful to Alan Menius, Norm Kaplan and Bruce Weir for helpful discussions. They gratefully acknowledge Michael Barnes, David Briley, Linda Briley, Daniel K. Burns, Gyan Chandra, Stephanie L. Chissoe, Gopi Gopalam, Kavita Gopalam, Kirby Gottschalk, Ray Grimaila, Steve Haneline, Tim Harris, Donald Holt, Jan Kopf, Eric H. Lai, Patrick McMillen, Mike Moore, Quan Nguyen, Melanie Reisman, Nancy Shepherd, Nigel Spurr, Pam St. Jean, Derek Traughber, and David Yarnall for technical assistance, advice and support. We thank two anonymous reviewers for helpful suggestions that improved the manuscript.

Appendix 1
The HTR and 2*N*-ANOVA Methods; Asymptotic Comparisons

Here we show the general specification of the method and demonstrate asymptotic equivalence to the double-*N* ANOVA approach under H_0, thus establishing its validity. Then we will demonstrate superiority of the HTR test under the hypothesis of association of haplotype with response.

The 2N-Based ANOVA Model

Let X denote a haplotype that may take any one of L values. Individual i has haplotypes X_{i1} and X_{i2}, and Y_i is an associated phenotype. The 'double-*N*' ANOVA model relating responses to haplotypes is $\mathbf{Y}_{(d)} = \mathbf{A}\alpha + \varepsilon$. Here $\mathbf{A}^T = (\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}, \mathbf{A}_{22}, ..., \mathbf{A}_{N1}, \mathbf{A}_{N2})$,where $\mathbf{A}_{ij}^T$ is the $1 \times L$ haplotype indicator vector for haplotype j of subject i. For example, if $X_{ij} = 2$, then $\mathbf{A}_{ij}^T = (0\ 1\ 0\ ...\ 0)$, indicating that gamete j has haplotype class 2. The elements of $\mathbf{Y}_{(d)}$ denote corresponding phenotypic observations: $\mathbf{Y}_{(d)}^T = (Y_1, Y_1, Y_2, Y_2, ..., Y_N, Y_N)$, so that the data are 'doubled'. The test statistic for associating haplotype with trait is the usual ANOVA F test

$$F' = \{SSA'/(L-1)\}/\{SSE'/(2N-L)\}$$

with

$$SSA' = \mathbf{Y}_{(d)}^T\left(\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T - \frac{1}{2N}\mathbf{J}_{2N\times 2N}\right)\mathbf{Y}_{(d)}$$

and

$$SSE' = \mathbf{Y}_{(d)}^T(\mathbf{I}_{2N} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{Y}_{(d)}$$

where $\mathbf{J}_{2N\times 2N}$ denotes the $2N \times 2N$ matrix of 1's and $\mathbf{I}_{2N}$ denotes the $2N \times 2N$ identity matrix.

The HTR Model

This model is an N-dimensional regression model $\mathbf{Y} = \mathbf{D}\beta + \varepsilon$, where Y_i is the trait value for individual i,

$$\mathbf{Y}^T = (Y_1, Y_2, ..., Y_N),$$

$$\mathbf{D}^T = (\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_N),\ \mathbf{D}_i^T = (D_{i1}, D_{i2}, ..., D_{iL}),$$

and where

$$D_{ij} = \begin{cases} 1 & \text{if } i\text{th subject is homozygous for haplotype } j \\ 1/2 & \text{if } i\text{th subject is heterozygous including haplotype } j \\ 0 & \text{otherwise} \end{cases}$$

The test statistic for associating haplotype with trait is the usual regression model F test

$$F'' = \{SSA''/(L-1)\}/\{SSE''/(N-L)\}$$

with

$$SSA'' = \mathbf{Y}^T\left(\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T - \frac{1}{N}\mathbf{J}_{N\times N}\right)\mathbf{Y}$$

and

$$SSE'' = \mathbf{Y}^T(\mathbf{I}_N - \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T)\mathbf{Y}$$

The haplotype regression model is that of Weir and Cockerham [35] in the case of no dominance deviations. Specifically, these authors assume the mean phenotypic response for genotype (j,k) is $\mu_{jk} = \mu + \alpha_j + \alpha_k + d_{jk}$, where α_j and α_k denote additive effects and d_{jk} is the dominance deviation.

Asymptotic Equivalence of HTR and 2N-ANOVA under H_0

Letting $\xrightarrow{p}$ denote convergence in probability, we show that $F'' - F' \xrightarrow{p} 0$ under H_0.

The sum of the Y_i corresponding to occurrences of haplotype j (h_j) is

$$Y^{j+} = 2\sum_{i\,\mathrm{D}\,N_{jj}} Y_i + \sum_{k\neq j}\sum_{i\,\mathrm{D}\,N_{jk}} Y_i,$$

where N_{jj} denotes the set of individuals who are homozygous for h_j, and where N_{jk} denotes the set of individuals who are heterozygous (h_j, h_k). The corresponding 'haplotype average' is $\overline{Y}^{j+} = Y^{j+}/n_j$, where n_j is jth haplotype count. Using this model, it can be seen that $\overline{Y}^{j+} = \mu_j + o_p(1) = \mu + \alpha_j + o_p(1)$ where $o_p(1)$ denotes a term that converges to zero in probability, and that $N^{1/2}(\overline{\mathbf{Y}}^+ - \mu) \xrightarrow{d} U$ where $\overline{\mathbf{Y}}^+$ denotes the vector of haplotype averages, U denotes a multivariate normal L vector, and $\xrightarrow{d}$ is convergence in distribution. We also have that $\mathbf{Y}^T\mathbf{Y}/N \xrightarrow{p} \mu^{(2)} < \infty$ and that $MSE'' \xrightarrow{p} \sigma^2 > 0$. Note that

$$F'' - F' = \frac{MSE'MSA'' - MSE''MSA'}{MSE''MSE'},$$

therefore if we can show $MSA'' - MSA' = o_p(1)$; $MSE'' - MSE' = o_p(1)$; $MSA'' \xrightarrow{d} Q$ for some random variable Q, then we can write

$$F'' - F' = \frac{o_p(1)\,MSA' + o_p(1)\,MSE'}{MSE'(MSE' + o_p(1))} = o_p(1)$$

to establish asymptotic equivalence under H_0. The H_0 condition implies that $\mu_{jk} \equiv \mu$ and for the remainder we assume without loss of generality that $\mu = 0$, since all quadratic forms in previous expressions are invariant to μ. To verify *MSE* difference conditions, it will suffice to show that corresponding differences in mean squares converge in probability to zero. Considering *SSA′* first, we note that $\overline{\mathbf{Y}}^{+} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}_{(d)}$. Letting $\mathbf{D}_A = \mathbf{A}^T\mathbf{A} = \text{diag}\{n_j\}$, we have $\mathbf{Y}^T_{(d)}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}_{(d)} = (\overline{\mathbf{Y}}^{+})^T\mathbf{D}_A\overline{\mathbf{Y}}^{+}$. Noting that $\mathbf{J}_{1\times 2N}\mathbf{Y}_{(d)} = 2\Sigma Y_i = 2\mathbf{J}_{1\times N}\mathbf{Y}$, and that $\mathbf{J}_{1\times L}\mathbf{D}_A\overline{\mathbf{Y}}^{+} = 2\Sigma Y_i$, we have

$$\frac{1}{2N}\mathbf{Y}^T_{(d)}\mathbf{J}_{2N\times 2N}\mathbf{Y}_{(d)} = \frac{1}{2N}(\overline{\mathbf{Y}}^{+})^T\mathbf{D}_A\mathbf{J}_{L\times L}D_A\overline{\mathbf{Y}}^{+}.$$

Thus

$$SSA' = (\overline{\mathbf{Y}}^{+})^T\left(\mathbf{D}_A - \frac{1}{2N}\mathbf{D}_A\mathbf{J}_{L\times L}\mathbf{D}_A\right)\overline{\mathbf{Y}}^{+}$$

$$= 2(N^{1/2}(\overline{\mathbf{Y}}^{+})^T)\left(\frac{\mathbf{D}_A}{2N} - \frac{\mathbf{D}_A\mathbf{J}_{L\times 1}\mathbf{J}_{1\times L}\mathbf{D}_A}{2N}\right)(N^{1/2}\overline{\mathbf{Y}}^{+})$$

$$= 2(N^{1/2}(\overline{\mathbf{Y}}^{+})^T)\,\mathbf{A}_N\,(N^{1/2}\overline{\mathbf{Y}}^{+})$$

and under HWE we have $\mathbf{A}_N = \mathbf{P} - \mathbf{p}\mathbf{p}^T + o_p(1)$, where $\mathbf{P} = \text{diag}\{p_j\}$ and $\mathbf{p}^T = (p_1, ..., p_L)$, with p_j denoting the population proportion of haplotype j. Now, considering *SSA″*, note that $\mathbf{D}'\mathbf{Y} = \mathbf{D}_A\overline{\mathbf{Y}}^{+}$, so that

$$SSA'' = (\overline{\mathbf{Y}}^{+})^T\left(\mathbf{D}_A(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}_A - \frac{1}{4N}\mathbf{D}_A\mathbf{J}_{L\times L}\mathbf{D}_A\right)\overline{\mathbf{Y}}^{+}$$

$$= 2(N^{1/2}(\overline{\mathbf{Y}}^{+})^T)\left(\frac{\mathbf{D}_A}{2N}\frac{(\mathbf{D}^T\mathbf{D})^{-1}}{2N}\frac{\mathbf{D}_A}{2N} - \frac{1}{2N}\frac{\mathbf{D}_A\mathbf{J}_{L\times 1}\mathbf{J}_{1\times L}\mathbf{D}_A}{2N}\right)(N^{1/2}\overline{\mathbf{Y}}^{+})$$

$$= 2(N^{1/2}(\overline{\mathbf{Y}}^{+})^T)\mathbf{B}_N(N^{1/2}\overline{\mathbf{Y}}^{+})$$

To find the limit of $\mathbf{B}_N$, we first calculate that

$$\frac{1}{2N}\mathbf{D}^T\mathbf{D} \overset{\text{HWE}}{=} \mathbf{P} + \mathbf{p}\mathbf{p}^T = o_p(1).$$

Since $\mathbf{P} + \mathbf{p}\mathbf{p}^T$ is invertable (positive definite in fact), under the random union of haplotypes we have

$$2N(\mathbf{D}^T\mathbf{D})^{-1} \overset{\text{HWE}}{=} (\mathbf{P} + \mathbf{p}\mathbf{p}^T)^{-1} + o_p(1) = \mathbf{P}^{-1} - \frac{1}{2}\mathbf{J}_{L\times L} + o_p(1)$$

Thus

$$\mathbf{B}_N = \mathbf{P}\left(\mathbf{P}^{-1} - \frac{1}{2}\mathbf{J}_{L\times L}\right)\mathbf{P} - \frac{1}{2}\mathbf{p}\mathbf{p}^T + o_p(1)$$

$$= \mathbf{P} - \mathbf{p}\mathbf{p}^T + o_p(1)$$

Hence, $\mathbf{A}_N - \mathbf{B}_N = o_p(1)$. Using this and the result that $N^{1/2}(\overline{\mathbf{Y}}^{+} - \mu) \xrightarrow{d} U$ we see that

$$SSA' - SSA'' = 2(N^{1/2}(\overline{\mathbf{Y}}^{+})^T)\,[o_p(1)]\,(N^{1/2}\overline{\mathbf{Y}}^{+}) = o_p(1).$$

Next, consider

$$MSE' = \frac{2N}{2N-L}\left(\frac{\mathbf{Y}^T_{(d)}\mathbf{Y}_{(d)}}{2N} - \frac{\mathbf{Y}^T_{(d)}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}_{(d)}}{2N}\right)$$

$$= (1 + o_p(1))\left[\frac{\mathbf{Y}^T\mathbf{Y}}{2N} - \frac{(\overline{\mathbf{Y}}^{+})^T\mathbf{D}_A\overline{\mathbf{Y}}^{+}}{2N}\right]$$

and

$$MSE'' = (1 + o_p(1))\left[\frac{\mathbf{Y}^T\mathbf{Y}}{N} - \frac{(\overline{\mathbf{Y}}^{+})^T\mathbf{D}_A(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}_A\overline{\mathbf{Y}}^{+}}{N}\right],$$

the difference is

$$MSE'' - MSE' = o_p(1) + (\mathbf{Y}^{+}_A)^T\left[\frac{\mathbf{D}_A(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}_A}{N} - \frac{\mathbf{D}_A}{N}\right]\mathbf{Y}^{+}_A$$

$$= o_p(1) + (\mathbf{Y}^{+}_A)^T\left[2(\mathbf{P} - \frac{1}{2}\mathbf{p}\mathbf{p}^T) - \mathbf{P} + o_p(1)\right]\mathbf{Y}^{+}_A$$

$$= o_p(1) + (\mathbf{Y}^{+}_A)^T\,[\mathbf{P} - \mathbf{p}\mathbf{p}^T]\,\mathbf{Y}^{+}_A$$

$$= o_p(1)$$

since $\mathbf{Y}^{+}_A \xrightarrow{p} 0$. Now, $MSA'' \xrightarrow{d} Q$ follows by noting that $N^{1/2}\overline{\mathbf{Y}}^{+}$ converges in distribution and that the elements of $\mathbf{B}_N$ converge in probability, establishing equivalence of HTR and 2*N*-ANOVA tests under H_0.

Superiority of HTR over 2N-ANOVA under the Alternative

Denote $V_\alpha = \Sigma\, p_j\alpha_j^2 - (\Sigma\, p_j\alpha_j)^2$. Then

$$\frac{1}{2N}SSA'' = (\overline{\mathbf{Y}}^{+})^T\mathbf{B}_N\overline{\mathbf{Y}}^{+} \xrightarrow{p} \mu(\mathbf{P} - \mathbf{p}\mathbf{p}^T)\,\mu = V_\alpha$$

and

$$\frac{1}{2N}SSA' = (\overline{\mathbf{Y}}^{+})^T\mathbf{A}_N\overline{\mathbf{Y}}^{+} \xrightarrow{p} V_\alpha$$

Hence, $SSA'' - SSA' \xrightarrow{p} 1$. Now when we take the ratio,

$$\frac{MSE'}{MSE''} = \frac{MSE' - MSE''}{MSE''} + 1 \xrightarrow{p} \frac{V_\alpha}{\sigma^2} + 1,$$

implying the result $\frac{F''}{F'} \xrightarrow{p} \text{const} > 1$.

Appendix 2
Power Loss due to Haplotype Phase Uncertainty

Generally, statistical techniques used to recover phase information will cause decrease in power as compared to the situation when haplotype phase is directly observed. Douglas et al. [9] studied the decrease of information due to phase uncertainty in the case of complete LE. The power may also be adversely affected by the specifics of statistical techniques used to relate traits with the haplotype information. The probabilistic assignment of haplotype pairs to individuals conditioned on sample frequency estimates may lead to some bias, especially in the case-control samples where frequencies in two groups can be substantially different. We investigated the loss of power under a phenotype model driven by haplotypes. The mean phenotypic response for genotype (j,k) was given by $\mu_{jk} = \alpha_j + \alpha_k$, where α_j and α_k denote effects of two haplotypes. These effects were sampled from $N(0,1)$ distribution [28] for each simulation run, forming a normal distribution of population effects across simulations. The phenotype value for ith individual in each sample was con-

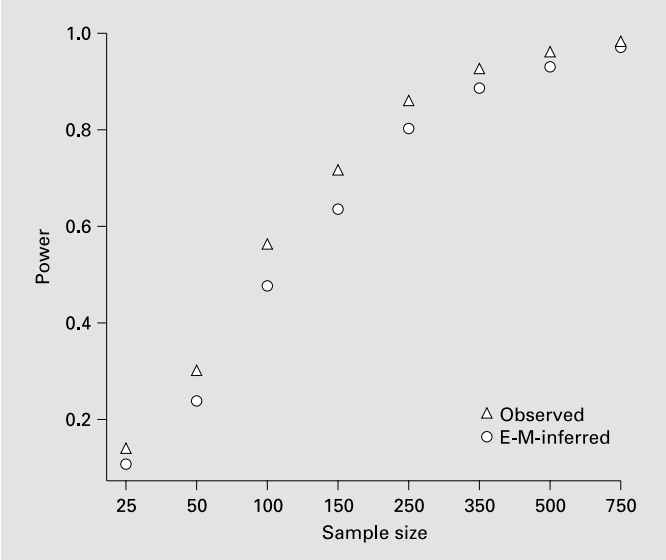

Fig. 5. Power values against the sample size for observed and E-M-inferred three marker haplotypes (HTR tests).

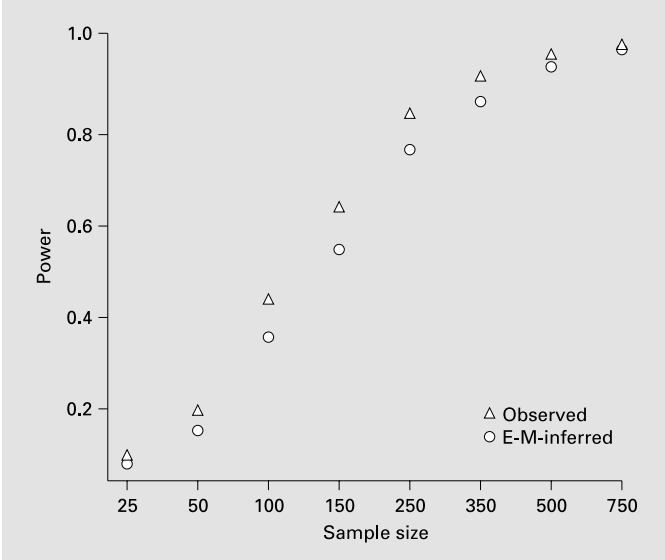

Fig. 6. Power values against the sample size for observed and E-M-inferred five marker haplotypes (HTR tests).

structed as $Y_i = m_{jk} + \varepsilon_i$, with $\varepsilon_i \sim N(0,2)$. Population frequencies of haplotypes were sampled from a symmetric Dirichlet distribution. The common parameter was obtained empirically to form the average population distribution of Lewontin's D' with the mean value of 0.36. We studied 3- and 5-SNP haplotypes. Multinomial samples of multilocus genotypes were obtained assuming population HWE. Prior to analysis, phenotypic values were dichotomized about the sample mean, thus leading to a case-control design. In the case of observed haplotypes and binary response, the HTR reduces to a test that compares allele frequencies between two categories and at least as powerful as the LRT (data not shown). Power values for different sample sizes (based on at least 8,500 simulations for each sample size) were obtained using samples of 50–1,500 haplotypes (fig. 5, 6). Triangles on the graphs show power values for the 'ideal' situation when haplotypes are directly observed, so the haplotypes are treated simply as multiallelic markers. Circles represent power values of HTR obtained for the same data sets. As expected, some power is lost for the range of sample sizes that result in intermediate power values. The loss is not great. The graphs illustrate that less than 100 extra individuals are required to match the power.

References

1 Akey J, Jin L, Xiong M: Haplotypes vs single marker linkage disequilibrium tests: What do we gain? Eur J Hum Genet 2000;9:291–300.

2 Almasy L, Terwilliger JD, Nielsen D, Dyer TD, Zaykin D, Blangero J: GAW12: Simulated genome scan, sequence, and family data for a common disease. Genet Epidemiol 2001;21 (suppl 1):S332–338.

3 Bennett JH: On the theory of random mating. Ann Eugen 1954;18:311–317.

4 Chiano MN, Clayton DG: Fine genetic mapping using haplotype analysis and the missing data problem. Am J Hum Genet 1998;62:55–65.

5 Cressie N, Read TRC: Multinomial goodness of fit tests. J R Stat Assoc B 1984;46:440–464.

6 Davidson S: Research suggests importance of haplotypes over SNPs. Nature Biotechnol 2000;18:1134–1135.

7 Dempster AP, Laird NM, Rubin D: Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 1977;39:1–38.

8 Devlin B, Roeder K: Genomic control for association studies. Biometrics 1999;55:997–1004.

9 Douglas JA, Boehnke M, Gillanders E, Trent JM, Gruber SB: Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. Nature Genet 2001;28:361–364.

10 Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB: Complex promoter and coding region b2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 2000;97:10483–10488.

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

12 Fallin D: Haplotype-Based Approaches for Genetic Case-Control Studies; dissertation Case Western Reserve University, Cleveland, 2000.

13 Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, Schork NJ: Genetic analysis of case/control data using estimated haplotype frequencies: Application to APOE locus variation and Alzheimer's disease. Genome Res 2001;11:143–151.

14 Fallin D, Schork NJ: Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am J Hum Genet 2000;67:947–959.

15 Fallin D, Schork NJ: Power of omnibus likelihood ratio test for haplotype-based case-control studies (abstract). Am J Hum Genet 2000; 67(S2):214.
16 Haley CS, Knott SA: A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 1992; 69:315–324.
17 Hawley ME, Kidd KK: HAPLO: A program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J Hered 1995;86:409–411.
18 Hill WG: Estimation of linkage disequilibrium in randomly mating populations. Heredity 1974;33:229–239.
19 Joosten PHLJ, Toepoel M, Mariman ECM, Van Zoelen EJJ: Promoter haplotype combinations of the platelet-derived growth factor α-receptor gene predispose to human neural tube defects. Nat Genet 2001;27:215–217.
20 Kaplan N, Morris R: Issues concerning association studies for fine mapping a susceptibility gene for a complex disease. Genet Epidemiol 2001;20:432–457.
21 Lewontin RC: The interaction of selection and linkage I. General considerations; heterotic models. Genetics 1964;49:49–67.
22 Li YJ: Characterizing the Structure of Genetic Populations; PhD thesis North Carolina State University, Raleigh, 1996.
23 Little RJA, Rubin DB: Statistical Analysis with Missing Data. Chichester, Wiley, 1987.
24 Long AD, Langley CH: The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res 1999;9:720–731.
25 Long JC, Williams RC, Urbanek M: An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 1995;56: 799–810.
26 MacLean CJ, Morton NE: Estimation of myriad haplotype frequencies. Genet Epidemiol 1985;2:263–272.
27 Maiste PJ: Comparison of Statistical Tests for Independence at Genetic Loci with Many Alleles; PhD thesis North Carolina State University, Raleigh, 1993.
28 Otto SP, Jones CP. Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. Genetics 2000;156:2093–2107.
29 Sasieni PD: From genotypes to genes: Doubling the sample size. Biometrics 1997;53: 1253–1261.
30 Sen PK, Singer JM: Large Sample Methods in Statistic: An Introduction with Applications. London, Chapman & Hall, 1993.
31 Slatkin M, Excoffier L: Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 1996;76:377–383.
32 Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001; 68:978–989.
33 Thomas DC, Borecki IB, Thomson G, Weiss K, Almasy L, Blangero J, Nielsen D, Terwilliger J, Zaykin D, Macluer J: Evolution of the simulated data problem. Genet Epidemiol, in press.
34 Weir BS: Genetic Data Analysis II. Sunderland, Ma., Sinauer, 1996.
35 Weir BS, Cockerham CC: Two-locus theory in quantitative genetics; in Pollak E, Kempthorne O, Bailey TB (eds): Proceedings of the International Conference on Quantitative Genetics. Ames, Iowa State University Press, 1977, pp 247–269.
36 Weir BS, Cockerham CC: Estimation of linkage disequilibrium in randomly mating populations. Heredity 1979;42:105–111.
37 Seltman H, Roeder K, Devlin B: Transmission/disequilibrium test meets measured haplotype analysis: Family-based association analysis guided by evolution of haplotypes. Am J Hum Genet 2001;68:1250–1263.
38 Wright S: The genetical structure of populations. Ann Eugen 1951;15:323–354.
39 Wokman PL, Niswander JD: Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. Am J Hum Genet 1970;22:24–49.
40 Xie X, Ott J: Testing linkage disequilibrium between a disease gene and marker loci (abstract). Am J Hum Genet 1993;53(S):1107.
41 Zhao JH, Curtis D, Sham PC: Model-free analysis and permutation tests for allelic associations. Hum Hered 2000;50:133–139.