

Enhancing the Power to Detect Low-Frequency Variants in Genome-Wide Screens

Chang-Yun Lin,^{*,†,1} Guan Xing,^{*,†,1} Hung-Chih Ku,^{*} Robert C. Elston,[§] and Chao Xing^{*,1,2}

^{*}McDermott Center of Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas 75390;

[†]Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung, Taiwan; [‡]Bristol-Myers Squibb Company, Pennington, New Jersey 08534; [§]Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

ABSTRACT In genetic association studies a conventional test statistic is proportional to the correlation coefficient between the trait and the variant, with the result that it lacks power to detect association for low-frequency variants. Considering the link between the conventional association test statistics and the linkage disequilibrium measure r^2 , we propose a test statistic analogous to the standardized linkage disequilibrium D' to increase the power of detecting association for low-frequency variants. By both simulation and real data analysis we show that the proposed D' test is more powerful than the conventional methods for detecting association for low-frequency variants in a genome-wide setting. The optimal coding strategy for the D' test and its asymptotic properties are also investigated. In summary, we advocate using the D' test in a dominant model as a complementary approach to enhancing the power of detecting association for low-frequency variants with moderate to large effect sizes in case-control genome-wide association studies.

IN testing for association between a trait and a genetic variant, conventional methods fit to data generalized linear models, *e.g.*, logistic regression models for a binary trait and linear regression models for a continuous trait (Balding 2006). Under the assumption of an additive genetic model, the power of the test is proportional to the locus-specific narrow sense heritability—the proportion of the trait variance explained by the additive genetic effect of the variant (Chapman *et al.* 2003)—which is essentially the square of the linear correlation coefficient r^2 between the trait and the variant. In this article we call tests based on the correlation coefficient r^2 -type tests. The power of an r^2 -type test can be shown to be proportional to effect size² $\times p(1-p)$, where p denotes the allele frequency (Chapman *et al.* 2003; Spencer *et al.* 2009). The power of an r^2 -type test decreases as the minor allele frequency (MAF) of the variant decreases,

which is one of the reasons genome-wide association studies (GWAS) mostly identify associations with common variants.

Association of alleles at different loci on the same chromosome is referred to as gametic phase disequilibrium, or linkage disequilibrium (LD). Various measures of LD are proposed in the literature (Hedrick 1987; Devlin and Risch 1995; Mueller 2004), among which two are commonly used, the correlation coefficient r^2 (Hill and Robertson 1968) and the standardized gametic disequilibrium coefficient D' (Lewontin 1964). Numerous studies compare the two measures from the perspectives of inferring population evolutionary history and mapping genes (for a review, see *e.g.*, Slatkin 2008). A consensus concerning the two measures is that D' is more robust to allele frequency. Although neither measure is independent of the allele frequency (Edwards 1963; Lewontin 1988), the range of D' is always between -1.0 and $+1.0$, regardless of allele frequency, whereas the limits of r^2 depend on the allele frequencies at both loci. D' more accurately reflects the recombination patterns (Hedrick 1988), in particular in the case of low allele frequencies, which can facilitate pinpointing disease-causing mutations (Devlin and Risch 1995; Guo 1997). Note that in the case of small sample sizes and low allele frequencies, the absolute magnitude of D' tends to be overestimated (Zapata *et al.* 2001; Teare *et al.* 2002; Pe'er *et al.* 2006).

Copyright © 2014 by the Genetics Society of America

doi: 10.1534/genetics.113.160739

Manuscript received December 17, 2013; accepted for publication January 26, 2014; published Early Online February 4, 2014.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.160739/-/DC1>.

¹Equal contribution.

²Corresponding author: MC 8591, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390. E-mail: chao.xing@utsouthwestern.edu

Inspired by these characteristics of D' , we propose an alternative approach to gene mapping using statistics based on D' instead of the conventional r^2 -type tests. We hypothesized that a D' -type test would be more powerful than r^2 -type tests in detecting association for low-frequency variants.

Method

Consider two diallelic loci A and B, with alleles A_1 and A_2 at locus A and alleles B_1 and B_2 at locus B, *i.e.*, the factors A and B in Table 1 stand for loci A and B, respectively. Denote by $p_{k\cdot}$ and $p_{\cdot k}$, where $k \in \{1, 2\}$, allele (relative) frequencies, and by p_{ij} , where $i, j \in \{1, 2\}$, haplotype (relative) frequencies. The gametic disequilibrium coefficient is defined as

$$D = p_{11}p_{22} - p_{12}p_{21} \quad (1)$$

(Lewontin and Kojima 1960). The measure

$$r^2 = \frac{D^2}{p_{1\cdot}p_{2\cdot}p_{\cdot 1}p_{\cdot 2}} \quad (2)$$

is the square of Pearson's correlation coefficient that measures the linear dependence between the two loci (Hill and Robertson 1968). The measure

$$D' = \frac{D}{D_{\max}}, \quad (3)$$

where $D_{\max} = \min(p_{1\cdot}p_{2\cdot}, p_{2\cdot}p_{1\cdot})$ when $D > 0$ and $D_{\max} = \min(p_{1\cdot}p_{\cdot 1}, p_{2\cdot}p_{\cdot 2})$ when $D < 0$, is a standardized D obtained on dividing it by its theoretical maximum given the allele frequencies (Lewontin 1964).

Let factor B in Table 1 stand for the affection status of a disease, with B_1 denoting cases and B_2 denoting controls, and consider association between the disease and the alleles, assumed to be independent, of a diallelic locus A. The widely used Pearson's χ^2 test statistic is defined as $\chi^2 = \sum_{i,j} [(n_{ij} - \hat{e}_{ij})^2 / \hat{e}_{ij}]$, where $\hat{e}_{ij} = n_i \cdot n_{\cdot j} / n$ and $i, j \in \{1, 2\}$.

It can be shown that $\chi^2 = n\hat{r}^2$, where \hat{r}^2 is the estimate of the squared Pearson correlation coefficient as defined in (2). Therefore, using this method to test association between the disease and a marker is analogous to measuring LD between two loci by r^2 . The range of r^2 for a 2×2 table depends on the frequencies of both factors and it reaches 1 only when the two factors are perfectly correlated. We speculated that this is one reason that conventional r^2 -type methods have low power to detect association for low-frequency variants, in particular in a case-control study where cases are over-sampled. In contrast, D' has the same range for all frequencies and its estimate tends to be inflated in the case of low allele frequencies. We hypothesized that a test based on D' will be more powerful than the r^2 -type methods for detecting association with low-frequency variants in a case-control GWAS.

Table 1 Notation for counts/probabilities in a 2×2 table

Factor A	Factor B		
	B_1	B_2	
A_1	n_{11}/p_{11}	n_{12}/p_{12}	$n_{1\cdot}/p_{1\cdot}$
A_2	n_{21}/p_{21}	n_{22}/p_{22}	$n_{2\cdot}/p_{2\cdot}$
	$n_{\cdot 1}/p_{\cdot 1}$	$n_{\cdot 2}/p_{\cdot 2}$	$n/1$

A Wald-type test statistic is defined as

$$\chi_{D'}^2 = \frac{[\hat{D}' - E_0(\hat{D}')]^2}{\text{Var}(\hat{D}')},$$

which asymptotically follows a χ^2 distribution with 1 degree of freedom. For simplicity we refer to this test as the D' test. D' can be estimated as \hat{D}' by plugging estimates into formula (3). It is a ratio of two random variables and both its mean and variance can be obtained by the delta method using the first-order Taylor's series expansions

$$E(\hat{D}') \approx \frac{E(\hat{D})}{E(\hat{D}_{\max})}$$

and

$$\begin{aligned} \text{Var}(\hat{D}') \approx & \frac{1}{E^2(\hat{D}_{\max})} \text{Var}(\hat{D}) + \frac{E^2(\hat{D})}{E^4(\hat{D}_{\max})} \text{Var}(\hat{D}_{\max}) \\ & - 2 \frac{E(\hat{D})}{E^3(\hat{D}_{\max})} \text{Cov}(\hat{D}, \hat{D}_{\max}) \end{aligned}$$

(Casella and Berger 2001). Under the null hypothesis of no association between the disease and marker, $E(\hat{D}) = 0$; therefore, $E_0(\hat{D}') \approx 0$. In particular, in a balanced case-control study, $\hat{D}_{\max} = \min(\hat{p}_{1\cdot}/2, \hat{p}_{2\cdot}/2)$ is a constant regardless of the sign of \hat{D} , and thus $E_0(\hat{D}') = 0$. $\text{Var}(\hat{D}')$ has been obtained (Zapata *et al.* 1997; Zapata *et al.* 2001) as

$$\text{Var}(\hat{D}') \approx \frac{(1 - |\hat{D}'|) [n \text{Var}(\hat{D}) - |\hat{D}'| \hat{D}_{\max} (\alpha \hat{p}_{1\cdot} + \beta \hat{p}_{2\cdot} - 2|\hat{D}'|)] + |\hat{D}'| \gamma (1 - \gamma)}{n \hat{D}_{\max}^2}, \quad (4)$$

where

$$\text{Var}(\hat{D}) \approx \frac{\hat{p}_{1\cdot} \hat{p}_{2\cdot} \hat{p}_{\cdot 1} \hat{p}_{\cdot 2} + \hat{D}(1 - 2\hat{p}_{1\cdot})(1 - 2\hat{p}_{\cdot 1}) - \hat{D}^2}{n}$$

(Hill 1974); $\alpha = \hat{p}_{\cdot 2}$ and $\beta = \hat{p}_{\cdot 1}$ when $\hat{D}' < 0$; $\alpha = \hat{p}_{\cdot 1}$ and $\beta = \hat{p}_{\cdot 2}$ when $\hat{D}' > 0$; and $\gamma = \hat{p}_{11}, \hat{p}_{12}, \hat{p}_{21}$, and \hat{p}_{22} when $\hat{D}_{\max} = \hat{p}_{1\cdot} \hat{p}_{\cdot 1}, \hat{p}_{1\cdot} \hat{p}_{\cdot 2}, \hat{p}_{2\cdot} \hat{p}_{\cdot 1}$, and $\hat{p}_{2\cdot} \hat{p}_{\cdot 2}$, respectively. Note

Table 2 Comparison of the D' -type and r^2 -type measures of association and corresponding test statistics, in a 2×2 table of low-frequency exposure data

n_{12}	\hat{D}'	$\chi_{D'}^2$	\hat{r}^2	$\chi_{r^2}^2$
10	0	0	0	0
9	0	0	0.000	0.055
8	0.056	0.058	0.001	0.233
7	0.118	0.249	0.001	0.553
6	0.188	0.606	0.003	1.042
5	0.267	1.188	0.004	1.732
4	0.357	2.107	0.007	2.665
3	0.462	3.602	0.010	3.896
2	0.583	6.297	0.014	5.498
1	0.727	12.48	0.019	7.572
0	0.900	42.79	0.026	10.26

The structure of data are as in Table 1, with $n_{\cdot 1} = n_{\cdot 2} = 200$, $n_{11} = 10$, and n_{12} equal to the number indicated in the first column; 0.5 was added to the cell in the table with the smallest number, as described in *Methods*, before calculating the entries of this table.

that at some points the parameters lie on the boundary of the parameter space (Mangin *et al.* 2008). When $D = 0$ the asymptotic variance of D' cannot be obtained by the delta method because D_{\max} is not continuous at this point; however, from a practical point of view $\hat{D}' = 0$ has a point mass with zero probability. For the sake of completeness, we define $\chi_{D'}^2 = 0$ at $\hat{D}' = 0$ with a probability of zero. When one cell equals 0, or when both diagonal/antidiagonal cells = 0, $|\hat{D}'| = 1$ and $\text{Var}(\hat{D}') = 0$, and the test statistic $\chi_{D'}^2$ is undefined. One way to circumvent this is to add 0.5 to each cell. Considering the known upward bias of \hat{D}' in the case of small sample sizes and low allele frequencies, we take a more conservative approach by adding 0.5 to the cell in the table with the smallest number while keeping the marginal totals fixed; when there are two such cells belonging to the same row or column, no penalty will be taken.

First we provide a numerical example illustrating differences between the D' -type measures and the r^2 -type measures for association of low-frequency exposures in a case-control study under both null and alternative hypotheses. Consider a balanced design with 200 cases and 200 controls and a risk factor with an exposure rate of 0.05, *i.e.*, $n_{\cdot 1} = n_{\cdot 2} = 200$ and $n_{11} = 10$ in Table 1. We calculated \hat{D}' and \hat{r}^2 , and their corresponding test statistics $\chi_{D'}^2$ and $\chi_{r^2}^2$, with $n_{11} = 10$ fixed and n_{12} decreasing from 10 to 0 (Table 2).

Since both the validity and power of a test depend on the coding schemes it employs in genetic studies (Sasieni 1997; Freidlin *et al.* 2002), we investigated what coding scheme to use when applying the D' test in the case of low-frequency variants. Consider a balanced case-control study with a total of N subjects and a diallelic locus with minor allele A and major allele a . Denote by x_i and y_i , where $i \in \{0, 1, 2\}$ indicates the copy number of allele A , the numbers of individuals with genotypes aa , Aa , and AA in cases and controls, respectively (Table 3). For simplicity, assume $x_2 = y_2 = 0$, *i.e.*, no individual with the AA genotype is observed, since the D' test is specifically designed for low-frequency var-

Table 3 Data distributions under different coding schemes for a low-frequency variant in a balanced case-control study

Genotype	Phenotype		Total
	Case	Control	
Distribution of co-dominant genotypes			
AA	$x_2 = 0$	$y_2 = 0$	0
Aa	x_1	y_1	n_{Aa}
aa	x_0	y_0	n_{aa}
Total	$N/2$	$N/2$	N
Distribution of alleles			
A	x_1	y_1	n_{Aa}
a	$2x_0 + x_1$	$2y_0 + y_1$	$2n_{aa} + n_{Aa}$
Total	N	N	$2N$
Distribution under a dominant model			
A ⁺	x_1	y_1	n_{Aa}
aa	x_0	y_0	n_{aa}
Total	$N/2$	$N/2$	N
Distribution under a recessive model			
AA	$x_2 = 0$	$y_2 = 0$	0
a ⁺	$x_0 + x_1$	$y_0 + y_1$	N
Total	$N/2$	$N/2$	N

A and a denote the minor and major alleles, respectively. For simplicity, we assume that no individual with the AA genotype is observed.

iants. As a result, we need only compare two coding schemes—one takes each allele as a unit, which we call the “allelic model”; the other takes each individual as a unit, which we call the “dominant model.”

We compared the power of the D' test with that of Pearson's χ^2 test and the likelihood-ratio test (LRT) to detect association of variants with various MAFs in a GWAS setting. To make a fair comparison, we simulated variants of different MAFs with equal chance of being detected by a standard method at a nominal level and compared the empirical power of the different methods in a genome-wide screen context. Specifically, for a given MAF f_0 in controls, the MAF f_1 in cases was computed, using the arcsine transformation, such that the power to detect the MAF difference between the cases and controls was 0.8 at a nominal 0.05 significance level (Cohen 1988). Three MAF categories—defined by the intervals $(0, 0.01)$, $[0.01, 0.10)$, and $[0.10, 0.50]$ —were considered. In each category 30 single-nucleotide polymorphisms (SNPs) associated with the disease were simulated by randomly choosing values of f_0 . To mimic a GWAS according to the MAF spectrum of the Affymetrix Genome-wide Human SNP Array 6.0 in Caucasians, we simulated 1 million null SNPs, among which 15, 20 and 65% were with MAFs randomly chosen from $(0, 0.01)$, $[0.01, 0.10)$, and $[0.10, 0.50]$, respectively. Again we considered a balanced design with an equal number (200, 500, 1000, or 2000) of cases and controls. In each setting, 100 replicates were generated. For each replicate, we recorded the statistics and performed the power analysis in two ways. One was to count the number of true positive SNPs among the top $m \in \{30, 40, \dots, 90, 100\}$ SNPs. The other way, adjusting for multiple testing, was to fit to the statistics an empirical null distribution, estimate the local false discovery

rate (FDR) (Efron 2004; Efron 2007a,b), and then count the number of true positive SNPs meeting a certain FDR criterion.

As a proof-of-principle example, we applied the D' test and the LRT in a GWAS of plasma levels of triglycerides with nonsynonymous SNPs in the African Americans (AAs) of the Dallas Heart Study (DHS) (Victor *et al.* 2004). There were 1722 individuals with complete phenotypes—triglycerides, body mass index (BMI), age, sex—after deleting those taking cholesterol-lowering medicine. We first fitted to data the linear regression model

$$\begin{aligned} \text{Triglycerides} = & \beta_0 + \beta_1 \times \text{Age} + \beta_2 \times \text{Sex} + \beta_3 \times \text{BMI} \\ & + \beta_4 \times \text{Ancestry} + \varepsilon, \end{aligned}$$

where the ancestry of each individual was inferred using ancestry-informative markers (Xing *et al.* 2011), to obtain the residuals ε . We then sampled the lowermost and uppermost quintiles of the residuals ($n = 345$) as cases and controls, respectively. There was a total of 8968 nonsynonymous SNPs assayed across the autosomes. We first filtered out singletons and those out of Hardy–Weinberg equilibrium (HWE; P -value $< 1.0 \times 10^{-5}$) in the whole AA population, and then filtered out monomorphic ones in the case-control sample, resulting in 8268 SNPs for further analysis. We performed genome-wide scans using both the D' test and the LRT. Fisher's exact test and Pearson's χ^2 test were omitted because the LRT is more powerful in detecting association for low-frequency variants (Xing *et al.* 2012, 2013), which was confirmed in the simulation study of this article.

Finally, we explored by simulation the asymptotic properties of the D' test, in particular whether the type I error rate is well calibrated for the D' test under the null hypothesis at different sample sizes. We considered a balanced design with equal numbers (200, 500, 1000, or 2000) of cases and controls. Variants with low frequencies were of particular interest and we simulated SNPs with MAFs equal to 0.005, 0.01, 0.05, and 0.1. In each setting 100,000 replicates, in each of which all the SNPs were polymorphic, were generated. Each data set could be summarized into a 2×2 table and we tested the allele frequency difference between cases and controls. The test statistic χ_D^2 was recorded to estimate its empirical distribution. We also performed Fisher's exact test, Pearson's χ^2 test, and the LRT, for comparison. Except in the case of Fisher's exact test, we calculated P -values assuming an asymptotic distribution of χ_1^2 . The empirical type I error rates at a nominal α significance level were calculated as the proportion of the 100,000 replicates for which the P -value was $\leq \alpha$.

Results

Compared with the r^2 -type test statistic, the D' -type test statistic under the alternative hypothesis is more distinct from that under the null hypothesis (Table 2). Under the null hypothesis of

no association— $n_{11} = n_{12} = 10$ —both \hat{D}' and \hat{r}^2 equal zero. As n_{12} decreases, \hat{D}' approaches one whereas \hat{r}^2 remains minimal, which is expected because the range of \hat{r}^2 is constrained by frequencies of both factors based on their definitions. In contrast, by definition χ_D^2 and $\chi_{r^2}^2$ should have the same range regardless of frequencies. Under the null hypothesis, both equal zero; however, as n_{12} decreases to 1 and 0, which correspond to situations of strong association, χ_D^2 increases more rapidly than $\chi_{r^2}^2$. This suggested the potential advantage of using the D' -type tests to detect association for low-frequency exposures, as further illustrated in a GWAS below.

Denote by \hat{D}'_a and \hat{D}'_d the sample estimates of D' , and by χ_a^2 and χ_d^2 the D' -type test statistics under the allelic and dominant models, respectively. It can be shown that $\hat{D}'_d = \hat{D}'_a$ and $\chi_d^2 > \chi_a^2$ in the case of a balanced case–control study for a low-frequency variant (see Appendix for details). Thus analyzing data under a dominant model leads to greater power than analyzing data under an allelic model for a low-frequency variant using the D' -type test, and therefore we employ the dominant coding scheme in what follows. It is of interest to note that by using the dominant model we also circumvent the potential issue of deviation from HWE of variants, which is a known issue impairing the validity of tests based on the allelic model coding scheme (Sasieni 1997).

In a genome-wide screen the D' test is superior to the LRT and Pearson's χ^2 test in detecting low-frequency variants, in particular when the sample size is small (Figure 1 and Figure 2); its gain in power diminishes as the sample size increases. Given the sample size, the D' test is most powerful in detecting association for low-frequency variants; its power decreases as the MAF increases. In contrast, the power of the other two tests increases as the MAF increases. The power of the three tests converges with increasing MAF and sample size. It is interesting to note that for common variants, when using ranking for the cutoff, the D' test selects fewer true positive variants than the other two methods (Figure 1); however, when using the FDR as the criterion to correct for multiple testing, the D' test selects more true positive variants, especially when the sample size is small (Figure 2). The underlying reason is that the null distribution and the alternative distribution of the D' test statistics are more distinguishable than those of the LRT statistics and Pearson's χ^2 test statistics (Table 4).

In Table 5 we list the union of the top 10 SNPs in the genome-wide scan of plasma triglyceride levels in the case–control study of DHS AAs using the D' test and the LRT. There was a total of 13 SNPs, among which 7 SNPs were ranked in the top 10 by both methods. All the top 10 SNPs by the LRT were ranked in the top 15 by the D' test; however, common variants tended to be ranked higher by the LRT than by the D' test. In contrast, there were two low-frequency variants (MAFs < 0.005) ranked in the top 10 by the D' test but ranked much lower by the LRT. There was one SNP with local FDR less than 0.05 using the D' test, but none using the LRT.

The D' test statistic converges to a distribution of χ_1^2 as both the MAF and the sample size increase; however,

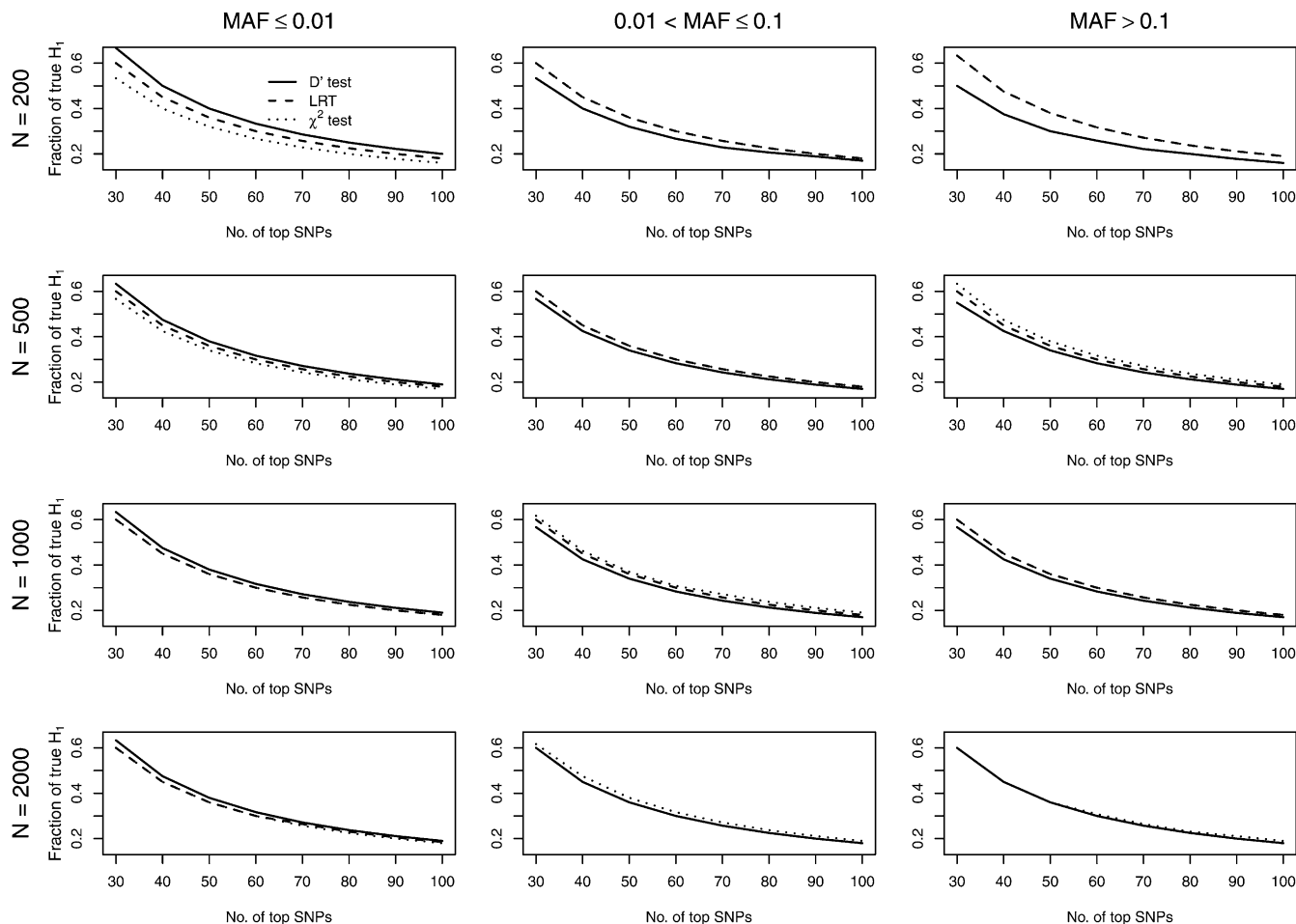


Figure 1 Power comparison of different methods using the ranking criterion. The x-axis is the number of top single-nucleotide polymorphisms (SNPs) in rank and the y-axis is the fraction of those SNPs associated with the disease given the ranking criterion. The plots are based on the mean of 100 replicates. Each row corresponds to a different sample size of N cases and controls, and each column corresponds to a minor allele frequency (MAF) category of SNPs.

when the MAF is low and the sample size is small, the distribution has a heavy tail (supporting information, Figure S1). Its type I error rate is well preserved at the nominal 0.05 significance level but inflated at more stringent test levels (Table 6). We previously compared Fisher's exact test, Pearson's χ^2 test, and the LRT, concluding that Fisher's exact test is the most conservative, whereas the LRT is the most liberal (Xing *et al.* 2012). Compared to the LRT, the D' test is more conservative at a loose criterion ($\alpha = 0.05$), but more liberal at a stringent criterion ($\alpha < 0.001$), especially in the case of testing for the association of low-frequency variants. This property suggests proper usage of the D' test: on one hand, the D' test is not a valid test at low test levels; on the other hand, the D' test statistic has the advantage of detecting association signals of low-frequency variants over that of common variants in a genome-wide screen setting, as we have demonstrated by both simulated and real data of genome-wide screens.

Discussion

Conventional wisdom regarding the LD measures D' and r^2 is that D' is more useful in population genetics, such as for

inferring recombination events, whereas r^2 is more useful in association mapping (Mueller 2004), because the power of a test between the causal variant and the proxy variant being examined is proportional to r^2 (Pritchard and Przeworski 2001; Chapman *et al.* 2003; Spencer *et al.* 2009; Lin *et al.* 2012). However, the r^2 -type association test lacks power for low-frequency variants. Instead, in this article we propose a D' -type test. Based on D' 's biological property of measuring recombination events, and by making use of its low sensitivity to allele frequencies and its upward bias estimate in the case of small sample sizes and low allele frequencies, we aimed to enhance the power of detecting association for low-frequency variants in genome-wide screens. In the D' -type test, the inflation of \hat{D}' under the null is largely controlled by the large accompanying variance (Zapata *et al.* 1997; Teare *et al.* 2002) and by adding 0.5 to the minimum cell while keeping the marginal totals fixed. The D' test is conservative at a loose criterion ($\alpha = 0.05$), but liberal at a stringent criterion ($\alpha < 0.001$), in particular when testing for association with low-frequency variants, which favors low-frequency variants with modest to large effect sizes in a genome-wide screen. As the sample size

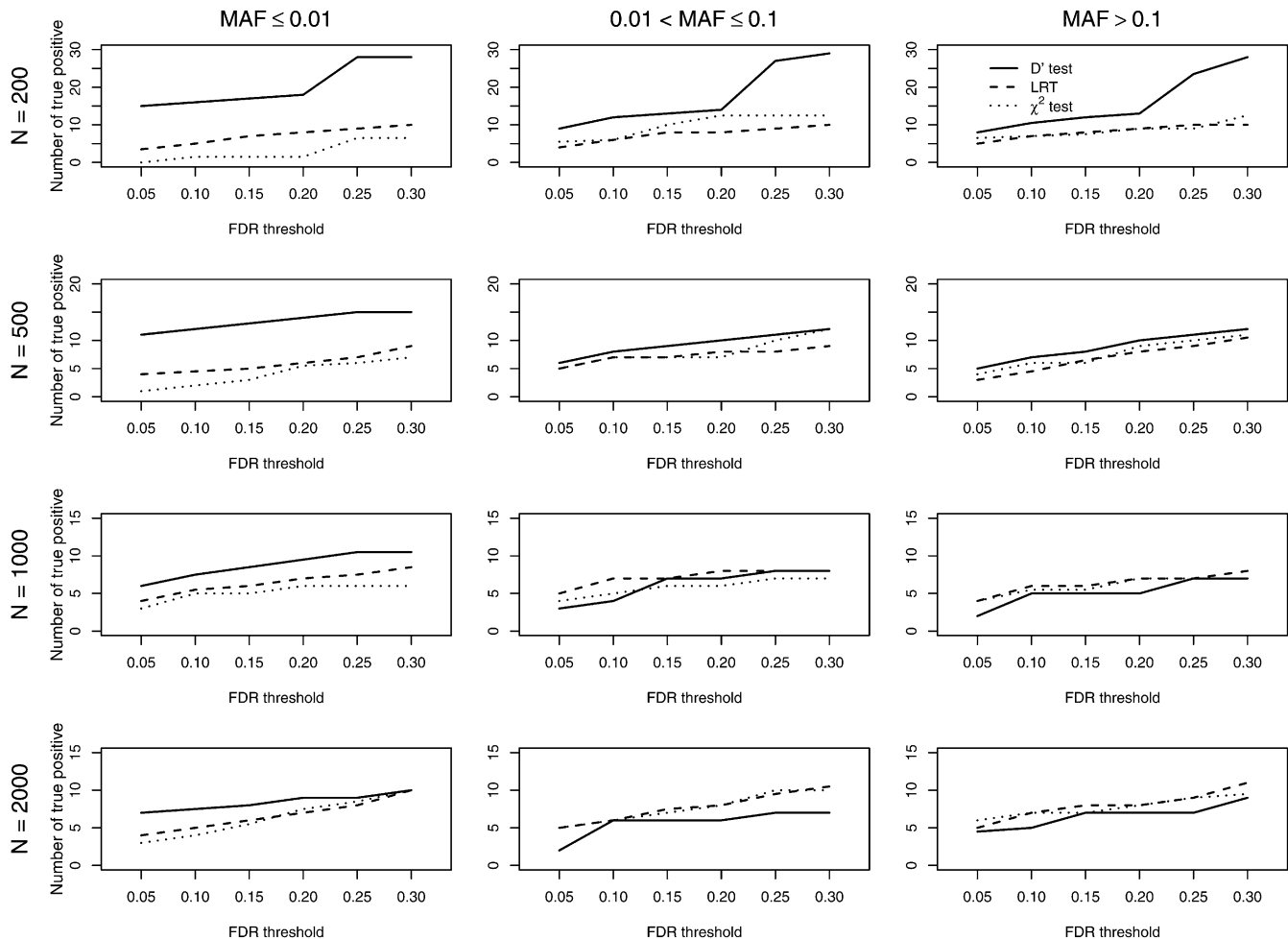


Figure 2 Power comparison of different methods using the false discovery rate (FDR) criterion. The x-axis is the local FDR and the y-axis is the number of SNPs associated with the disease given the FDR criterion. The plots are based on the mean of 100 replicates. Each row corresponds to a different sample size of N cases and controls, and each column corresponds to a MAF category of SNPs.

and the MAF increase, the test statistic converges to a χ_1^2 distribution under the null hypothesis. After standardization, \hat{D}' and \hat{r}^2 are comparable in terms of precision, accuracy, and efficiency, but the former has an advantage in the fact that its range of variation is always the same regardless of MAFs (Zapata 2011), which, we speculate, explains its superior power for low-frequency variants in genome-wide screens.

In the literature, a metric closely related to D' is the so-called “optimal measure of allelic association” ρ , which equals $|D'|$. It was proposed in a genetic mapping method based on the Malécot model for isolation by distance (Collins and Morton 1998; Morton *et al.* 2001; Morton *et al.* 2007). The Malécot model uses a composite likelihood to combine association information from multiple variants; however, at each site the association information is still summarized by Pearson’s χ^2 statistic, which is fundamentally different from the D' -type test we propose. The metric ρ originated from population genetic theories with an intuitive appeal—it is an estimate of the probability that a random haplotype has descended without recombination from

a founder population (Morton *et al.* 2001; Shete 2003). This feature sheds light on the D' -type test in detecting association for low-frequency variants; a highly penetrant, low-frequency, causal variant is expected to be in complete LD/association, *i.e.*, $D' = 1$, with disease affection status. Another metric closely related to D' is the robust attributable risk δ (Bengtsson and Thomson 1981; Devlin and Risch 1995). In the case of rare diseases and random sampling, such that the disease frequency is smaller than the MAF of a variant, these two metrics are approximately equal (Devlin and Risch 1995). In a case–control study design, the two metrics have the same numerator, which is D , but different denominators— $p_{\cdot 1}p_{22}$ and $p_{\cdot 1}p_{\cdot 2}$ for δ and D' , respectively. In a case–control design, $p_{\cdot 1} \approx p_{\cdot 2}$, but for a low-frequency variant $p_{1\cdot} < p_{22}$ so that $D' > \delta$. Therefore, D' should be a more sensitive metric for detecting association of low-frequency variants. Both ρ and δ were applied in a composite likelihood approach to locating a disease causal gene/variant assuming an evolutionary model (Terwilliger 1995; Devlin *et al.* 1996; Collins and Morton 1998; Morton *et al.* 2001, 2007). On one hand, we suspect that imposing a population

Table 4 Median of the test statistics for Pearson's χ^2 test, the likelihood-ratio test (LRT) and the D' test in genome-wide screens

Sample size	Test	Alternative SNPs ^a			Null SNPs ^b
		MAF < 0.01	0.01 ≤ MAF < 0.10	MAF ≥ 0.10	
200	D' test	19.24	12.77	12.00	0.35
	LRT	12.11	11.92	12.27	0.45
	χ^2 test	10.25	11.57	12.22	0.45
500	D' test	15.95	12.61	12.32	0.39
	LRT	12.04	12.34	12.54	0.45
	χ^2 test	10.98	12.15	12.51	0.45
1000	D' test	14.57	12.53	12.24	0.41
	LRT	12.36	12.47	12.42	0.46
	χ^2 test	11.66	12.37	12.41	0.46
2000	D' test	13.43	12.34	12.21	0.42
	LRT	12.35	12.40	12.35	0.46
	χ^2 test	11.95	12.34	12.34	0.46

In each replicate the medians for each category of SNPs were calculated and the mean of the medians from 100 replicates were reported.

^a There were 30 SNPs in each category.

^b There was a total of 1 million null SNPs, among which 15, 20, and 65% were with MAFs randomly chosen from (0, 0.01), [0.01, 0.10), and [0.10, 0.50], respectively.

genetic model may harm its power in gene mapping for complex diseases with locus and allelic heterogeneity; on the other hand, we speculate that using the information directly from ρ/D' , as the D' -type test, instead of from Pearson's χ^2 statistic at each site to form the composite likelihood, will increase its power, especially in isolated populations with relatively homogenous disease etiology.

The D' test is highly sensitive to association signals of low-frequency variants and so it can capture association missed by the r^2 -type test; however, to avoid spurious results, it is important to replicate findings in independent samples; to estimate significance accurately, a permutation procedure is needed. In the case of low-frequency variants, we proved that the D' test is more powerful in a dominant coding scheme than in an allelic coding scheme; however, this does not hold over the full allele frequency spectrum. Thus we advocate using the D' test in a dominant model as a complementary approach to the r^2 -type test instead of a replacement for it in GWAS. By doing so we also circumvent the potential issues brought up when a variant is out of HWE. Since the advantage of the D' test lies in low-frequency variants, it is not a primary interest of this article to construct a D' -type genotypic test for a 2×3 table, although this is under investigation. In this article we present a single-variant D' test to enhance the power of detecting association for low-frequency variants with moderate to large effect sizes. Previously, we compared the power of the single-variant LRT and two multivariate analysis approaches—the burden test and variance component method—for analyzing sequence data and showed that as long as there is a variant with a moderate to large effect size and stable expressivity, the single-variant LRT has the potential to identify it (Xing *et al.* 2012). Given that the D' test is more powerful than the LRT for low-frequency variants, this conclusion will also hold for the D' test. A further potential research topic would be to extend the D' test to be multi-variant, which, we expect, will enhance the power of detect-

ing loci of multiple low-frequency variants with small effect sizes.

Variants attaining genome-wide significance explain only a small fraction of the genetic variance of complex traits (Manolio *et al.* 2009), and it was shown that a large proportion of the heritability is not so much missing as hidden because the individual effects are too small to pass stringent significance tests (Yang *et al.* 2010); this is relevant to not only GWAS, but also sequencing studies. Apart from increasing the sample size, more efficient study designs and more powerful methods are needed to extract from random noise true associations, in particular, low-frequency variants with moderate to large effect sizes. For example, employing a weighted FDR control procedure we previously identified,

Table 5 Comparison of the D' test and the LRT for nonsynonymous variants in a genome-wide scan of plasma triglyceride levels in a case-control study of the Dallas Heart Study (DHS) African Americans

SNP ^a	MAF ^b	D' test		LRT	
		Rank ^c	FDR	Rank ^c	FDR
p233619	0.008	1	0.01	6	0.30
p4753017	0.121	2	0.12	1	0.10
p4781961	0.004	3	0.12	20	0.66
p710835	0.051	4	0.12	3	0.17
p970231	0.088	5	0.12	2	0.17
p3137745	0.195	6	0.13	4	0.17
p2508007	0.021	7	0.14	9	0.60
p4167740	0.129	8	0.19	5	0.30
p4776211	0.003	9	0.19	40	0.66
p4682552	0.025	10	0.30	11	0.60
p274255	0.061	11	0.41	7	0.60
p4563895	0.143	12	0.60	8	0.60
p4585346	0.330	13	0.61	10	0.60

The lowermost and uppermost quintiles of the DHS African American triglyceride levels were treated as cases and controls ($n = 345 \times 2$), respectively.

^a Perlegen identification.

^b MAF in the DHS African American population.

^c Out of 8268 variants in ascending order of FDR.

Table 6 Empirical type I error rates for Fisher's exact test, Pearson's χ^2 test, the LRT, and the D' test at different test levels

Sample size	MAF ^a	$\alpha = 0.05$				$\alpha = 0.001$			
		Exact test	χ^2 test	LRT	D' test	Exact test	χ^2 test	LRT	D' test
200	0.005	0.005	0.041	0.097	0.042	0.0000	0.0000	0.0003	0.0043
	0.01	0.017	0.044	0.075	0.043	0.0001	0.0001	0.0019	0.0086
	0.05	0.033	0.048	0.053	0.040	0.0006	0.0009	0.0012	0.0016
	0.1	0.038	0.052	0.053	0.041	0.0007	0.0010	0.0012	0.0011
500	0.005	0.021	0.042	0.070	0.041	0.0002	0.0002	0.0020	0.0072
	0.01	0.032	0.050	0.053	0.040	0.0004	0.0007	0.0014	0.0035
	0.05	0.039	0.049	0.050	0.041	0.0008	0.0010	0.0011	0.0010
	0.1	0.041	0.050	0.051	0.042	0.0008	0.0010	0.0010	0.0009
1000	0.005	0.029	0.051	0.053	0.041	0.0005	0.0007	0.0013	0.0031
	0.01	0.033	0.051	0.051	0.042	0.0005	0.0008	0.0011	0.0017
	0.05	0.042	0.050	0.051	0.044	0.0007	0.0009	0.0010	0.0011
	0.1	0.043	0.049	0.049	0.044	0.0008	0.0010	0.0010	0.0008
2000	0.005	0.035	0.049	0.053	0.042	0.0006	0.0009	0.0011	0.0019
	0.01	0.039	0.050	0.052	0.043	0.0007	0.0010	0.0011	0.0012
	0.05	0.044	0.050	0.050	0.045	0.0008	0.0009	0.0009	0.0009
	0.1	0.046	0.050	0.050	0.047	0.0008	0.0010	0.0010	0.0010

Based on 100,000 replicates, in each of which a variant was polymorphic and the number of cases/controls each equaled the indicated sample size.

^a MAF of a diallelic variant.

with a sample size of ~7400, low-frequency variants at *FOXA2* associated with plasma glucose levels (Xing *et al.* 2010), which was missed by the regular method with sample sizes of ~6000 (Banasik *et al.* 2012) and ~46,000 (Dupuis *et al.* 2010), but was confirmed later with an even larger sample size of >100,000 (Scott *et al.* 2012) and by accounting for interaction with BMI in a sample of >50,000 (Manning *et al.* 2012). In this study we proposed a test based on the standardized LD measure D' that is powerful in detecting association for low-frequency variants in genome-wide screens. We recommend using the D' test as a complementary approach to the conventional r^2 -type test to detect association for variants over the full allele frequency spectrum.

Acknowledgments

We thank the editor and reviewers for constructive comments that improved the manuscript, as well as Dr. Jonathan Cohen for critically reading the manuscript. We also thank Dr. Helen Hobbs for granting us permission to use the DHS data. This study was supported by the American Heart Association Scientist Development Grant (No. 10SDG4220051) to C.X., by the National Research Foundation of Korea Grant (No. NRF-2011-220-C00004) to R.C.E., and by the National Science Council of Taiwan Grant (No. NSC-102-2118-M-005-002) to C.-Y.L.

Literature Cited

Balding, D. J., 2006 A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7: 781–791.
 Banasik, K., M. Hollensted, E. Andersson, T. Sparso, A. Sandbaek *et al.*, 2012 The effect of *FOXA2* rs1209523 on glucose-related phenotypes and risk of type 2 diabetes in Danish individuals. *BMC Med. Genet.* 13: 10.

Bengtsson, B. O., and G. Thomson, 1981 Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* 18: 356–363.
 Casella, G., and R. L. Berger, 2001 *Statistical Inference*. Ed. 2. Duxbury, Pacific Grove, CA.
 Chapman, J. M., J. D. Cooper, J. A. Todd, and D. G. Clayton, 2003 Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum. Hered.* 56: 18–31.
 Cohen, J., 1988 *Statistical Power Analysis for The Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.
 Collins, A., and N. E. Morton, 1998 Mapping a disease locus by allelic association. *Proc. Natl. Acad. Sci. USA* 95: 1741–1745.
 Devlin, B., and N. Risch, 1995 A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29: 311–322.
 Devlin, B., N. Risch, and K. Roeder, 1996 Disequilibrium mapping: composite likelihood for pairwise disequilibrium. *Genomics* 36: 1–16.
 Dupuis, J., C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo *et al.*, 2010 New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* 42: 105–116.
 Edwards, A. W. F., 1963 The measure of association in a 2x2 table. *J. R. Stat. Soc. Ser. A* 126: 109–114.
 Efron, B., 2004 Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Stat. Assoc.* 99: 96–104.
 Efron, B., 2007a Correlation and large-scale simultaneous significance testing. *J. Am. Stat. Assoc.* 102: 93–103.
 Efron, B., 2007b Size, power and false discovery rates. *Ann. Stat.* 35: 1351–1377.
 Freidlin, B., G. Zheng, Z. Li, and J. L. Gastwirth, 2002 Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Hum. Hered.* 53: 146–152.
 Guo, S. W., 1997 Linkage disequilibrium measures for fine-scale mapping: a comparison. *Hum. Hered.* 47: 301–314.
 Hedrick, P. W., 1987 Gametic disequilibrium measures: proceed with caution. *Genetics* 117: 331–341.
 Hedrick, P. W., 1988 Inference of recombinational hotspots using gametic disequilibrium values. *Heredity* 60(3): 435–438.
 Hill, W. G., 1974 Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33: 229–239.

- Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38: 226–231.
- Lewontin, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49: 49–67.
- Lewontin, R. C., 1988 On measures of gametic disequilibrium. *Genetics* 120: 849–852.
- Lewontin, R. C., and K. Kojima, 1960 The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 458–472.
- Lin, C. Y., G. Xing, and C. Xing, 2012 Measuring linkage disequilibrium by the partial correlation coefficient. *Heredity* 109: 401–402.
- Mangin, B., P. Garnier-Gere, and C. Cierco-Ayrolles, 2008 The estimator of the optimal measure of allelic association: mean, variance and probability distribution when the sample size tends to infinity. *Stat. Appl. Genet. Mol. Biol.* 7: 20.
- Manning, A. K., M. F. Hivert, R. A. Scott, J. L. Grimsby, N. Bouatia-Naji *et al.*, 2012 A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycaemic traits and insulin resistance. *Nat. Genet.* 44: 659–669.
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorf *et al.*, 2009 Finding the missing heritability of complex diseases. *Nature* 461: 747–753.
- Morton, N. E., W. Zhang, P. Taillon-Miller, S. Ennis, P. Y. Kwok *et al.*, 2001 The optimal measure of allelic association. *Proc. Natl. Acad. Sci. USA* 98: 5217–5221.
- Morton, N., N. Maniatis, W. Zhang, S. Ennis, and A. Collins, 2007 Genome scanning by composite likelihood. *Am. J. Hum. Genet.* 80: 19–28.
- Mueller, J. C., 2004 Linkage disequilibrium for different scales and applications. *Brief. Bioinform.* 5: 355–364.
- Pe'er, I., Y. R. Chretien, P. I. de Bakker, J. C. Barrett, M. J. Daly *et al.*, 2006 Biases and reconciliation in estimates of linkage disequilibrium in the human genome. *Am. J. Hum. Genet.* 78: 588–603.
- Pritchard, J. K., and M. Przeworski, 2001 Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 69: 1–14.
- Sasieni, P. D., 1997 From genotypes to genes: doubling the sample size. *Biometrics* 53: 1253–1261.
- Scott, R. A., V. Lagou, R. P. Welch, E. Wheeler, M. E. Montasser *et al.*, 2012 Large-scale association analyses identify new loci influencing glycaemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* 44: 991–1005.
- Shete, S., 2003 A note on the optimal measure of allelic association. *Ann. Hum. Genet.* 67: 189–191.
- Slatkin, M., 2008 Linkage disequilibrium: understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 9: 477–485.
- Spencer, C. C., Z. Su, P. Donnelly, and J. Marchini, 2009 Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 5: e1000477.
- Teare, M. D., A. M. Dunning, F. Durocher, G. Rennart, and D. F. Easton, 2002 Sampling distribution of summary linkage disequilibrium measures. *Ann. Hum. Genet.* 66: 223–233.
- Terwilliger, J. D., 1995 A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am. J. Hum. Genet.* 56: 777–787.
- Victor, R. G., R. W. Haley, D. L. Willett, R. M. Peshock, P. C. Vaeth *et al.*, 2004 The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. *Am. J. Cardiol.* 93: 1473–1480.
- Xing, C., J. C. Cohen, and E. Boerwinkle, 2010 A weighted false discovery rate control procedure reveals alleles at FOXA2 that influence fasting glucose levels. *Am. J. Hum. Genet.* 86: 440–446.
- Xing, G., C. Y. Lin, and C. Xing, 2011 A comparison of approaches to control for confounding factors by regression models. *Hum. Hered.* 72: 194–205.
- Xing, G., C. Y. Lin, S. P. Wooding, and C. Xing, 2012 Blindly using Wald's test can miss rare disease-causal variants in case-control association studies. *Ann. Hum. Genet.* 76: 168–177.
- Xing, G., H. C. Ku, and C. Xing, 2013 A comparison of the likelihood ratio test and the variance-stabilising transformation-based tests for detecting association of rare variants. *Ann. Hum. Genet.* 77: 333–335.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42: 565–569.
- Zapata, C., 2011 On the uses and applications of the most commonly used measures of linkage disequilibrium from the comparative analysis of their statistical properties. *Hum. Hered.* 71: 186–195.
- Zapata, C., G. Alvarez, and C. Carollo, 1997 Approximate variance of the standardized measure of gametic disequilibrium D' . *Am. J. Hum. Genet.* 61: 771–774.
- Zapata, C., C. Carollo, and S. Rodriguez, 2001 Sampling variance and distribution of the D' measure of overall gametic disequilibrium between multiallelic loci. *Ann. Hum. Genet.* 65: 395–406.

Communicating editor: N. Yi

Appendix

Under the assumptions of a balanced case-control study with a total of N subjects and a diallelic locus with the minor allele A being rare such that no individual with the AA genotype is observed (Table 3), one can analytically compare the D' -type test statistics— χ_a^2 in the allelic model coding scheme and χ_d^2 in the dominant model coding scheme.

Denote by \hat{D}_d , $\hat{D}_{d \max}$, and \hat{D}'_d the sample estimates of D , D_{\max} , and D' , respectively, under the dominant model. By formulae (1), (3), and (4), it can be shown that

$$\hat{D}_d = \frac{x_1 y_0 - x_0 y_1}{N^2}, \quad \hat{D}_{d \max} = \frac{n_{Aa}}{2N}, \quad \hat{D}'_d = \frac{2(x_1 y_0 - x_0 y_1)}{N n_{Aa}},$$

and

$$\text{Var}(\hat{D}'_d) = \frac{(1 - |\hat{D}'_d|) \left[\frac{n_{Aa} n_{aa}}{4N^2} - \hat{D}_d^2 - |\hat{D}'_d| D_{d \max} (0.5 - 2|\hat{D}_d|) \right] + |\hat{D}'_d| \frac{n_{Aa} n_{aa}}{N^2}}{N \hat{D}_{d \max}^2}.$$

Denote by \hat{D}_a , $\hat{D}_{a \max}$, and \hat{D}'_a the sample estimates of D , D_{\max} , and D' , respectively, under the allelic model. Again by formulae (1), (3), and (4), it can be shown that

$$\hat{D}_a = \frac{x_1 y_0 - x_0 y_1}{2N^2}, \quad \hat{D}_{a \max} = \frac{n_{Aa}}{4N}, \quad \hat{D}'_a = \frac{2(x_1 y_0 - x_0 y_1)}{N n_{Aa}},$$

and

$$\text{Var}(\hat{D}'_a) = \left\{ (1 - |\hat{D}'_a|) \left[\frac{n_{Aa}(n_{Aa} + 2n_{aa})}{16N^2} - \hat{D}_a^2 - |\hat{D}'_a| D_{a \max} (0.5 - 2|\hat{D}_a|) \right] + |\hat{D}'_a| \frac{n_{Aa}(n_{Aa} + 2n_{aa})}{4N^2} \right\} / 2N \hat{D}_{a \max}^2.$$

Note that $\hat{D}_d = 2\hat{D}_a$, $\hat{D}_{d \max} = 2\hat{D}_{a \max}$, and $\hat{D}'_d = \hat{D}'_a$, i.e., the two coding schemes lead to the same D' estimates. By manipulation, it can be shown that

$$\text{Var}(\hat{D}'_a) = \text{Var}(\hat{D}'_d) + \left[(1 - |\hat{D}'_d|) \left(\frac{n_{Aa}^2}{8N^2} - \frac{\hat{D}_d^2}{2} \right) + |\hat{D}'_d| \frac{n_{Aa}^2}{2N} \right] / N \hat{D}_{d \max}^2.$$

Based on $\hat{D}_{d \max} = n_{Aa}/2N$, we know $n_{Aa}/N - 2|\hat{D}_d| \geq 0$, and therefore $(n_{Aa}/N + 2|\hat{D}_d|)(n_{Aa}/N - 2|\hat{D}_d|) \geq 0$, i.e., $n_{Aa}^2/N^2 - 4\hat{D}_d^2 \geq 0$. Based on this inequality, it follows that $\text{Var}(\hat{D}'_a) > \text{Var}(\hat{D}'_d)$. Given $\chi_d^2 = \hat{D}'_d^2 / \text{Var}(\hat{D}'_d)$, $\chi_a^2 = \hat{D}'_a^2 / \text{Var}(\hat{D}'_a)$, and $\hat{D}'_d = \hat{D}'_a$, we can conclude $\chi_d^2 > \chi_a^2$.

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.160739/-/DC1>

Enhancing the Power to Detect Low-Frequency Variants in Genome-Wide Screens

Chang-Yun Lin, Guan Xing, Hung-Chih Ku, Robert C. Elston, and Chao Xing

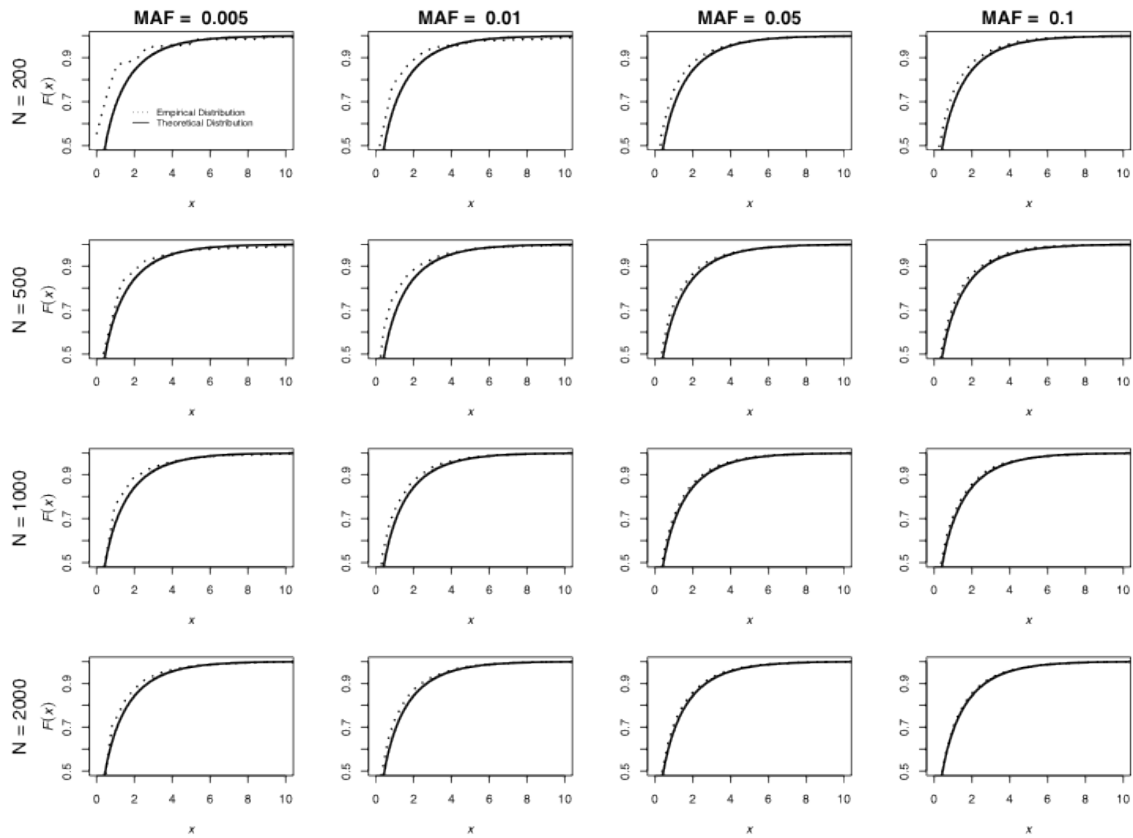


Figure S1 Empirical distribution of the D' test statistic

The x-axis is the test statistic and the y-axis is the cumulative distribution function. The dotted line is the empirical distribution of the D' test statistic based on 100,000 replicates, and the solid line is the theoretical distribution of χ_1^2 . Each row corresponds to a different sample size of N cases and controls, and each column corresponds to a minor allele frequency (MAF) of a diallelic variant.