Genetic Epidemiology 30: 301–313 (2006)

Improving the Power of Association Tests for Quantitative Traits in Family Studies

G. Diao and D.Y. Lin*

Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina

Association mapping based on family studies can identify genes that influence complex human traits while providing protection against population stratification. Because no gene is likely to have a very large effect on a complex trait, most family studies have limited power. Among the commonly used family-based tests of association for quantitative traits, the quantitative transmission-disequilibrium tests (QTDT) based on the variance-components model is the most flexible and most powerful. This method assumes that the trait values are normally distributed. Departures from normality can inflate the type I error and reduce the power. Although the family-based association tests (FBAT) and pedigree disequilibrium tests (PDT) do not require normal traits, nonnormality can also result in loss of power. In many cases, approximate normality can be achieved by transforming the trait values. However, the true transformation is unknown, and incorrect transformations may compromise the type I error and power. We propose a novel class of association tests for arbitrarily distributed quantitative traits by allowing the true transformation function to be completely unspecified and empirically estimated from the data. Extensive simulation studies showed that the new methods provide accurate control of the type I error and can be substantially more powerful than the existing methods. We applied the new methods to the Collaborative Study on the Genetics of Alcoholism and discovered significant association of single nucleotide polymorphisms (SNP) tsc0022400 on chromosome 7 with the quantitative electrophysiological phenotype TTTH1, which was not detected by any existing methods. We have implemented the new methods in a freely available computer program. *Genet. Epidemiol*. 30:301–313, 2006. © 2006 Wiley-Liss, Inc.

Key words: association studies; variance components; linkage disequilibrium; transformation; semiparametric models; QTDT; FBAT; PDT

Contract grant sponsor: NIH; Contract grant number: ROI CA 82859.
*Correspondence to: Dr. Danyu Lin, Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, CB #7420, Chapel Hill, North Carolina 27599–7420. E-mail: lin@bios.unc.edu
Received 9 November 2005; Accepted 11 January 2006
Published online 7 April 2006 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20145

INTRODUCTION

Complex human diseases are likely influenced by multiple genetic and environmental factors, with no particular gene having a singly large effect. It is widely believed that the genetic dissection of complex diseases requires association studies to explore the correlations between genetic variants, particularly single nucleotide polymorphisms (SNPs), and disease phenotypes [Risch, 2000; Botstein and Risch, 2003]. In fact, the availability of dense SNP maps across the human genome [International SNP Map Working Group, 2001; International HapMap Consortium, 2005] has led to a proliferation of SNP-based association studies worldwide.

The case-control study is a widely adopted strategy for association mapping. This study and other population-based studies are relatively easy to conduct but are prone to detecting spurious association arising from population stratification. Detection of the spurious association can be avoided by employing family-based association studies. Because it is difficult to collect a large number of families, most family studies have limited power to detect the moderate effects likely to be contributed by disease genes. Thus, it is critically important to employ the most powerful statistical methods for family-based association studies.

The transmission-disequilibrium test (TDT) [Spielman et al., 1993] and various extensions

[Martin et al., 1997, 2000; Boehnke and Langefeld, 1998; Horvath and Laird, 1998; Spielman and Ewens, 1998] can be used to detect association for dichotomous traits. The likelihood-based methods can improve the power over the TDT [Schaid and Sommer, 1993, 1994]. For many complex diseases, quantitative phenotypes are more informative than diagnostic categories in genetic analysis. There have been tremendous recent interests in the development of association tests for quantitative traits and in the use of these methods for mapping complex diseases.

The commonly used family-based tests of association for quantitative traits fall into two broad categories. The first category is based on regression models. Allison [1997] introduced a test for parents-offspring trios by using parental genotypes to construct family-matched controls in linear regression models. Allison et al. [1999] developed a test for siblings by using a linear regression model with random sibship effects. Fulker et al. [1999] proposed a variance-components method for the combined analysis of linkage and association for sib pairs. Their method involves modeling of the allelic means for testing association, with simultaneous modeling of the sib-pair covariance structure for testing linkage, and controls for spurious association due to population stratification by partitioning of the mean effect of a locus into between- and within-sibship components. Abecasis et al. [2000a,b] generalized this method to nuclear families and extended pedigrees and developed a program called quantitative transmission-disequilibrium tests (QTDT), which is widely used. Kistner and Weinberg [2004, 2005] proposed a polytomous logistic regression model by modeling the offspring genotype conditioning on the quantitative trait and the parents' genotypes.

The second category compares more directly the transmissions among offspring with high trait values to those of offspring with low trait values [Rabinowitz, 1997; Lunetta et al., 2000; Monks and Kaplan, 2000; Rabinowitz and Laird, 2000]. Laird et al. [2000] represented the test statistic as the covariance between genotype transmissions and trait values and developed the popular FBAT program. Monks and Kaplan [2000] proposed the pedigree disequilibrium tests (PDT) to allow missing parents and tests of association in the presence of linkage. Lange et al. [2002] extended the FBAT to incorporate familial phenotypic correlations by using the variance-components model of Fulker et al. [1999].

The QTDT is particularly attractive because it accommodates arbitrary pedigrees with or without parental genotypes and allows simultaneous analysis of linkage and association. The performance of the QTDT relies heavily on the normality assumption of the quantitative traits. Nonnormality can cause inflated type I error and diminished power. The FBAT and PDT are valid tests for arbitrarily distributed quantitative traits; however, these tests may not be powerful for nonnormal traits.

One approach to achieving (approximate) normality is to transform the trait values. However, it is very challenging to identify an appropriate transformation, especially when the data contain outlying trait values. Different transformations may yield conflicting results of analysis, and incorrect transformations can adversely affect the type I error and power.

In this article, we develop powerful tests of association for arbitrarily distributed quantitative traits. Specifically, we extend the QTDT, FBAT, and PDT by allowing a completely unspecified transformation function for the trait values, which is estimated empirically from the observed data. We also develop a procedure to properly adjust for multiple comparisons when testing several markers. We implement the new methods in a computer program for public use. Extensive simulation studies demonstrate that the new methods can be substantially more powerful than the existing ones while providing accurate control of the type I error; see Tables I, III, IV, V, VI and Figure 2. When applied to the Collaborative Study on the Genetics of Alcoholism (COGA), the new QTDT method detected significant association of SNP tsc0022400 on chromosome 7 with the quantitative electrophysiological phenotype TTTH1, whereas the existing methods did not; see Figure 3.

METHODS

NOTATION

Suppose that the study collects n general pedigrees or families, with n_i individuals in the ith pedigree. Let Y_{ij} be the trait value for the jth individual of the ith pedigree, and $\mathbf{x}_{ij}$ a vector of observed covariates. Consider a candidate diallelic marker, with frequencies p and $q = 1-p$ for alleles A and B. Define the marker genotype score for the jth individual of the ith pedigree as

$Z_{ij} = -1$, 0, or 1 according to whether this individual has genotype B/B, A/B, or A/A, respectively. Clearly, the number of transmitted A alleles is $Z_{ij}+1$.

MODELS

Assuming additive genetic effects, we propose the following variance-components model:

$$H(Y_{ij}) = \beta Z_{ij} + \gamma^{\mathrm{T}}\mathbf{x}_{ij} + g_{ij} + G_{ij} + e_{ij} \qquad (1)$$

where H is an unknown increasing function, β is the additive genetic effect, γ is a set of fixed covariate effects, g_{ij} is a random effect due to the major gene after accounting for the marker association, G_{ij} is a random effect due to other genes at unlinked loci, and e_{ij} is an individual-specific residual environmental effect. In this model, association and covariate effects are represented by the mean parameters, while linkage is represented by the covariance structure. The random effects g_{ij}, G_{ij}, and e_{ij} are assumed to be normally distributed with mean zero and variances σ_g^2, σ_G^2, and σ_e^2. Because H is an arbitrary function, we constraint the residual variance σ_e^2 to be 1 and absorb the intercept in H.

Suppose that g_{ij}, G_{ij}, and e_{ij} are uncorrelated. Then the expected phenotypic variance for $H(Y_{ij})$ is $\sigma^2 = (\sigma_a^2 + \sigma_g^2) + \sigma_G^2 + \sigma_e^2$, where $\sigma_a^2 = 2pq\beta^2$ is the phenotypic variance explained by the association with the candidate marker, and $\sigma_a^2 + \sigma_g^2$ is the overall additive genetic variance explained by both linkage and association. The overall heritability of the trait is $(\sigma_a^2 + \sigma_g^2 + \sigma_G^2)/\sigma^2$ and the heritability attributable to the examined locus is $(\sigma_a^2 + \sigma_g^2)/\sigma^2$.

Let $\mathbf{H}_i$ denote the transformed trait values $[H(Y_{i1}), \ldots, H(Y_{in_i})]^{\mathrm{T}}$ for the ith family. The phenotypic covariance matrix of $\mathbf{H}_i$, after accounting for association, can be expressed as

$$\mathbf{V}_i = \sigma_g^2\Sigma_{gi} + 2\sigma_G^2\Sigma_{Gi} + \sigma_e^2\mathbf{I}_i$$

where $\mathbf{\Sigma}_{gi}$ contains the proportions of alleles at the major locus that are identity-by-descent (IBD) among the relative pairs in the ith family, $\mathbf{\Sigma}_{Gi}$ is the matrix of kinship coefficients which depends only on the relatedness of the relative pairs, and $\mathbf{I}_i$ is an identity matrix. Several computer programs, such as GENEHUNTER [Kruglyak et al., 1996], SOLAR [Almasy and Blangero, 1998], and MERLIN [Abecasis et al., 2002], are available for estimating the IBD allele sharing probabilities.

To avoid detecting spurious association introduced by population stratification, we follow Fulker et al. [1999] and Abecasis et al. [2000a,b] to decompose the marker genotype score Z_{ij} into orthogonal between- and within-family components: b_{ij} denotes the expected genotype score conditional on family data, and w_{ij} denotes the deviation from this expectation. Let M_{ij} and F_{ij} represent specific indexes for the male and female parents of the jth individual in the ith family. In nuclear families, b_{ij} is defined as $(Z_{F_{ij}} + Z_{M_{ij}})/2$ if parental genotypes are available and as the average of the Z_{ij} among the siblings of the ith family otherwise. In general pedigrees, we set $b_{ij} = Z_{ij}$ for genotyped founders; for nonfounders, b_{ij} is $(b_{F_{ij}} + b_{M_{ij}})/2$ if both $b_{F_{ij}}$ and $b_{M_{ij}}$ are defined and is the average genotype score among the full siblings of the jth individual in the ith pedigree otherwise.

Given the above orthogonal decomposition of the genotype scores, we modify model (1) as

$$\begin{aligned} H(Y_{ij}) =& \beta_{\mathrm{b}} b_{ij} + \beta_{\mathrm{w}} w_{ij} + \gamma^{\mathrm{T}}\mathbf{x}_{ij} \\ &+ g_{ij} + G_{ij} + e_{ij} \end{aligned} \qquad (2)$$

where β_{b} and β_{w} are the between- and within-family effects. Positive values of β_{w} imply that excess transmission of A allele has a positive effect on the trait values. By the arguments of Abecasis et al. [2000a], β_{b} accounts for all the spurious association between genotype score and phenotype, and β_{w} provides a direct measure of the additive genetic value. We refer to (1) and (2) as semiparametric variance-components models because the function H is unspecified. The existing variance-components models are parametric in that the transformation is assumed to be known or incorporated into the definition of Y.

MAXIMUM LIKELIHOOD ESTIMATION

Define $\Lambda(y) = e^{H(y)}$. Let ξ denote the variance parameters σ_g^2 and σ_G^2, and let $\boldsymbol{\theta}$ denote the complete set of parameters β_{b}, β_{w}, γ, ξ, and Λ. The log likelihood for $\boldsymbol{\theta}$ takes the form

$$\begin{aligned} & c - \frac{1}{2}\sum_{i=1}^{n} \log|\det(\mathbf{V}_i)| \\ & - \frac{1}{2}\sum_{i=1}^{n} (\mathbf{H}_i - \beta_{\mathrm{b}}\mathbf{b}_i - \beta_{\mathrm{w}}\mathbf{w}_i - \mathbf{X}_i\gamma)^{\mathrm{T}} \\ & \times \mathbf{V}_i^{-1}(\mathbf{H}_i - \beta_{\mathrm{b}}\mathbf{b}_i - \beta_{\mathrm{w}}\mathbf{w}_i - \mathbf{X}_i\gamma) \\ & + \sum_{i=1}^{n}\sum_{j=1}^{n_i} \log \frac{\lambda(\mathrm{Y_{ij}})}{\Lambda(\mathrm{Y_{ij}})} \end{aligned} \qquad (3)$$

where c is a constant, $\mathbf{X}_i$ is the matrix of covariates for the ith family, $\mathbf{b}_i = (b_{i1}, \ldots, b_{in_i})^{\mathrm{T}}$, $\mathbf{w}_i = (w_{i1}, \ldots, w_{in_i})^{\mathrm{T}}$, and λ is the derivative of Λ. This is a nonparametric likelihood [Bickel et al., 1993] in that the function H or Λ is completely arbitrary.

It seems natural to estimate $\boldsymbol{\theta}$ by maximizing the nonparametric log likelihood given in (3). The maximum does not exist if Λ is restricted to be absolutely continuous. Thus, we regard Λ as a right-continuous function and maximize the function

$$\begin{aligned}\log L(\boldsymbol{\theta}) = c &- \frac{1}{2}\sum_{i=1}^{n}\log|\det(\mathbf{V}_i)| \\ &- \frac{1}{2}\sum_{i=1}^{n}(\mathbf{H}_i - \beta_{\mathrm{b}}\mathbf{b}_i - \beta_{\mathrm{w}}\mathbf{w}_i - \mathbf{X}_i\gamma)^{\mathrm{T}} \\ &\times \mathbf{V}_i^{-1}(\mathbf{H}_i - \beta_{\mathrm{b}}\mathbf{b}_i - \beta_{\mathrm{w}}\mathbf{w}_i - \mathbf{X}_i\gamma) \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{n_i}\log\frac{\Lambda\{Y_{ij}\}}{\Lambda(Y_{ij})}\end{aligned} \tag{4}$$

where $\Lambda\{Y_{ij}\}$ is the jump size of $\Lambda(y)$ at $y = Y_{ij}$, i.e., the value of $\Lambda(y)$ at $y = Y_{ij}$ minus its value right before Y_{ij}. The resulting estimator, denoted by $\widehat{\boldsymbol{\theta}} = (\hat{\beta}_{\mathrm{b}}, \hat{\beta}_{\mathrm{w}}, \widehat{\gamma}, \widehat{\xi}, \widehat{\Lambda})$, is the nonparametric maximum likelihood estimator of $\boldsymbol{\theta}$ [Bickel et al., 1993].

We can show that $\widehat{\Lambda}(\cdot)$ is a step function with jumps at Y_{ij} only. Thus, we maximize (4) over β_{b}, β_{w}, γ, ξ, and $\Lambda\{Y_{ij}\}$ $(i = 1, \ldots, n;\ j = 1, \ldots, n_i)$ through the quasi-Newton algorithm [Press et al., 1992]. The unknown transformation $H(y)$ is then estimated by $\widehat{H}(y) = \log\widehat{\Lambda}(y)$.

The (nonparametric) maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ has many desirable properties. First, the estimators of the regression and variance parameters depend on the Y_{ij} only through their ranks, so the estimators are rank-based and thus insensitive to outliers. Second, $\hat{\boldsymbol{\theta}}$ is approximately unbiased, normally distributed, and statistically efficient, implying that the unknown transformation is correctly estimated from the data and the likelihood-based test statistics are the most powerful among all valid test statistics. We can prove the forgoing results by adopting the arguments in Appendix B of Diao and Lin [2005b]; the actual proofs are available from the authors.

TEST STATISTICS

We can perform various hypothesis tests under model (2). Specifically, we can assess whether there is association between the candidate marker and quantitative trait by testing the null hypothesis H_0: $\beta_{\mathrm{w}} = 0$ against the alternative H_{A}: $\beta_{\mathrm{w}} \neq 0$; we will refer to this test as the semiparametric QTDT (SQTDT). We can also test for the presence of population stratification, H_0: $\beta_{\mathrm{b}} = \beta_{\mathrm{w}}$. In addition, we can test H_0: $\sigma_g^2 = 0$ against H_A: $\sigma_g^2 > 0$ for genetic linkage. For each hypothesis test, we can calculate the likelihood ratio statistic

$$\mathrm{LR} = -2[\log L(\widetilde{\boldsymbol{\theta}}) - \log L(\widehat{\boldsymbol{\theta}})]$$

where $\widetilde{\boldsymbol{\theta}}$ is the restricted maximum likelihood estimator of $\boldsymbol{\theta}$ under the null hypothesis. For testing association, LR is approximately χ_1^2 distributed. For testing linkage, the distribution of LR is approximated by a mixture of χ^2 distributions [Self and Liang, 1987].

We can use model (2) to develop a semiparametric version of the general FBAT [Laird et al., 2000; Lange et al., 2002]. Specifically, we obtain the restricted maximum likelihood estimator $\widetilde{\boldsymbol{\theta}}$ under model (2) with the constraint of $\beta_{\mathrm{w}} = 0$, and calculate the residuals $\mathbf{R}_i = \widetilde{\mathbf{H}}_i - \tilde{\beta}_{\mathrm{b}}\mathbf{b}_i - \mathbf{X}_i\tilde{\gamma}$. We then define

$$S = \sum_{i=1}^{n} S_i$$

where $S_i = \mathbf{w}_i^{\mathrm{T}}\widetilde{\mathbf{V}}_i^{-1}\mathbf{R}_i$, and $\widetilde{\mathbf{V}_i}$ is the estimated phenotypic covariance matrix for the ith family. By treating the marker genotype scores Z_{ij} as random and the trait values or covariate-adjusted residuals as fixed, we propose the following semiparametric FBAT:

$$\mathrm{SFBAT} = \frac{S^2}{\mathrm{var}(S)}$$

where

$$\mathrm{var}(S) = \sum_{i=1}^{n}(\widetilde{\mathbf{V}}_i^{-1}\mathbf{R}_i)^{\mathrm{T}}\,\mathrm{Cov}(\mathbf{Z}_i)\widetilde{\mathbf{V}}_i^{-1}\mathbf{R}_i$$

and $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{in_i})^{\mathrm{T}}$.

One can calculate the conditional mean and covariance matrix of $\mathbf{Z}_i$ under the null hypothesis of no association, regardless of whether the parental genotypes are available or not [Rabinowitz and Laird, 2000]. Note that the between-family component b_{ij} is the conditional expectation of the marker genotype score Z_{ij}. As in the case of the general FBAT, we can incorporate an offset vector $\boldsymbol{\mu}_i = (\mu_1, \ldots, \mu_{n_i})$ for $\mathbf{R}_i$ into the calculation of S_i.

There are two key differences between the general FBAT and the proposed SFBAT. First, the transformation function H is assumed to be known in the FBAT, but is completely unspecified and nonparametrically estimated in the SFBAT.

Second, we use the correlation-adjusted residuals $\widetilde{\mathbf{V}}_i^{-1}\mathbf{R}_i$ instead of the $\mathbf{R}_i$ in the construction of the SFBAT so as to incorporate familial correlation due to the linkage of the major gene locus and the polygenic effects at unlinked loci. Lange et al. [2002] extended the general FBAT by using a simple structure for $\mathbf{V}_i$ to account for the environmental correlation within families and showed that disregarding the within-family correlation results in loss of power.

Motivated by the fact that the covariance between the marker residuals w_{ij} and the phenotypic residuals R_{ij} is zero in the absence of association, we propose the semiparametric PDT

$$\text{SPDT} = \frac{S^2}{\widetilde{\text{var}}(S)}$$

where

$$\widetilde{\text{var}}(S) = \sum_{i=1}^{n} \widetilde{S}_i^2$$

$\widetilde{S}_i = S_i + \mathbf{\Gamma}^{\mathrm{T}}\Sigma^{-1}\mathbf{U}_i$, $\mathbf{U}_i$ is the ith family's score function for $\boldsymbol{\eta} = (\beta_b, \boldsymbol{\gamma}, \boldsymbol{\xi}, \Lambda)$, $\mathbf{\Sigma}$ is the Fisher information matrix of $\boldsymbol{\eta}$, and

$$\mathbf{\Gamma} = \sum_{i=1}^{n} \frac{\partial \mathbf{w}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{R}_i}{\partial \boldsymbol{\eta}}.$$

The unknown parameter $\boldsymbol{\theta}$ in $\mathbf{\Gamma}$, $\mathbf{\Sigma}$ and $\mathbf{U}_i$ is evaluated at $\widetilde{\boldsymbol{\theta}}$. The PDT [Monks and Kaplan, 2000] assumes that the transformation function H is known and ignores the within-family correlation. The denominator of the SPDT takes a more complicated form than its counterpart in the PDT because the S_i are not independent.

Under the null hypothesis of no association, both the SFBAT and SPDT are approximately χ_1^2 distributed. Both tests are valid, at least in large samples, even when model (2) is incorrect. Note that the two tests differ only in the variance calculation: the SFBAT computes the variances of the marker scores on the basis of Mendelian transmissions, whereas the SPDT estimates the variance of S empirically.

Because the QTDT of Abecasis et al. [2000a,b] is a likelihood ratio statistic under a parametric variance-components model, its performance depends critically on the normality assumption. The FBAT and PDT can be represented as the score statistics under parametric variance-components models and thus may not have good power for nonnormal traits, although they are valid (in large samples).

Technically speaking, the normality assumption is imposed on the unobserved residuals or the conditional distribution of the trait values (given the genetic marker and covariates) rather than on the marginal trait distribution. In fact, the marginal distribution will not be normal if there is any marker or covariate effect. Thus, the common practice of making the distribution of the trait values normal-looking through transformation can be counter-productive.

ADJUSTMENTS FOR MULTIPLE TESTING

In association studies, one often examines a number of SNPs in a chromosomal region. Failure to account for the effects of multiple comparisons would result in an abundance of false positive results. The commonly adopted Bonferroni correction is overly conservative because the test statistics for SNPs in linkage disequilibrium (LD) are correlated. We describe below a Monte Carlo procedure to properly adjust for multiple testing.

Suppose that one is interested in testing association with m SNPs. At the kth SNP site, $k = 1, \ldots, m$, the test statistic can be written as or be approximated by $T_k = U_k^2/V_k$, where

$$U_k = \sum_{i=1}^{n} U_{ki}$$

U_{ki} involves only the data from the ith family, and

$$V_k = \sum_{i=1}^{n} U_{ki}^2.$$

For the SFBAT and SPDT, U_{ki} takes the form of $\widetilde{S}_i$. For the SQTDT, U_{ki} pertains to the ith family's efficient score function [Bickel et al., 1993; Diao and Lin, 2005b] for β_w, which has a similar expression to that of $\widetilde{S}_i$. If none of the m SNPs is associated with the trait, then the joint distribution of $(U_1, \ldots, U_m)$ is approximately multivariate normal with mean zero and with covariance

$$V_{kl} = \sum_{i=1}^{n} U_{ki} U_{li}$$

between U_k and U_l. Define

$$\widetilde{U}_k = \sum_{i=1}^{n} U_{ki} \mathcal{G}_i$$

and $\widetilde{T}_k = \widetilde{U}_k^2/V_k$, where $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent standard normal random variables. The joint

distribution of $(T_1, \ldots, T_m)$ is approximately the same as the conditional distribution of $(\widetilde{T}_1, \ldots, \widetilde{T}_m)$ given the data. One can obtain realizations from the latter by generating a large number of normal random samples $(\mathcal{G}_1, \ldots, \mathcal{G}_n)$. Given these realizations, one can obtain the P-values for the test statistics adjusted for multiple comparisons [Lin, 2005].

RESULTS

SIMULATION STUDIES

We carried out a number of simulation studies to investigate the properties of the new methods and to compare them with those of the existing methods. We assumed a diallelic QTL, Q, with additive effects and simulated a tightly linked diallelic marker locus, M, with recombination fraction of 0. The frequencies of the minor alleles Q_1 and M_1 of Q and M are 0.25, i.e., $p_{Q_1} = p_{M_1} = 0.25$. We introduced LD between the QTL and marker locus in the parental chromosomes. LD is measured by $D = p_{M_1Q_1} - p_{M_1}p_{Q_1}$, where $p_{M_1Q_1}$ is the frequency of haplotype M_1Q_1. The maximum of D is $D_{\max} = \min(p_{M_1}, p_{Q_1}) - p_{M_1}p_{Q_1}$, and the standardized LD coefficient is $D' = D/D_{\max}$. We considered different levels of D', the case of $D' = 0$ pertaining to the null hypothesis of no association. For each scenario, we simulated 10,000 data sets, each with 100 nuclear families. Each family consisted of 2, 3, 4, or 5 siblings with probabilities 0.3, 0.3, 0.2, and 0.2, respectively. The parental genotypes were assumed to be known.

In the first set of studies, we generated trait values from the model

$$H(Y_{ij}) = \beta Z_{ij} + \gamma_1 X_{1ij} + \gamma_2 X_{2ij} + G_{ij} + e_{ij} \quad (5)$$

where $H(y) = \log(2y-2)$, $\beta = 0.73$, $\gamma_1 = -1$, $\gamma_2 = 1$, Z_{ij} is the QTL genotype score, X_{1ij} is a binary variable with 0.5 probability of being 1, X_{2ij} is an independent standard normal variable, and G_{ij} and e_{ij} are independent zero-mean normal variables with variances $\sigma_G^2 = 0.6$ and $\sigma_e^2 = 1.2$. The overall heritability is 0.4, and the major-gene heritability is 0.1.

We evaluated the proposed semiparametric methods (i.e., SQTDT, SFBAT, and SPDT), as well as the existing parametric methods (i.e., QTDT, FBAT, and PDT) with various transformations, including the true transformation, log transformation, square-root transformation, and no transformation. We also included the permutation test of Abecasis et al. [2000a]. The parametric method with the true transformation is an ideal situation in which the normality assumption holds after a known transformation.

The results of these studies are presented in Table I. The new methods provide accurate control of the type I error in all cases and have virtually the same power as their parametric counterparts with the true transformation. As expected, the SQTDT tends to be more powerful than the SFBAT and SPDT. Although the QTDT with an incorrect transformation has reasonable type I error, the power is drastically reduced. Without transformation, the power is extremely low. In the case of complete LD (i.e., $D' = 1$), the power of the SQTDT is 95.0% at the 1% nominal significance level, as compared to 14.7% for the QTDT without transformation. The permutation test does not improve the power. With incorrect transformations, the FBAT and PDT tend to be conservative and have low power.

The additive effect of the marker on the phenotype is $\alpha = \beta D/p_{M_1}p_{M_2}$ [Cardon and Abecasis, 2000]. The results for the estimation of this parameter are summarized in Table II. The parameter estimator is virtually unbiased. The standard error estimator reflects accurately the true variation, and the confidence intervals have proper coverage probabilities. As expected, the effect size of the marker decreases as the LD between the QTL and marker alleles becomes weaker.

The residual genetic variance, i.e., the difference between the additive genetic variance of the QTL and the variance of the QTL explained by association with the marker allele is given by $\sigma_g^2 = 2p_{Q_1}p_{Q_2}\beta^2 - 2p_{M_1}p_{M_2}\alpha^2$ [Cardon and Abecasis, 2000]. In the above studies, the type I error/power of the proposed method to detect residual linkage are approximately 0.19, 0.18, 0.14, 0.09, and 0.05 at the nominal significance level of 0.05 when D' equals 0, 0.25, 0.5, 0.75, and 1, respectively.

In the second set of studies, we considered the same model as the first one except that the genetic and environmental factors are correlated. Specifically, we generated X_{2ij} from a normal distribution with mean $|Z_{ij}|$ and variance 1. The results are shown in Table III. The new methods continue to perform as well as their parametric counterparts with the true transformation and greatly outperform the parametric methods with incorrect transformations. The type I error of the QTDT with incorrect transformations is inflated. The permutation test does not correct the type I error

TABLE I. Type I error and power (%) of the association tests at the nominal significance level of 1% for nonnormal traits when the genetic and environmental factors are uncorrelated

Method	Transformation	$D' = 0.00$	0.25	0.50	0.75	1.00
SQTDT	Unspecified	0.93	5.80	29.13	69.87	95.03
QTDT	True	1.03	6.16	30.41	71.14	95.41
	None	1.01	1.46	3.40	7.40	14.69
	Square root	0.92	2.24	7.87	20.32	41.90
	Log	0.95	3.67	15.27	41.80	74.38
	Permutation	0.98	1.50	3.83	8.47	16.09
SPDT	Unspecified	0.89	4.93	24.69	63.60	93.19
PDT	True	0.86	4.96	24.98	64.13	93.41
	None	0.37	0.62	2.36	6.97	17.17
	Square root	0.56	1.67	7.24	21.01	46.34
	Log	0.78	3.16	13.72	39.51	73.83
SFBAT	Unspecified	0.63	4.31	23.85	63.84	93.27
FBAT	True	0.63	4.24	24.27	64.20	93.53
	None	0.34	0.64	2.54	7.42	17.81
	Square root	0.50	1.64	7.22	21.71	47.36
	Log	0.53	2.79	13.71	40.03	74.82

Note: QTDT, PDT, and FBAT are the parametric tests due to Abecasis et al. [2000a], Monks and Kaplan [2000], and Laird et al. [2000]. SQTDT, SPDT, and SFBAT are the proposed semiparametric QTDT, PDT, and FBAT tests.

TABLE II. Summary statistics for the estimation of the marker effects

D'	True value	Mean	SE	SEE	CP(%)
0.00	0.000	−0.003	0.160	0.162	95.5
0.25	0.167	0.159	0.161	0.162	95.4
0.50	0.333	0.322	0.160	0.162	95.2
0.75	0.500	0.485	0.161	0.162	94.9
1.00	0.667	0.649	0.158	0.162	95.2

Note: The original value of α is divided by $\sigma_e = \sqrt{1.2}$ so that α and $\hat{\beta}_w$ are compared on the same scale. Mean and SE are the sampling mean and sampling standard error of the parameter estimator; SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95% confidence interval.

or improve the power. Without transformation, the FBAT and PTDT have inflated type I error; with incorrect transformations, the power is reduced.

In the third set of studies, we generated population admixture by mixing in equal proportions families drawn from two populations (A and B) with different QTL and marker allele frequencies: in population A, $p_{Q_1} = p_{M_1} = 0.25$; in population B, $p_{Q_1} = p_{M_1} = 0.75$. When there is no LD in either population, LD exists in the pooled population with $D' = 0.25$. We considered the same model as above except that the value of D' is population-specific rather than for the pooled population. The results, as shown in Table IV, are similar to those without population admixture. The new methods are robust to the spurious association introduced by population admixture and provide accurate control of the type I error.

For positive quantitative traits, one may employ the Box-Cox transformation. However, the Box-Cox transformation may compromise the type I error and power unless it provides a good approximation to the true transformation. In the above three sets of studies, the Box-Cox transformation turned out to be the log-transformation and thus failed to alleviate the adverse effects of nonnormality.

To investigate the robustness against outlying trait values, we considered model (5) with identity H but generated the residual error for 2% of the families from the exponential distribution with mean of 4. Figure 1 shows the distribution of trait values for the first simulated data set. The results are summarized in Table V. While still providing accurate control of the type I error, the new methods are much more powerful than the existing methods, especially under strong LD.

Instead of the familiar parametric transformations, one may use the rank or normal score transformation. For example, Wang and Huang [2002] described a multivariate normal copula model by using normal score transformation.

TABLE III. Type I error and power (%) of the association tests at the nominal significance level of 1% for nonnormal traits when the genetic and environmental factors are correlated

Method	Transformation	$D' = 0.00$	0.25	0.50	0.75	1.00
SQTDT	Unspecified	0.94	5.83	28.08	68.05	94.03
QTDT	True	1.07	5.85	29.00	69.00	94.57
	None	1.95	1.34	1.74	3.48	6.51
	Square root	1.30	1.78	5.65	15.67	32.40
	Log	1.26	3.72	16.57	45.33	78.14
	Pernutation	1.53	1.14	1.66	3.59	6.97
SPDT	Unspecified	0.74	4.52	22.38	59.43	91.02
PDT	True	0.69	4.35	22.56	59.97	91.03
	None	1.60	0.82	0.65	1.51	3.73
	Square root	0.73	0.68	3.40	11.64	29.40
	Log	0.69	2.50	11.92	37.70	73.16
SFBAT	Unspecified	0.59	3.72	21.15	59.63	91.26
FBAT	True	0.55	3.70	21.55	59.95	91.45
	None	2.11	1.23	1.25	2.86	6.24
	Square root	1.14	1.25	5.08	15.96	37.25
	Log	0.83	2.99	14.71	43.17	77.77

See the Note to Table I.

TABLE IV. Type I error and power (%) of the association tests at the nominal significance level of 1% for nonnormal traits in the presence of population admixture when the genetic and environmental factors are correlated

Method	Transformation	$D' = 0.00$	0.25	0.50	0.75	1.00
SQTDT	Unspecified	1.15	6.22	29.14	69.68	94.97
QTDT	True	1.29	6.57	30.27	70.43	95.31
	None	1.61	3.66	8.18	16.61	27.94
	Square root	1.71	6.39	18.50	40.88	66.01
	Log	1.75	8.65	30.81	65.79	90.93
	Pernutation	1.19	3.03	6.45	12.97	22.11
SPDT	Unspecified	0.97	5.18	24.60	63.69	92.76
PDT	True	0.96	5.26	24.76	63.81	93.06
	None	0.54	1.86	4.92	12.27	24.68
	Square root	0.95	4.21	14.32	35.88	62.72
	Log	1.25	6.52	25.05	58.05	87.25
SFBAT	Unspecified	0.81	4.65	23.96	62.95	92.95
FBAT	True	0.83	4.70	24.14	63.33	93.04
	None	0.99	3.09	7.44	17.33	31.89
	Square root	1.37	5.86	18.78	42.82	70.72
	Log	1.54	7.60	28.39	63.22	90.05

See the Note to Table I.

However, such transformations tend to destroy the relationship between the trait and the marker (or covariates) and thus reduce the power. To demonstrate this point, we considered the same model as (5) but with identity H and log-normal X_{2ij}. In this case, no transformation is needed, although the marginal distribution of the trait values is not normal. As shown in Table VI, the

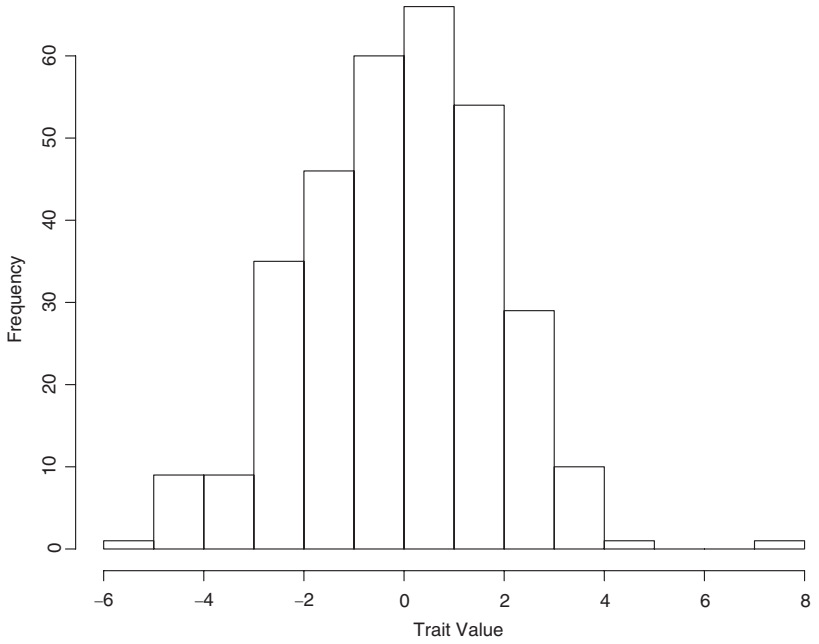

Fig. 1. Histogram of trait values for a simulated data set with outliers.

TABLE V. Type I error and power (%) of the association tests at the nominal significance level of 1% in the presence of outliers

Method	$D' = 0.00$	0.25	0.50	0.75	1.00
SQTDT	1.04	5.70	26.36	64.19	92.20
QTDT	1.02	5.03	22.08	55.07	85.05
SPDT	0.91	4.59	22.76	58.18	89.11
PDT	0.66	3.09	14.68	39.33	66.31
SFBAT	0.77	4.04	21.51	57.36	88.98
FBAT	0.70	3.52	18.06	48.86	79.71

See the Note to Table I.

rank and normal score transformations result in appreciable loss of power, as compared to the new methods and the existing methods without transformation.

Our final set of simulation studies was designed to assess the performance of the proposed Monte Carlo procedure for multiple testing. We generated 11 tightly linked, evenly distributed SNPs in a chromosome region, with successive recombination fractions of 0.02. The minor allele frequency for each SNP was set to 0.3. We assumed that the QTL is located in the middle, which was either in linkage equilibrium (i.e., $D' = 0$) or in complete LD with the sixth SNP. We generated the trait values from model (5) with $(\sigma_a^2, \sigma_G^2, \sigma_e^2) = (0.1, 0.7, 1.2)$. We considered different degrees of LD between successive SNPs. The results for the SQTDT are shown in Figure 2. Under the null hypothesis of no association, the type I error based on the Monte Carlo procedure is always close to the nominal level whereas the Bonferroni correction is overly conservative. Under the alternative hypothesis, the Monte Carlo procedure is more powerful than the Bonferroni correction, especially under strong LD. For the pairwise LD coefficient D' of 0.8, the power associated with the Bonferroni correction is 62.6% at the nominal significance level of 5% whereas that of the Monte Carlo procedure is 69.0%. Similar results were obtained for the SFBAT, SPDT, and the parametric methods (data not shown).

COGA STUDY

COGA is a multi-center study designed to identify and characterize genes that affect susceptibility to alcohol dependence and related

TABLE VI. Type I error and power (%) of the association tests with rank and normal score transformation at the nomial significance level of 1% for normal traits

Method	Transformation	$D' = 0.00$	0.25	0.50	0.75	1.00
SQTDT	Unspecified	0.94	5.75	28.83	69.34	94.89
QTDT	None	1.00	6.07	30.52	70.97	95.43
	Rank	1.11	5.04	20.94	54.30	85.06
	Normal score	1.07	5.60	25.00	61.97	90.28
SPDT	Unspecified	0.88	4.88	24.42	63.08	92.99
PDT	None	0.91	4.83	24.95	63.92	93.46
	Rank	0.96	3.89	17.26	48.23	80.70
	Normal score	0.81	4.37	19.76	54.32	86.05
SFBAT	Unspecified	0.68	4.13	23.36	63.10	93.09
FBAT	None	0.65	4.28	24.25	64.11	93.51
	Rank	0.73	3.72	16.77	47.78	81.56
	Normal score	0.72	3.92	19.38	54.27	86.46

See the Note to Table I.

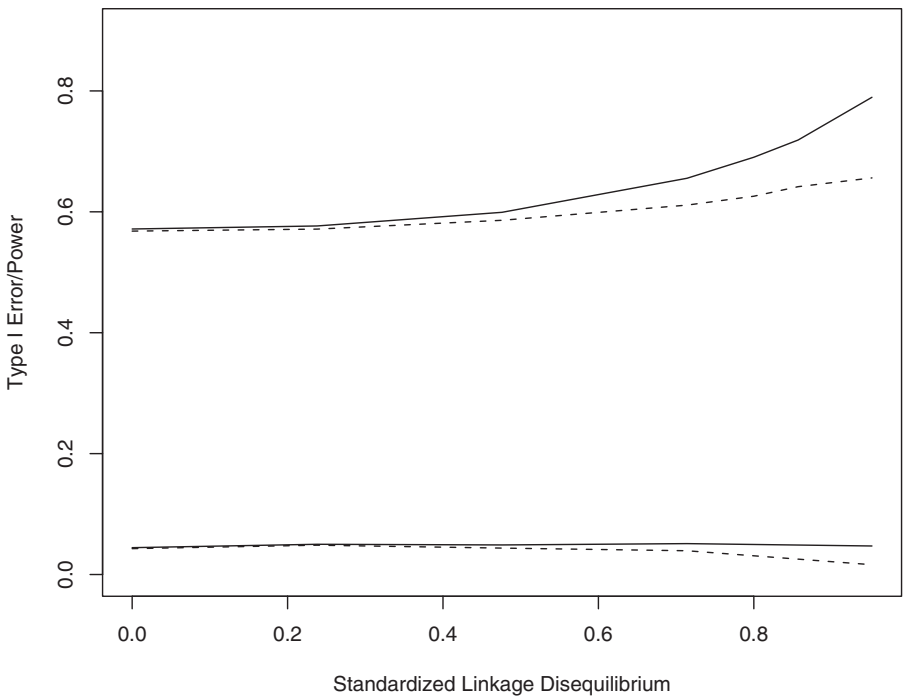

Fig. 2. Type I error and power of the semiparametric QTDT at the overall nominal significance level of 0.05 based on the Bonferroni correction and the proposed Monte Carlo method. The horizontal axis pertains to the linkage disequilibrium between two successive loci. The lower and upper solid curves pertain to the type I error and power of the Monte Carlo method; the lower and upper dashed curves pertain to the type I error and power of the Bonferroni correction.

phenotypes [Begleiter et al., 1995]. The study contains 143 multi-generation pedigrees with a total of 1,614 individuals and with family sizes ranging from 5 to 32. A total of 1,353 individuals were selected for genotyping of SNPs conducted by Affymetrix and Illumina. We considered the quantitative electrophysiological phenotype TTTH1 (electric potential FP1, far frontal left side channel). Strong linkage signals of TTTH1 on chromosome 7 were previously discovered by

Porjesz et al. [2002]. Of the 1,353 genotyped individuals, 901 had the TTTH1 measurements. The mean TTTH1 value was 2.41, with SD of 0.68 and median of 2.32. The distribution of the TTTH1 values was slightly right skewed with skewness of 0.62 and kurtosis of 0.28.

We performed a chromosome-wide association analysis (on chromosome 7) using the genotype data provided by Affymetrix for 577 SNPs. Of the total 578 Affymetrix SNPs on chromosome 7, SNP tsc0047552 contained the same genotype for all individuals and was thus ignored in the analysis. We included age, age^2, gender, and maximum number of drinks consumed in a 24-hr period as covariates in model (2). We estimated the IBD allele-sharing probabilities at each SNP site by the computer package SOLAR [Almasy and Blangero, 1998].

Figure 3 displays the LOD scores of four association tests: SQTDT, QTDT, SPDT, and PDT. (The SFBAT and FBAT require nuclear families and thus were excluded.) The LOD scores were obtained by dividing the original statistics by $2\log 10$. The LOD scores reached their peaks at the same location of 130.405 cM for SNP tsc0022400, with peak values of 3.27 and 2.73 for SQTDT and SPDT, as opposed to 3.07 and 2.35 for QTDT and PDT. The corresponding *P*-values without adjustment of multiple testing were 1.04×10^{-4}, 3.99×10^{-4}, 1.73×10^{-4}, and 1.0×10^{-3}. With the Bonferroni correction, no SNPs were found to be significantly associated with the trait at the chromosome-wide significance level of 0.05, the Bonferroni-adjusted *P*-values of the SQTDT, QTDT, SPDT, and PDT being 0.06, 0.10, 0.23, and 0.58, respectively, for SNP tsc0022400. Using the proposed Monte Carlo procedure to account for multiple testing, we obtained the adjusted p-values of 0.05, 0.08, 0.17, and 0.33 for the SQTDT, QTDT, SPDT, and PDT, respectively, at SNP tsc0022400. Thus, the SQTDT with the Monte Carlo adjustment of multiple testing is the only test that is significant at the chromosome-wide significance level of 0.05. The proposed estimate of the effect size for SNP tsc0022400 is -0.525 with estimated standard error of 0.135.

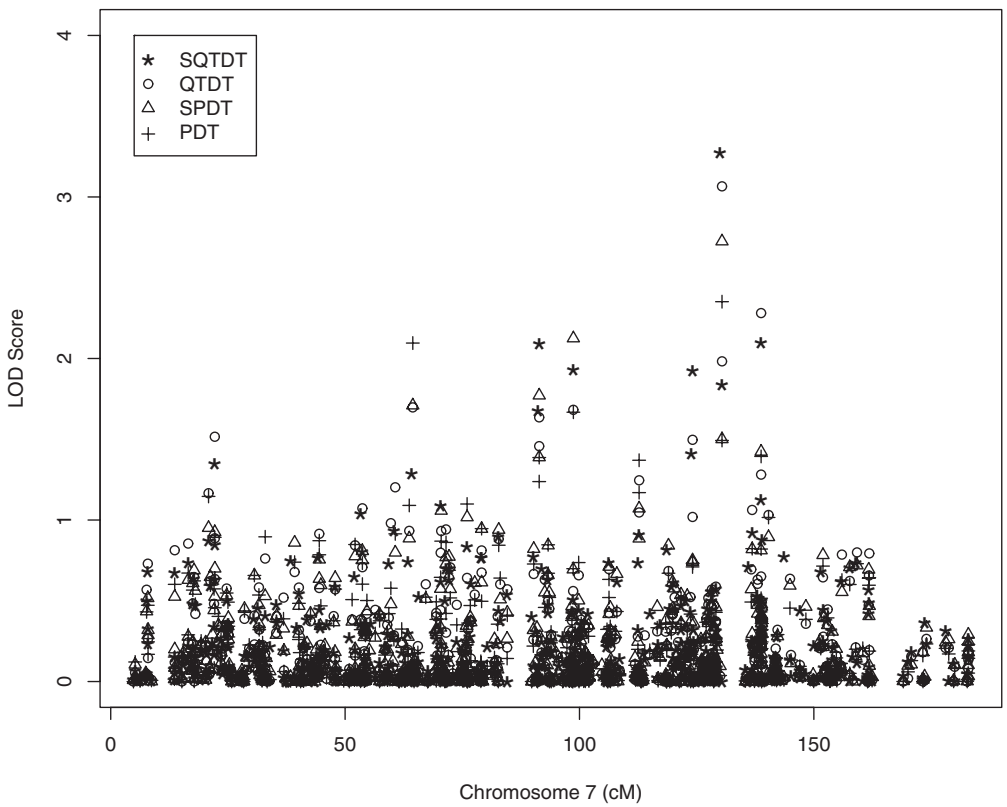

Fig. 3. LOD scores from the new and existing association tests for quantitative trait TTTH1 on chromosome 7 in the COGA study: QTDT and PDT are the parametric tests due to Abecasis et al. [2000a] and Monks and Kaplan [2000]; SQTDT and SPDT are the proposed semiparametric QTDT and PDT tests.

DISCUSSION

Family-based tests of association for quantitative traits are playing an increasingly important role in identifying genetic determinants of complex diseases. The performance of the existing methods depends critically on the normality assumption. Most quantitative traits are not normally distributed. The semiparametric methods described in this article are considerably more powerful than the existing methods under nonnormality and have the same power as the existing methods under normality.

The biological rationale for variance-components models was provided by Amos [1994], Fulker et al. [1999] and Abecasis et al. [2000a,b] among others. Our main contribution is to allow an unknown transformation of the trait values. Misspecification of other aspects of the model may also affect the performance of the current methods. The validity of the SFBAT and SPDT does not depend on correct model specification, although their power does.

We have implemented the new methods in a cost-free computer program, which is posted on our website site: www.bios.unc.edu/∼lin. Although it is more time consuming to perform a semiparametric test than a parametric test, the computing time is not a concern with the current computing power. It took less than 6 s on an IBM BladeCenter HS-20 machine to perform the SQTDT and SPDT tests at one SNP locus for the COGA data. For the simulation studies, the analysis at one position took only 1 s for the three semiparametric tests. Our computer program is efficient and reliable even for very large samples. We conducted some simulation studies with 500 sibships, and it took less than 15 sec to perform the semiparametric tests; the results are similar to those reported in the Simulation Studies section.

Recently, Diao and Lin [2005a] extended the traditional variance-components model [Amos, 1994] by allowing an arbitrary transformation and developed a powerful and robust method for linkage analysis. Model (2) is more general than the model of Diao and Lin [2005a] in that it formulates both linkage and association. If the association effects are disregarded, then model (2) reduces to model (1) of Diao and Lin [2005a].

The SQTDT is versatile in that it can handle extended pedigrees with missing genotype data and perform joint linkage and association analysis. Model (2) allows one to test for population stratification, as described in the Methods section. If no population stratification is detected, one can use the more powerful test based on model (1). For the SFBAT and SPDT, one can flexibly choose the offset vector and the structure of $\mathbf{V}_i$. Lange et al. [2002] showed that the FBAT and PDT are valid tests and are more powerful than the QTDT when only offsprings with trait values in the upper 10% tail of the trait distribution are selected. Similar conclusions hold for the SFBAT, SPDT, and SQTDT (data not shown).

For simplicity of description, we assumed that the markers are diallelic and the genetic effects are additive. It is straightforward to incorporate multiallelic markers and dominant effects. In addition, we can extend model (2) to include gene-gene interactions and gene-environment interactions. It is also possible to extend the single-trait model (2) to longitudinal data and multiple traits.

In some studies, the trait values may be censored. When the trait pertains to the age at onset of a disease, censoring is inevitable because of loss to follow-up and limited study duration. Censoring also arises if the assay cannot detect values smaller (or larger) than some threshold. Lange et al. [2004] considered the FBAT approach for the age-at-onset data based on the log-rank and Wilcoxon statistics. We are currently extending the SQTDT, SFBAT, and SPDT to censored traits.

ACKNOWLEDGMENTS

The authors are grateful to the COGA investigators and Jean W. MacCluer for providing the COGA data from GAW14, which was supported in part by the NIH grant GM31575.

REFERENCES

Abecasis GR, Cardon LR, Cookson WOC. 2000a. A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292.

Abecasis GR, Cookson WOC, Cardon LR. 2000b. Pedigree tests of transmission disequilibrium. Eur J Hum Genet 8:545–551.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101.

Allison DB. 1997. Transmission-disequilibrium tests for quantitative traits. Am J Hum Genet 60:676–690.

Allison DB, Heo M, Kaplan N, Martin ER. 1999. Sibling-based tests of linkage and association for quantitative traits. Am J Hum Genet 64:1754–1764.

Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211.

Amos CI. 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54: 535–543.

Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, Schuckit MA, Edenberg HJ, Rice JP. 1995. The collaborative study on the genetics of alcoholism. Alcohol Health Res World 19:228–236.

Bickel PJ, Klassen CAJ, Ritov Y, Wellner JA. 1993. Efficient and adaptive estimation in semiparametric models. Baltimore: Johns Hopkins University Press.

Boehnke M, Langefeld CD. 1998. Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet 62:950–961.

Botstein D, Risch N. 2003. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat Genet (Suppl. 33): 228–237.

Cardon LR, Abecasis GR. 2000. Some properties of a variance components model for fine-mapping quantitative trait loci. Behav Genet 30:235–243.

Diao G, Lin DY. 2005a. A powerful and robust method for mapping quantitative trait loci in general pedigrees. Am J Hum Genet 77:97–111.

Diao G, Lin DY. 2005b. Semiparametric methods for mapping quantitative trait loci with censored data. Biometrics 61:789–798.

Fulker DW, Cherny SS, Sham PC, Hewitt JK. 1999. Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 64:259–267.

Horvath S, Laird NM. 1998. A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 63:1886–1897.

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

International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933.

Kistner EO, Weinberg CR. 2004. Method for using complete and incomplete trios to identify genes related to quantitative trait. Genet Epidemiol 27:33–42.

Kistner EO, Weinberg CR. 2005. A method for identifying genes related to a quantitative trait, incompleting multiple siblings and missing parents. Genet Epidemiol 29:155–165.

Kruglyak L, Daly M, Reeve-Daly M, Lander ES. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363.

Laird NM, Horvath S, Xu X. 2000. Implementing a unified approach to family-based tests of association. Genet Epidemiol (Suppl. 19):S36–S42.

Lange C, DeMeo DL, Laird NM. 2002. Power and design considerations for a general class of family-based association tests: quantitative traits. Am J Hum Genet 71: 1330–1341.

Lange C, Blacker D, Laird NM. 2004. Family-based association tests for survival and times-to-onset analysis. Stat Med 23: 179–189.

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

Lunetta KL, Faraone SV, Biederman J, Laird NM. 2000. Family-based tests of association and linkage that use unaffected sibs, covariates, and interaction. Am J Hum Genet 66:605–614.

Martin ER, Kaplan NL, Weir BS. 1997. Tests of linkage and association in nuclear families. Am J Hum Genet 61:439–448.

Martin ER, Monks SA, Warren LL, Kaplan NL. 2000. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146–154.

Monks SA, Kaplan NL. 2000. Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. Am J Hum Genet 66:576–592.

Porjesz B, Begleiter H, Wang K, Almasy L, Chorlian DB, Stimus AT, Kuperman S, O'Connor SJ, Rohrbaugh J, Bauer LO, Edenberg HJ, Goate A, Rice JP, Riech T. 2002. Linkage and linkage disequilibrium mapping of ERP and EEG phenotypes. Biol Psychol 61:229–248.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. Numerical Recipes in C: The Art of Scientific Computing, 2nd edition. New York: Cambridge University Press.

Rabinowitz D. 1997. A transmission disequilibrium test for quantitative trait loci. Hum Hered 47:342–350.

Rabinowitz D, Laird NM. 2000. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 50:211–223.

Risch N. 2000. Searching for genetic determinants in the new millennium. Nature 405:847–856.

Schaid DJ, Sommer SS. 1993. Genotype relative risks: methods for design and analysis of candidate-gene association studies. Am J Hum Genet 53:1114–1125.

Schaid DJ, Sommer SS. 1994. Comparison of statistics for candidate-gene associate studies using genes and parents. Am J Hum Genet 55:402–409.

Self SG, Liang KL. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J Am Stat Assoc 82:605–610.

Spielman RS, McGinnis RE, Ewens WJ. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506–516.

Spielman RS, Ewens WJ. 1998. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–458.

Wang K, Huang J. 2002. A score-statistic approach for the mapping of quantitative-trait loci with sibships of arbitrary size. Am J Hum Genet 70:412–424.