

High-Resolution Joint Linkage Disequilibrium and Linkage Mapping of Quantitative Trait Loci Based on Sibship Data

Ruzong Fan Jeusun Jung

Department of Statistics, Texas A&M University, College Station, Tex., USA

Key Words

QTL · LD mapping · Linkage analysis · IBD · Variance component models

Abstract

This paper proposes variance component models for high resolution joint linkage disequilibrium (LD) and linkage mapping of quantitative trait loci (QTL) based on sibship data; this can include population data if independent individuals are treated as single sibships. One application of these models is late onset complex disease gene mapping, when parental data are not available. The models simultaneously incorporate both LD and linkage information. The LD information is contained in mean coefficients of sibship data. The linkage information is contained in the variance-covariance matrices of trait values for sibships with at least two siblings. We derive formulas for calculating the probability of sharing two trait alleles identical by descent (IBD) for sibpairs in interval mapping of QTL; this is the coefficient of dominant variance of the trait covariance of sibpairs on major QTL. To investigate the performance of the formulas, we calculate the numerical values via the formulas and get satisfactory approximations. We compare the power and sample sizes for both LD and linkage mapping. By simu-

lation and theoretical analysis, we compare the results with those of Fulker and Abecasis 'AbAw' approach. It is well known that the resolution of linkage analysis can be low for complex disease gene mapping. LD mapping, on the other hand, can increase mapping precision and is useful in high resolution mapping. Linkage analysis is less sensitive to population subdivisions and admixtures. The level of LD is sensitive to population stratification which may easily lead to spurious association. Performing a joint analysis of LD and linkage mapping can help to overcome the limits of both approaches. Moreover, the advantages of the two complementary strategies can be utilized maximally. In practice, linkage analysis may be performed using pedigree data to identify suggestive linkage between markers and trait loci based on a sparse marker map. In the presence of linkage, joint LD and linkage mapping can be carried out to do fine gene mapping based on a dense genetic map using both pedigree and population data. Population and pedigree data of any type can be combined to perform a joint analysis of high resolution LD and linkage mapping of QTL by generalizing the method.

Copyright © 2003 S. Karger AG, Basel

KARGERFax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com© 2003 S. Karger AG, Basel
0001-5652/03/0564-0166\$19.50/0Accessible online at:
www.karger.com/hheDr. Ruzong Fan
Department of Statistics
Texas A&M University
TAMUS, College Station, TX 77843 (USA)
Tel. +1 979 845 3156 or 3141, Fax +1 979 845 3144, E-Mail rfan@stat.tamu.edu

Introduction

There has been growing interest in joint analyses of linkage disequilibrium (LD) mapping (or association studies) and linkage studies of genetic traits. Separate analyses of either LD mapping or linkage analysis utilize only part of the available information; LD mapping uses LD information, and linkage analysis uses linkage information. A unified analysis simultaneously utilizes both LD and linkage information, and hence, is potentially more powerful. For qualitative traits, several studies have shown that joint LD and linkage mapping is advantageous over separate analyses [Göring and Terwillinger, 2000; Xiong and Jin, 2000]. Almasy et al. [1999] propose variance component models in quantitative trait locus (QTL) detection using combined linkage/disequilibrium analysis. Fulker et al. [1999] also utilize variance component models to perform combined linkage and LD mapping based on sibpairs for quantitative traits. Sham et al. [2000] perform analytical analyses of power of linkage versus association mapping of quantitative traits for sibship data. Abecasis et al. [2000, 2001] generalize the method of Fulker et al. [1999] to analyze data of nuclear families and general pedigrees. For natural populations, Wu et al. [2002] utilize mixture models in joint linkage and LD mapping of QTL. In these studies, the investigators usually use only one marker in their analyses. Since dense marker maps such as single nucleotide polymorphisms (SNPs) and high resolution microsatellite markers are available [The International SNP Map Work Group, 2001; Broman et al., 1998; Kong et al., 2002], it is natural and practical to generalize a single marker joint LD and linkage mapping to high resolution multiple marker mapping. Fan and Xiong [2002] propose a linear regression model using two markers in high resolution LD mapping of QTL based on population data. Fan and Xiong [2003] propose a variance component model for combined high resolution linkage and LD mapping of QTL using two flanking markers based on both population and nuclear family data. To our knowledge, Fan and Xiong [2002, 2003] are the only papers in literature using two markers in high resolution joint LD and linkage mapping of QTL.

For late-onset disorders such as Alzheimer's disease, heart disease, many forms of cancer, non-insulin-dependent diabetes mellitus (NIDDM), and osteoporosis, it is difficult to collect parental data. One way to study late-onset disorders is to perform sibpair or sibship analyses [Cardon 2000; Horvath and Laird, 1998; Schaid and Li, 1997; Schaid and Rowland, 1998; Spielman and Ewens, 1998]. This motivates us to explore models in high resolu-

tion joint LD and linkage mapping of QTL based on sibship data. We can include population data by treating an independent individual as a single sibship.

Linkage analysis is a family-based method, which uses recent recombination events in pedigrees to estimate distances among genetic loci. Here genetic loci can be either genetic markers or trait loci. If two genetic loci are very close, e.g., ≤ 2.5 cM, the number of recombination events between the two loci in a pedigree is unlikely to be high. For this reason, linkage analysis is usually suitable for broad chromosome region mapping (≤ 10 cM), but not suitable for high resolution mapping (≤ 2.5 cM). LD mapping, on the other hand, is based on both population data and pedigree data; it uses historical recombination events between genetic loci since non-random association of alleles at genetic loci was introduced into a population. Thus, LD can operate over short map distances, and can increase mapping precision in high resolution mapping. Linkage analysis is less sensitive to the population structures of subdivisions and admixtures, although its resolution can be low. The level of LD is sensitive to population stratification, although it can increase resolution in dissecting genetic traits when the association between markers and trait loci is introduced by events such as mutations at trait loci. Performing a joint analysis of LD and linkage mapping can help to overcome the limitations of the individual methods. Moreover, the advantages of the two complementary strategies can be utilized maximally. In practice, linkage analysis can be performed using pedigree data to identify suggestive linkage between markers and trait loci based on a sparse marker map. Then, joint LD and linkage mapping can be carried out to do fine gene mapping based on a dense genetic map using both pedigree and population data in the presence of linkage. In this paper, we propose variance component models to perform joint LD and linkage mapping of QTL based on sibship data. We justify the validity of the method theoretically; the method can also be generalized to use any type of pedigree data.

Following Fan and Xiong [2002, 2003], we introduce a linear regression model to describe a quantitative trait. Using variance component models, we perform model building and analytical analysis of the proposed methods based on sibship data. In the model, linkage information is contained in the variance-covariance structure, and LD information is contained in the mean coefficients. Hence, we simultaneously perform joint LD and linkage interval mapping of QTL using two flanking markers. The interval mapping studies published to date are mainly limited to use only the additive genetic variance in the analysis.

There are no explicit formulas to include both additive and dominant genetic variances in the interval mappings. In this paper, we derive formulas to calculate the trait covariance of sibships using both additive and dominant variances. To investigate the performance of the formulas, we calculate the numerical values via the formulas and get satisfactory approximations. To compare the power and sample sizes for both LD and linkage mapping, we calculate the noncentrality parameters of sibpairs and general sibships. We adopt an idea from Sham et al. [2000] to calculate the noncentrality parameters for linkage analysis based on standard statistical theory [Stuart and Ord, 1991]. For LD mapping, we use general theory of linear model to calculate the noncentrality parameters [Graybill, 1976]. Then we compare the power and sample sizes of LD mapping, and the power of linkage mapping using two flanking markers with or without dominant variance. By simulation and theoretical analysis, we compare the results with those of Fulker and Abecasis 'AbAw' approach. The method is applied to the Genetic Analysis Workshop (Gaw) 12 German asthma data [Myers, Wjst and Aber, 2001].

Methods

Models

Consider a quantitative trait which is influenced by a quantitative trait locus Q . Assume that there are two alleles Q_1 and Q_2 at the trait locus with frequencies q_1 and q_2 . Suppose that trait locus Q is flanked by two markers A and B in an order of AQB . At the marker locus A , assume there are two alleles A and a with frequencies P_A and P_a , respectively; for the marker B , assume that there are two alleles B and b with frequencies P_B and P_b . Suppose that trait locus Q and markers A and B are individually in Hardy-Weinberg equilibrium. For sibship data, variance component models can be used for high resolution joint LD and linkage mapping of QTL. For a sibship of l children, denote their quantitative traits by a vector $\mathbf{y} = (y_1, \dots, y_l)^\tau$, genotypes at marker A by a vector $(A_1, A_2, \dots, A_l)^\tau$, and genotypes at marker B by a vector $(B_1, B_2, \dots, B_l)^\tau$. Here y_i is the trait value of the i -th offspring, A_i is the genotype of the i -th offspring at marker A , and B_i is the genotype of the i -th offspring at marker B . The log-likelihood function for these data is

$$L = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{y} - X\mu)^\tau \Sigma^{-1} (\mathbf{y} - X\mu). \quad (1)$$

The notations of model (1) are defined as follows. Σ is a $l \times l$ variance-covariance matrix defined as

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1l} \\ \rho_{21} & 1 & \dots & \rho_{2l} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{l1} & \rho_{l2} & \dots & 1 \end{pmatrix} \sigma^2,$$

where $\sigma^2 = \sigma_g^2 + \sigma_G^2 + \sigma_s^2 + \sigma_e^2$, σ_g^2 is the variance explained by the putative QTL Q , σ_G^2 is the polygenic variance, σ_s^2 is the shared environment residual variance, and σ_e^2 is the error variance. The genetic variances $\sigma_g^2 = \sigma_{ga}^2 + \sigma_{gd}^2$ and $\sigma_G^2 = \sigma_{Ga}^2 + \sigma_{Gd}^2$ are decomposed into additive and dominant components, respectively. $\rho_{ij} = \rho_{ji} = (\pi_{ijQ}\sigma_{ga}^2 + \Delta_{ijQ}\sigma_{gd}^2 + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4 + \sigma_s^2)/\sigma^2$ is the correlation between the i -th child and the j -th child, π_{ijQ} is the proportion of alleles shared identical by descent (IBD) at putative QTL Q by the i -th child and the j -th child, and Δ_{ijQ} is the probability that both alleles at the putative QTL Q shared by the i -th child and the j -th child are IBD [Pratt et al., 2000; Zhu and Elston, 2000]. To introduce the mean component $X\mu$ for log-likelihood (1), we consider the following regression [Fan and Xiong, 2002, 2003]

$$y_i = \beta + w_i\gamma + x_{Ai}\alpha_A + x_{Bi}\alpha_B + z_{Ai}\delta_A + z_{Bi}\delta_B + G_i + H_i + e_i, \quad (2)$$

where β is the overall mean, w_i is a row vector of covariates such as sex and age, γ is a column vector of regression coefficients of w_i , G_i is the polygenic effect, H_i is the shared environment residual effect, and e_i is the error term. Assume that G_i is normal $N(0, \sigma_G^2)$, H_i is normal $N(0, \sigma_s^2)$, and e_i is normal $N(0, \sigma_e^2)$. Moreover, G_i , H_i and e_i are independent. x_{Ai} , x_{Bi} , z_{Ai} and z_{Bi} are dummy random variables that are independent of G_i , H_i , and e_i defined by

$$x_{Ai} = \begin{cases} 2P_a & \text{if } A_i = AA \\ P_a - P_A & \text{if } A_i = Aa \\ -2P_A & \text{if } A_i = aa \end{cases}, \quad z_{Ai} = \begin{cases} -P_a^2 & \text{if } A_i = AA \\ P_a P_A & \text{if } A_i = Aa \\ -P_A^2 & \text{if } A_i = aa \end{cases}$$

$$x_{Bi} = \begin{cases} 2P_b & \text{if } B_i = BB \\ P_b - P_B & \text{if } B_i = Bb \\ -2P_B & \text{if } B_i = bb \end{cases}, \quad z_{Bi} = \begin{cases} -P_b^2 & \text{if } B_i = BB \\ P_b P_B & \text{if } B_i = Bb \\ -P_B^2 & \text{if } B_i = bb \end{cases}$$

α_A , α_B , δ_A and δ_B are the coefficients of the dummy variables x_{Ai} , x_{Bi} , z_{Ai} and z_{Bi} . X is the model matrix based on regression (2), and $\mu = (\beta, \gamma^\tau, \alpha_A, \alpha_B, \delta_A, \delta_B)^\tau$ is a vector of coefficients.

Fan and Xiong [2002] provide an intuitive rationale for model (2) as follows. Let μ_{ij} be the effect of genotype Q_iQ_j , $i, j = 1, 2$, $\mu_{12} = \mu_{21}$. Denote the overall population mean by $\mu_0 = \mu_{11}q_1^2 + 2\mu_{12}q_1q_2 + \mu_{22}q_2^2$, the average effect of gene substitution by $\alpha_Q = q_1\mu_{11} + (q_2 - q_1)\mu_{12} - q_2\mu_{22}$, and the dominant deviation by $\delta_Q = 2\mu_{12} - \mu_{11} - \mu_{22}$. Assume that marker A coincides with the trait locus Q , marker allele A is trait allele Q_1 and marker allele a is trait allele Q_2 . Fan and Xiong [2002] show that the trait value can be expressed as $y_i = \mu_0 + x_{Qi}\alpha_Q + z_{Qi}\delta_Q + e_i$, where $x_{Qi} = x_{Ai}$ and $z_{Qi} = z_{Ai}$. In practice, information about trait locus Q is unknown, but the information at marker loci is available. This prompts us to propose regression model (2) to describe the trait values. For the population data considered in Fan and Xiong [2002], the trait values are independent of each other. However, the trait values of a sibship are correlated to each other with variance covariance matrix Σ .

Suppose there are I sibships, in which some may contain only one offspring. Denote their log-likelihoods as L_1, \dots, L_I , where L_i is the log-likelihood of trait values \mathbf{y}_i of the i -th sibship or individual. Let Σ_i be variance-covariance matrix of \mathbf{y}_i , and X_i be its model matrix. Denote the total trait values $\mathbf{y} = (\mathbf{y}_1^\tau, \dots, \mathbf{y}_I^\tau)^\tau$, the total variance-covariance matrix by $\Sigma = \text{diag}(\Sigma_1, \dots, \Sigma_I)$, and model matrix $X = (X_1^\tau, \dots, X_I^\tau)^\tau$. Combining all sibships together, the overall log-likelihood is $L = \sum_{i=1}^I L_i = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{y} - X\mu)^\tau \Sigma^{-1} (\mathbf{y} - X\mu)$, where N is the total number of individuals of the I sibships. The

unknown parameters are $\mu = (\beta, \gamma, \alpha_A, \alpha_B, \delta_A, \delta_B)^T$, σ_{ga}^2 , σ_{gd}^2 , σ_{Ga}^2 , σ_{Gd}^2 , σ_s^2 , and σ_e^2 . Likelihood ratio tests (LRT) can be used to test significance of the parameters of interest.

Denote $a = \mu_{11} - (\mu_{11} + \mu_{22})/2$ and $d = \mu_{12} - (\mu_{11} + \mu_{22})/2$. In terms of traditional quantitative genetics [Falconer and Mackay, 1996], $\alpha_Q = a + (q_2 - q_1)d$ and $\delta_Q = 2d$. The additive variance $\sigma_{ga}^2 = 2q_1q_2\alpha_Q^2$ and the dominant variance $\sigma_{gd}^2 = (q_1q_2)^2\delta_Q^2$. To test the linkage of the trait locus to a particular position in the genome, the null hypothesis is $H_0: \sigma_{ga}^2 = \sigma_{gd}^2 = 0$ and the alternative hypothesis is $H_A: \sigma_{ga}^2 > 0$ or $\sigma_{gd}^2 > 0$. The corresponding LRT is a mixture of χ^2 variables [Self and Liang, 1987]. If only the additive variance σ_{ga}^2 (or dominant variance σ_{gd}^2) is modeled, the null hypothesis is $H_0: \sigma_{ga}^2 = 0$ (or $H_0: \sigma_{gd}^2 = 0$), and the alternative hypothesis is $H_A: \sigma_{ga}^2 > 0$ (or $\sigma_{gd}^2 > 0$). Then the corresponding LRT is a $\frac{1}{2} : \frac{1}{2}$ mixture of χ_1^2 and a point mass at 0 [Self and Liang, 1987].

Denote the measure of LD between QTL Q and marker A by $D_{AQ} = P(AQ_1) - q_1P_A$, the measure of LD between QTL Q and marker B by $D_{QB} = P(BQ_1) - q_1P_B$, and the measure of LD between marker A and marker B by $D_{AB} = P(AB) - P_AP_B$ [Hartl and Clark, 1989; Hedrick, 1987; Lewontin, 1964]. Let the additive and dominant variance-covariance matrices be

$$V_A = \begin{pmatrix} 2P_aP_A & 2D_{AB} \\ 2D_{AB} & 2P_bP_B \end{pmatrix} \quad \text{and} \quad V_D = \begin{pmatrix} P_a^2P_A^2 & D_{AB}^2 \\ D_{AB}^2 & P_b^2P_B^2 \end{pmatrix}. \quad (3)$$

As in Fan and Xiong [2002], we can show that the coefficients of regression equation (2) are

$$\begin{pmatrix} \alpha_A \\ \alpha_B \end{pmatrix} = V_A^{-1} \begin{pmatrix} 2D_{AQ} \\ 2D_{QB} \end{pmatrix} \alpha_Q, \quad \begin{pmatrix} \delta_A \\ \delta_B \end{pmatrix} = V_D^{-1} \begin{pmatrix} D_{AQ}^2 \\ D_{QB}^2 \end{pmatrix} \delta_Q. \quad (4)$$

Equations (4) imply that regression (2) simultaneously accounts for the LD and the effects of the putative QTL Q . The parameters of LD (i.e., D_{AQ} and D_{QB}) and gene effect (i.e., α_Q and δ_Q) are contained in the mean coefficients. In the presence of linkage to a particular position, the association between the trait locus and the markers can be tested based on equations (4). First, suppose that the presence of linkage is supported by both $\sigma_{ga}^2 > 0$ and $\sigma_{gd}^2 > 0$, which implies that both α_Q and δ_Q are not equal to 0. Then testing $H_0: \alpha_A = \alpha_B = \delta_A = \delta_B = 0$ vs. H_A : at least one of $\alpha_A, \alpha_B, \delta_A$, and δ_B is not 0, shows the association between the trait locus and the markers. Notice that this test will lead to 4 degrees of freedom, but the number of parameters D_{AQ} and D_{QB} is only 2. Hence, there should be only one or two coefficients of $\alpha_A, \alpha_B, \delta_A$, and δ_B , which is/are significantly different from 0 in the data analysis. Second, suppose that the presence of linkage is supported by additive variance $\sigma_{ga}^2 > 0$, but the dominant variance σ_{gd}^2 is not significantly larger than 0. Then testing $H_0: \alpha_A = \alpha_B = 0$ vs. H_A : at least one of α_A and α_B is not 0, shows the association between the trait locus and the markers. In this case, it is possible that only one of α_A and α_B is significantly different from 0 in the data analysis. Third, suppose that the presence of linkage is supported by the dominant variance $\sigma_{gd}^2 > 0$, but the additive variance σ_{ga}^2 is not significantly larger than 0. Then testing $H_0: \delta_A = \delta_B = 0$ vs. H_A : at least one of δ_A and δ_B is not 0, shows the association between the trait locus and the markers.

Suppose that only one marker A is used in the analysis. Then equations (4) can be replaced by $\alpha_A = D_{AQ}\alpha_Q/(P_aP_A)$, $\delta_A = D_{AQ}^2\delta_Q/(P_a^2P_A^2)$. Suppose that the presence of linkage is supported by both $\sigma_{ga}^2 > 0$ and $\sigma_{gd}^2 > 0$. Then testing $H_0: \alpha_A = \delta_A = 0$ vs. H_A : at least one of α_A and δ_A is not 0, shows the association between the trait locus and marker A . Again, there should be only one coefficient of α_A and

δ_A which is significantly different from 0 in data analysis, since only one parameter D_{AQ} is being tested. Suppose that the presence of linkage is supported by additive variance $\sigma_{ga}^2 > 0$, but the dominant variance σ_{gd}^2 is not significantly larger than 0. Then a test of $H_0: \alpha_A = 0$ vs. $H_A: \alpha_A \neq 0$ shows the association between the trait locus and marker A . On the other hand, if the presence of linkage is supported by the dominant variance $\sigma_{gd}^2 > 0$, but the additive variance σ_{ga}^2 is not significantly larger than 0, then a test of $H_0: \delta_A = 0$ vs. $H_A: \delta_A \neq 0$ shows the association between the trait locus and the marker A .

In practice, it may be reasonable to start with a variance component model which includes the covariates, but does not include the dummy variables x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi} . That is, to fit a reduced model $y_i = \beta + w_i\gamma + G_i + H_i + e_i$, instead of model (2) directly [Pratt et al., 2000]. This can achieve the initial objective of identifying linkage of trait values to a particular position in a region. Then, the dummy variables x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi} of markers A and B in the region can be included in the model to fit regression (2) for high resolution joint LD and linkage mapping. In this second step, the significant variables among $\sigma_{ga}^2, \sigma_{gd}^2, \alpha_A, \alpha_B, \delta_A$ and δ_B can be identified. Keeping only the significant variables in the final model, the likelihood ratio test of the final model against the model which assumes neither linkage nor association between the trait values and the markers can be calculated. By performing the analysis in this way, both linkage and LD information are used simultaneously to get a joint mapping of QTL.

Trait Variance-Covariance Matrix

For two siblings i and j in a sibship of size l , their trait covariance, conditional on the information of markers A and B , is $\text{Cov}(y_1, y_2 | I_A, I_B) = \hat{\pi}_{ijQ}\sigma_{ga}^2 + \Delta_{ijQ}\sigma_{gd}^2 + \sigma_{Gd}^2/2 + \sigma_{Gs}^2/4 + \sigma_s^2 = \hat{\rho}_{ij}\sigma^2$, where $\hat{\pi}_{ijQ} = E(\pi_{ijQ} | I_A, I_B)$, π_{ijQ} is the proportion of allele IBD at putative QTL Q , $\hat{\Delta}_{ijQ} = E(\Delta_{ijQ} | I_A, I_B)$ and Δ_{ijQ} is the probability that both alleles at the locus Q are IBD in the two offspring. The notations I_A and I_B represent the information on marker A and marker B . In the following paragraph, we use the interval mapping method given by Fulker and Cardon [1994] to estimate π_{ijQ} . In addition, we provide methods to estimate Δ_{ijQ} by the information on marker loci, which is not available in the literature.

Denote the recombination fraction between trait locus Q and marker A by θ_{AQ} , the recombination fraction between trait locus Q and marker B by θ_{QB} , and the recombination fraction between marker A and marker B by θ_{AB} . Fulker and Cardon [1994] propose calculating the proportion $\hat{\pi}_{ijQ}$ of alleles which are IBD at putative QTL Q for a sibpair i and j by $\hat{\pi}_{ijQ} = \alpha_\pi + \beta_{\pi A}\pi_{ijA} + \beta_{\pi B}\pi_{ijB}$, where π_{ijA} and π_{ijB} are the proportions of IBD alleles shared at marker A and marker B by sibpair i and j , respectively. The coefficients $\alpha_\pi, \beta_{\pi A}$ and $\beta_{\pi B}$ are functions of θ_{AQ}, θ_{QB} and θ_{AB} given by

$$\begin{aligned} \beta_{\pi A} &= \frac{(1 - 2\theta_{AQ})^2 - (1 - 2\theta_{AB})^2(1 - 2\theta_{QB})^2}{1 - (1 - 2\theta_{AB})^4} \\ \beta_{\pi B} &= \frac{(1 - 2\theta_{QB})^2 - (1 - 2\theta_{AB})^2(1 - 2\theta_{AQ})^2}{1 - (1 - 2\theta_{AB})^4} \\ \alpha_\pi &= \frac{1 - \beta_{\pi A} - \beta_{\pi B}}{2}. \end{aligned} \quad (5)$$

Let $\Delta_{ijA}, \Delta_{ijB}$ be the probability of sharing 2 alleles IBD at markers A and B for the sibpair i and j , respectively. We propose to estimate Δ_{ijQ} by

$$\hat{\Delta}_{ijQ} = \alpha + \beta_A\pi_{ijA} + \beta_B\pi_{ijB} + r_A\Delta_{ijA} + r_B\Delta_{ijB}. \quad (6)$$

In Appendices A, B and C, we show that under the assumption of no interference,

$$\begin{aligned}
 r_A &= \frac{(1 - 2\theta_{AQ})^4 - (1 - 2\theta_{QB})^4(1 - 2\theta_{AB})^4}{1 - (1 - 2\theta_{AB})^8} \\
 r_B &= \frac{(1 - 2\theta_{QB})^4 - (1 - 2\theta_{AQ})^4(1 - 2\theta_{AB})^4}{1 - (1 - 2\theta_{AB})^8} \\
 \beta_A &= \beta_{\pi A} - r_A, \beta_B = \beta_{\pi B} - r_B \\
 \alpha &= \frac{(1 - \psi_A)^2(1 - \psi_B)^2}{[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)]^2}, \tag{7}
 \end{aligned}$$

where $\beta_{\pi A}, \beta_{\pi B}$ are given in equations (5) [Fulker and Cardon, 1994], $\psi_A = \theta_{AQ}^2 + (1 - \theta_{AQ})^2$ and $\psi_B = \theta_{QB}^2 + (1 - \theta_{QB})^2$. Assuming that the positions of marker *A* and marker *B* are known, θ_{AB} can be calculated through an appropriate mapping function. Under the assumption of no interference, the mapping function is a Haldane's function $\theta = [1 - \exp(-2\lambda)]/2$ for a map distance λ . Then only one of θ_{AQ} and θ_{QB} is an unknown parameter since the other can be calculated through Trow's formula $\theta_{AB} = \theta_{AQ}(1 - \theta_{QB}) + (1 - \theta_{AQ})\theta_{QB}$ [Lange, 1997, p. 117].

Non-Centrality Parameters for Linkage Analysis

To calculate the non-centrality parameters of likelihood ratio tests, we follow an idea of Sham et al. [2000] according to the general statistical theory [Stuart and Ord, 1991]. Under the null or alternative hypothesis, the maximum-likelihood estimates of the parameters can be calculated. Taking the expectations of the log-likelihoods, the non-centrality parameters are then calculated as twice the difference between the log-likelihoods under the null and alternative hypotheses. The details are presented in Appendices D and E.

Non-Centrality Parameters for Association Study

In the following, we assume that the data are composed of three sub-samples: *n* independent individuals, *m* independent sibpairs, and *k* independent tri-sibships, each having 3 sibs. Moreover, we assume that *n*, *m* and *k* are sufficiently large, so that large sample theory applies. In practice, the sizes *n* and *m* of individuals and sibpairs are likely to be large. The size *k* of tri-sibships can be large. However, it is difficult to collect a large sample of sibships each having more than 3 sibs. In the event that a large sample of sibships each having more than 3 sibs is available, the following principle is still valid, but the corresponding formulas must be calculated accordingly. Assuming that there are no covariates, the regression coefficients are $\mu = (\beta, \alpha_A, \alpha_B, \delta_A, \delta_B)^T$. Consider the overall log-likelihood $L = \sum_{i=1}^I L_i$, $I = n + m + k$. Denote the total number of individuals by *N*, i.e., $N = n + 2m + 3k$. Let $\hat{\beta}, \hat{\alpha}_A, \hat{\alpha}_B, \hat{\delta}_A, \hat{\delta}_B, \hat{\Sigma}_i, \hat{\Sigma}$ be the maximum likelihood estimators of $\beta, \alpha_A, \alpha_B, \delta_A, \delta_B, \Sigma_i, \Sigma$. The estimate of μ is $\hat{\mu} = [X^T \hat{\Sigma}^{-1} X]^{-1} X^T \hat{\Sigma}^{-1} \vec{y} = [\sum_{i=1}^I X_i^T \hat{\Sigma}_i^{-1} X_i]^{-1} \sum_{i=1}^I X_i^T \hat{\Sigma}_i^{-1} \vec{y}_i$. Let *H* be a $q \times 5$ test matrix of rank *q* ($q \leq 5$). By Graybill [1976], chapter 6, the test statistic of a hypothesis $H\mu = 0$ is non-central *F*(*q*, *N* - 5) defined by

$$F = \frac{(H\hat{\mu})^T [H(X^T \hat{\Sigma}^{-1} X)^{-1} H^T]^{-1} (H\hat{\mu})}{Y^T [\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} X (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1}] Y} \frac{N - 5}{q}.$$

The non-centrality parameter of the test statistic *F* can be calculated by $\lambda = (H\mu)^T [H(X^T \Sigma^{-1} X)^{-1} H^T]^{-1} H\mu$. Under the assumption of large sample sizes *n*, *m* and *k*, we show in Appendix F that

$$\sum_{i=1}^{n+m+k} X_i^T \Sigma_i^{-1} X_i \approx \text{diag}(a_1, a_2 V_A, a_3 V_D) / \sigma^2, \tag{8}$$

where a_1, a_2 and a_3 are constants given by equations (19) in Appendix F.

In the presence of an additive effect, i.e., $\sigma_{ga}^2 > 0$ or $\alpha_Q \neq 0$, we may test the null hypothesis $H_{AB,a} : \alpha_A = \alpha_B = 0$ or $D_{AQ} = D_{QB} = 0$. The test matrix *H* is defined by

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Let us denote the corresponding *F* test statistic by $F_{AB,a}$, and the non-centrality parameter by $\lambda_{AB,a}$. Then we have from (4) and (8) that

$$\begin{aligned}
 \lambda_{AB,a} &\approx \frac{1}{\sigma^2} a_2 (\alpha_A \quad \alpha_B) V_A \begin{pmatrix} \alpha_A \\ \alpha_B \end{pmatrix} \\
 &= \frac{2a_2}{\sigma^2} \alpha_Q^2 \frac{P_b P_B D_{AQ}^2 - 2D_{AQ} D_{AB} D_{QB} + P_a P_A D_{QB}^2}{P_a P_A P_b P_B - D_{AB}^2} \\
 &= \frac{a_2}{\sigma^2} \sigma_{ga}^2 \frac{R_{AQ}^2 - 2R_{AQ} R_{AB} R_{QB} + R_{QB}^2}{1 - R_{AB}^2},
 \end{aligned}$$

where $R_{AB} = D_{AB} / \sqrt{P_a P_A P_b P_B}$, $R_{AQ} = D_{AQ} / \sqrt{P_a P_A q_1 q_2}$, and $R_{QB} = D_{QB} / \sqrt{q_1 q_2 P_b P_B}$ are three ratios [Almasy et al., 1999; Fan and Xiong, 2002, 2003; Sham et al., 2000].

In the presence of a dominant effect, i.e., $\sigma_{gd}^2 > 0$ or $\delta_Q \neq 0$, we may test the null hypothesis $H_{AB,d} : \delta_A = \delta_B = 0$ or $D_{AQ} = D_{QB} = 0$. The test matrix *H* is defined by

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Denote the corresponding *F* test statistic by $F_{AB,d}$, and the non-centrality parameter by $\lambda_{AB,d}$. Then we have from (4) and (8) that

$$\begin{aligned}
 \lambda_{AB,d} &\approx \frac{a_3}{\sigma^2} (\delta_A \quad \delta_B) V_D \begin{pmatrix} \delta_A \\ \delta_B \end{pmatrix} \\
 &= \frac{a_3}{\sigma^2} \delta_Q^2 \frac{P_b^2 P_B^2 D_{AQ}^4 - 2D_{AQ}^2 D_{AB}^2 D_{QB}^2 + P_a^2 P_A^2 D_{QB}^4}{P_a^2 P_A^2 P_b^2 P_B^2 - D_{AB}^4} \\
 &= \frac{a_3}{\sigma^2} \sigma_{gd}^2 \frac{R_{AQ}^4 - 2R_{AQ}^2 R_{AB}^2 R_{QB}^2 + R_{QB}^4}{1 - R_{AB}^4}.
 \end{aligned}$$

In the presence of both additive and dominant effects, i.e., $\sigma_{ga}^2 > 0$ and $\sigma_{gd}^2 > 0$, we may test the null hypothesis $H_{AB,ad} : \alpha_A = \alpha_B = \delta_A = \delta_B = 0$. The test matrix *H* is defined by

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Denote the corresponding *F* test statistic by $F_{AB,ad}$, and the non-centrality parameter by $\lambda_{AB,ad}$. Then, $\lambda_{AB,ad} = \lambda_{AB,a} + \lambda_{AB,d}$. Assume that only one marker *A* is used in the analysis. The non-centrality parameter is

$$\lambda_{A,ad} \approx [1/\sigma^2] [a_2 \sigma_{ga}^2 R_{AQ}^2 + a_3 \sigma_{gd}^2 R_{AQ}^4]$$

for the null hypothesis $H_{A,ad} : \alpha_A = \delta_A = 0$. Correspondingly, we denote the *F* test statistic by $F_{A,ad}$. Similarly, $\lambda_{A,a} \approx [a_2/\sigma^2] \sigma_{ga}^2 R_{AQ}^2$ is the non-centrality parameter of the test statistic $F_{A,a}$ for the

Table 1. Joint distribution of π_Q , π_A and π_B of a sibpair. Here subscripts ij are omitted from π_{ijQ} , π_{ijA} and π_{ijB}

Markers		Trait locus			Total probability
π_A	π_B	$\pi_Q = 1$	$\pi_Q = 1/2$	$\pi_Q = 0$	
1	1	$\psi_A^2/4 \cdot \psi_B^2$	$\psi_A(1 - \psi_A)/2 \cdot \psi_B(1 - \psi_B)$	$(1 - \psi_A)^2/4 \cdot (1 - \psi_B)^2$	$[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)]^2/4$
	1/2	$\psi_A^2/4 \cdot 2\psi_B(1 - \psi_B)$	$\psi_A(1 - \psi_A)/2 \cdot (1 - 2\psi_B + 2\psi_B^2)$	$(1 - \psi_A)^2/4 \cdot 2\psi_B(1 - \psi_B)$	$[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)] \cdot [(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]/2$
	0	$\psi_A^2/4 \cdot (1 - \psi_B)^2$	$\psi_A(1 - \psi_A)/2 \cdot \psi_B(1 - \psi_B)$	$(1 - \psi_A)^2/4 \cdot \psi_B^2$	$[(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]^2/4$
1/2	1	$\psi_A(1 - \psi_A)/2 \cdot \psi_B^2$	$(1 - 2\psi_A + 2\psi_A^2)/2 \cdot \psi_B(1 - \psi_B)$	$\psi_A(1 - \psi_A)/2 \cdot (1 - \psi_B)^2$	$[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)] \cdot [(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]/2$
	1/2	$\psi_A(1 - \psi_A)/2 \cdot 2\psi_B(1 - \psi_B)$	$(1 - 2\psi_A + 2\psi_A^2)/2 \cdot (1 - 2\psi_B + 2\psi_B^2)$	$\psi_A(1 - \psi_A)/2 \cdot 2\psi_B(1 - \psi_B)$	$[[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)]^2 + [(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]^2]/2$
	0	$\psi_A(1 - \psi_A)/2 \cdot (1 - \psi_B)^2$	$(1 - 2\psi_A + 2\psi_A^2)/2 \cdot \psi_B(1 - \psi_B)$	$\psi_A(1 - \psi_A)/2 \cdot \psi_B^2$	$[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)] \cdot [(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]/2$
0	1	$(1 - \psi_A)^2/4 \cdot \psi_B^2$	$\psi_A(1 - \psi_A)/2 \cdot \psi_B(1 - \psi_B)$	$\psi_A^2/4 \cdot (1 - \psi_B)^2$	$[(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]^2/4$
	1/2	$(1 - \psi_A)^2/4 \cdot 2\psi_B(1 - \psi_B)$	$\psi_A(1 - \psi_A)/2 \cdot (1 - 2\psi_B + 2\psi_B^2)$	$\psi_A^2/4 \cdot 2\psi_B^2(1 - \psi_B)$	$[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)] \cdot [(1 - \psi_A)\psi_B + \psi_A(1 - \psi_B)]/2$
	0	$(1 - \psi_A)^2/4 \cdot (1 - \psi_B)^2$	$\psi_A(1 - \psi_A)/2 \cdot \psi_B(1 - \psi_B)$	$\psi_A^2/4 \cdot \psi_B^2$	$[\psi_A\psi_B + (1 - \psi_A)(1 - \psi_B)]^2/4$
Total probability		1/4	1/2	1/4	1

null hypothesis $H_{A,a} : \alpha_A = 0$. The non-centrality parameter of the test statistic $F_{A,d}$ for the null hypothesis $H_{A,d} : \delta_A = 0$ is $\lambda_{A,d} \approx [a_3/\sigma^2]\sigma_{gd}^2 R_{AQ}^4$.

Results and Power Comparisons

To investigate the performance of the approximations given by the equations (6) and (7), we calculate the numerical values of the estimates of probability sharing 2 alleles IBD for sibs. We then perform power calculations and comparisons of LD and linkage mapping.

Estimates of the Probability of Sharing 2 Alleles IBD for Sibs

Tables 2 and 3 give the interval estimates of $\hat{\Delta}_Q$ by π_A , π_B , Δ_A and Δ_B under Haldane's function. Table 2 takes a map distance $\lambda_{AB} = 20$ cM, and table 3 takes $\lambda_{AB} = 100$ cM (i.e., marker A and marker B are unlinked). In each table, the interval is divided to be four equally spaced subintervals. This gives five equally spaced locations for the trait locus. In each table, the estimates of $\hat{\Delta}_Q$ are equal to Δ_A on the first location. Hence, Δ_A can fully estimate $\hat{\Delta}_Q$ on the first location. On the other hand, the estimates of $\hat{\Delta}_Q$ are equal to Δ_B on the fifth location. Hence, Δ_B can fully estimate $\hat{\Delta}_Q$ on the fifth location. In both tables, the estimates of $\hat{\Delta}_Q$ on the second location are intermediates between location 1 and location 3. The estimates of $\hat{\Delta}_Q$ on the fourth

location are intermediates between location 3 and location 5. In table 2, the estimates of $\hat{\Delta}_Q$ on the third location are close to the average of Δ_A and Δ_B (see the discussion in the following paragraph). In table 3, the estimates $\hat{\Delta}_Q$ on the third location tend to the expected value 0.25 since the location is unlinked to both markers.

Assume the two markers A and B are close, for instance ≤ 20 cM as suggested in Fulker and Cardon [1994]. By taking the first order approximation $(1 - x)^n \approx 1 - nx$ for small x , we have an approximation

$$\begin{aligned} r_A &\approx \frac{(1 - 4 \cdot 2\theta_{AQ}) - (1 - 4 \cdot 2\theta_{QB})(1 - 4 \cdot 2\theta_{AB})}{1 - (1 - 8 \cdot 2\theta_{AB})} \\ &\approx \frac{(1 - 8\theta_{AQ}) - (1 - 8\theta_{QB} - 8\theta_{AB})}{16\theta_{AB}} \\ &\approx \frac{-\theta_{AQ} + \theta_{QB} + (\theta_{AQ} + \theta_{QB})}{2\theta_{AB}} = \frac{\theta_{QB}}{\theta_{AB}}. \end{aligned}$$

Similarly, we can show that $r_B \approx \theta_{AQ}/\theta_{AB}$. Combining these results with equation (10) in Fulker and Cardon [1994], we have that $\beta_A \approx 0$ and $\beta_B \approx 0$. Using the small map interval approximations to replace the recombination fraction, we have $\beta_A \approx 0$, $\beta_B \approx 0$, $r_A \approx \lambda_{QB}/\lambda_{AB}$, $r_B \approx \lambda_{AQ}/\lambda_{AB}$, where λ_{ij} is the map distance between locus i and locus j . When the two markers A and B are close, $\psi_A \approx 1$ and $\psi_B \approx 1$, which implies that $\alpha \approx 0$. Therefore, the estimates $\hat{\Delta}_Q$ on the third location in table 2 are approximately equal to the average of Δ_A and Δ_B .

Table 2. Interval estimates of $\hat{\Delta}_Q$ by π_A , π_B , Δ_A and Δ_B , for the flanking markers separated by $\lambda_{AB} = 20$ cM under Haldane's mapping function

Parameters				Locations				
π_A	Δ_A	π_B	Δ_B	1	2	3	4	5
1	1	1	1	1.00	0.94	0.93	0.94	0.00
1	1	1/2	1/2	1.00	0.83	0.70	0.59	0.50
1	1	1/2	1/4	1.00	0.79	0.60	0.43	0.25
1	1	1/2	0	1.00	0.75	0.51	0.27	0.00
1	1	1/4	0	1.00	0.73	0.49	0.25	0.00
1	1	0	0	1.00	0.72	0.46	0.23	0.00
1/2	1/2	1	1	0.50	0.59	0.70	0.83	1.00
1/2	1/2	1/2	1/2	0.50	0.47	0.46	0.47	0.50
1/2	1/2	1/2	1/4	0.50	0.43	0.37	0.31	0.25
1/2	1/2	1/2	0	0.50	0.39	0.28	0.16	0.00
1/2	1/2	1/4	0	0.50	0.37	0.26	0.14	0.00
1/2	1/2	0	0	0.50	0.36	0.23	0.11	0.00
1/2	1/4	1	1	0.25	0.43	0.60	0.79	1.00
1/2	1/4	1/2	1/2	0.25	0.31	0.37	0.43	0.50
1/2	1/4	1/2	1/4	0.25	0.27	0.28	0.27	0.25
1/2	1/4	1/2	0	0.25	0.23	0.18	0.11	0.00
1/2	1/4	1/4	0	0.25	0.21	0.16	0.09	0.00
1/2	1/4	0	0	0.25	0.20	0.14	0.07	0.00
1/2	0	1	1	0.00	0.27	0.51	0.75	1.00
1/2	0	1/2	1/2	0.00	0.16	0.28	0.39	0.50
1/2	0	1/2	1/4	0.00	0.11	0.18	0.23	0.25
1/2	0	1/2	0	0.00	0.07	0.09	0.07	0.00
1/2	0	1/4	0	0.00	0.06	0.07	0.05	0.00
1/2	0	0	0	0.00	0.04	0.05	0.03	0.00
1/4	0	1	1	0.00	0.25	0.49	0.73	1.00
1/4	0	1/2	1/2	0.00	0.14	0.26	0.37	0.50
1/4	0	1/2	1/4	0.00	0.09	0.16	0.21	0.25
1/4	0	1/2	0	0.00	0.05	0.07	0.06	0.00
1/4	0	1/4	0	0.00	0.04	0.05	0.04	0.00
1/4	0	0	0	0.00	0.02	0.02	0.01	0.00
0	0	0	0	0.00	0.00	0.00	0.00	0.00
r_A				1.00	0.64	0.37	0.17	0.00
r_B				0.00	0.17	0.37	0.64	1.00
β_A				0.00	0.08	0.09	0.05	0.00
β_B				0.00	0.05	0.09	0.08	0.00

Table 3. Interval estimates of $\hat{\Delta}_Q$ by π_A , π_B , Δ_A and Δ_B , for the flanking markers separated by $\lambda_{AB} = 100$ cM under Haldane's mapping function

Parameters				Locations				
π_A	Δ_A	π_B	Δ_B	1	2	3	4	5
1	1	1	1	1.00	0.50	0.40	0.50	1.00
1	1	1/2	1/2	1.00	0.48	0.33	0.31	0.50
1	1	1/2	1/4	1.00	0.48	0.33	0.28	0.25
1	1	1/2	0	1.00	0.47	0.33	0.25	0.00
1	1	1/4	0	1.00	0.46	0.30	0.19	0.00
1	1	0	0	1.00	0.45	0.27	0.13	0.00
1/2	1/2	1	1	0.50	0.31	0.33	0.48	1.00
1/2	1/2	1/2	1/2	0.50	0.29	0.27	0.29	0.50
1/2	1/2	1/2	1/4	0.50	0.29	0.26	0.26	0.25
1/2	1/2	1/2	0	0.50	0.29	0.26	0.22	0.00
1/2	1/2	1/4	0	0.50	0.28	0.23	0.17	0.00
1/2	1/2	0	0	0.50	0.27	0.20	0.11	0.00
1/2	1/4	1	1	0.25	0.28	0.33	0.48	1.00
1/2	1/4	1/2	1/2	0.25	0.26	0.26	0.29	0.50
1/2	1/4	1/2	1/4	0.25	0.26	0.26	0.26	0.25
1/2	1/4	1/2	0	0.25	0.26	0.25	0.22	0.00
1/2	1/4	1/4	0	0.25	0.25	0.23	0.17	0.00
1/2	1/4	0	0	0.25	0.24	0.20	0.11	0.00
1/2	0	1	1	0.00	0.25	0.33	0.47	1.00
1/2	0	1/2	1/2	0.00	0.22	0.26	0.29	0.50
1/2	0	1/2	1/4	0.00	0.22	0.25	0.26	0.25
1/2	0	1/2	0	0.00	0.22	0.25	0.22	0.00
1/2	0	1/4	0	0.00	0.21	0.22	0.17	0.00
1/2	0	0	0	0.00	0.20	0.19	0.11	0.00
1/4	0	1	1	0.00	0.19	0.30	0.46	1.00
1/4	0	1/2	1/2	0.00	0.17	0.23	0.28	0.50
1/4	0	1/2	1/4	0.00	0.17	0.23	0.25	0.25
1/4	0	1/2	0	0.00	0.17	0.22	0.21	0.00
1/4	0	1/4	0	0.00	0.16	0.19	0.16	0.00
1/4	0	0	0	0.00	0.15	0.16	0.10	0.00
0	0	0	0	0.00	0.09	0.14	0.09	0.00
r_A				1.00	0.14	0.02	0.00	0.00
r_B				0.00	0.00	0.02	0.14	1.00
β_A				0.00	0.23	0.12	0.04	0.00
β_B				0.00	0.04	0.12	0.23	0.00

Comparisons with the 'AbAw' Approach of Fulker, Abecasis et al.

To compare the method developed in this paper with the 'AbAw' approach developed by Fulker and Abecasis et al., we present the theoretical expectations of the statistics for LD mapping of 1,000 sibpairs in table 4. The results of 'AbAw' approach by Fulker and Abecasis et al. are directly taken from table 5, p 1625, Sham et al. [2000]. The QTL is

assumed to be additive with $\sigma_{ga}^2 = 0.2$. The shared residual environment variance, σ_s^2 , is set to be either 0 or 0.4, such as those in table 3 and 5, Fulker et al. [1999], or table 5, Sham et al. [2000]. The error variance is set to be either 0.8 or 0.4, correspondingly. Moreover, it is assumed that there is no polygenic effects, and there is no putative dominant variance; thus, the total variance is 1. The QTL Q and marker A are assumed to be biallelic with equal allele fre-

Table 4. Empirical values vs. theoretical expectations of statistics, compared with results of table 5, Sham et al. [2000], when $\sigma_{ga}^2 = 0.2$, $\sigma_{gd}^2 = \sigma_{Ga}^2 = \sigma_{Gd}^2 = 0$

D_{AQ}	Sham et al. [2000], theory		Current Method											
	BP	WP	theory			$\theta_{AQ} = 0$			$\theta_{AQ} = 0.02$			$\theta_{AQ} = 0.04$		
			$\lambda_{A,a} + 1$	α_A	LRT	$F_{A,a}$	$\hat{\alpha}_A$	LRT	$F_{A,a}$	$\hat{\alpha}_A$	LRT	$F_{A,a}$	$\hat{\alpha}_A$	
$\sigma_s^2 = 0.0, \sigma_c^2 = 0.8$														
0.25	319.45	118.78	384.84	0.632	424.52	495.96	0.632	406.64	473.40	0.620	386.68	448.80	0.607	
0.20	192.82	74.77	246.66	0.506	259.18	285.82	0.507	249.38	274.38	0.498	238.20	261.52	0.488	
0.10	45.62	18.95	62.41	0.253	62.28	63.86	0.256	60.44	61.94	0.252	58.00	59.40	0.247	
0.05	11.97	5.45	16.35	0.127	14.74	14.82	0.125	14.32	14.42	0.123	13.86	13.94	0.121	
0.025	3.73	2.11	4.84	0.063	3.54	3.54	0.061	3.49	3.50	0.061	3.42	3.42	0.060	
$\sigma_s^2 = 0.4, \sigma_c^2 = 0.4$														
0.25	224.14	224.14	401.0	0.632	422.80	499.58	0.630	399.66	470.72	0.615	373.56	438.68	0.598	
0.20	137.97	137.97	257.0	0.506	258.44	289.24	0.506	246.08	274.48	0.494	231.56	257.04	0.480	
0.10	33.52*	33.52*	65.0	0.253	60.04	61.70	0.247	57.72	59.26	0.242	54.78	56.16	0.235	
0.05	9.03	9.03	17.0	0.127	13.96	14.04	0.119	13.40	13.48	0.117	12.84	12.91	0.114	
0.025	3.0	3.0	5.0	0.063	3.44	3.44	0.059	3.36	3.36	0.058	3.26	3.28	0.057	

A sample of 50,000 sibpairs are generated by simulation program Ldsimul. The reported values of statistics $F_{A,a}$ and likelihood ratio test (LRT) are divided by 50 to make comparison with results of table 5, Sham et al. [2000], where the simulation results are averages of 100 replicate samples of 1,000 sib pairs.

BP = Between pairs; WP = within pairs. LRT is calculated by $2[\ln L_A - \ln L_N]$, where L_A is maximum likelihood under $H_A: \alpha_A \neq 0$, and L_N is maximum likelihood under $H_N: \alpha_A = 0$.

$$F = \frac{(H\hat{\mu})^T [H(X^T \hat{\Sigma}^{-1} X)^{-1} H^T]^{-1} (H\hat{\mu})(N-2)}{Y^T [\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} X (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1}] Y}, \mu = (\beta, \alpha_A)^T, \text{ and } H = (0, 1).$$

* 36.52 in Sham et al. [2000], table 5, should be 33.52.

quencies. The measure D_{AQ} of LD varies from complete disequilibrium, 0.25, to weak disequilibrium, 0.025.

In table 4, the statistic $F_{A,a}$ is approximately distributed as non-central $\chi^2(1)$, since the sample size of 1,000 sibpairs is large enough for asymptotic property to hold. The theoretical expectations of the χ^2 statistics are the non-centrality parameters plus 1, i.e., $\lambda_{A,a} + 1$. To perform simulation studies, samples of 50,000 sibpairs are generated by the simulation program Ldsimul. The reported values of statistics $F_{A,a}$ and LRT are divided by 50 to be comparable with the results of table 5, Sham et al. [2000], where the simulation results are averages of 100 replicate samples of 1,000 sib pairs. From the results of table 4, it is clear that either $F_{A,a}$ or LTR is more powerful than any of between-pairs and within-pairs approaches of Fulker and Abecasis et al. 'AbAw' approach [Fulker et al., 1999; Sham et al., 2000]. The empirical estimates, $\hat{\alpha}_A$, of the parameter α_A are fairly close. In the presence of strong disequilibrium $D_{AQ} \geq 0.20$, both LRTs and F statistics tend to overestimate the theoretical expectations of the χ^2 statistics. In the weak disequilibrium $D_{AQ} \leq 0.10$, both LRTs and F statistics tend to underestimate the theoretical expectations of the χ^2 statistics.

Comparisons of Sample Sizes and Power for LD Mapping

In the sample size and power calculations, we take an additive polygenic variance $\sigma_{Ga}^2 = 0.10$, polygenic dominant variance $\sigma_{Gd}^2 = 0.05$, and shared environment residual variance $\sigma_s^2 = 0$. For sibpairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sibpair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5, \Delta_A = 0.25$ for sibpair 1 and 3; and $\pi_A = \pi_B = 0.5, \Delta_A = \Delta_B = 0.25$ for sibpair 2 and 3. Suppose that $\mu_{11} = a, \mu_{12} = \mu_{21} = d$ and $\mu_{22} = -a$. Denote heritability by h^2 which is defined by $h^2 = \sigma_{ga}^2 / \sigma^2$. Let λ_{AB} be the map distance between marker A and marker B . Under the assumption of no interference, we may calculate the recombination fraction $\theta_{AB} = [1 - \exp(-2\lambda_{AB})]/2$. Similarly, we may calculate the recombination fractions θ_{AQ} and θ_{QB} by the map distances λ_{AQ} and λ_{QB} .

Figure 1 gives the required number of sibpairs (fig. 1A, B) and tri-sibships (fig. 1C, D) of test statistics $F_{AB,ad}, F_{AB,as}, F_{AB,bs}, F_{A,ad}, F_{A,as},$ and $F_{A,d}$ against the heritability h^2 at 0.01 significant level and 0.80 power, for a mode of dominant inheritance $a = d = 1.0$ (fig. 1A, B), and a mode of recessive inheritance $a = 1.0, d = -0.5$ (fig. 1C, D), respectively. In the figure, we take equal allele frequencies

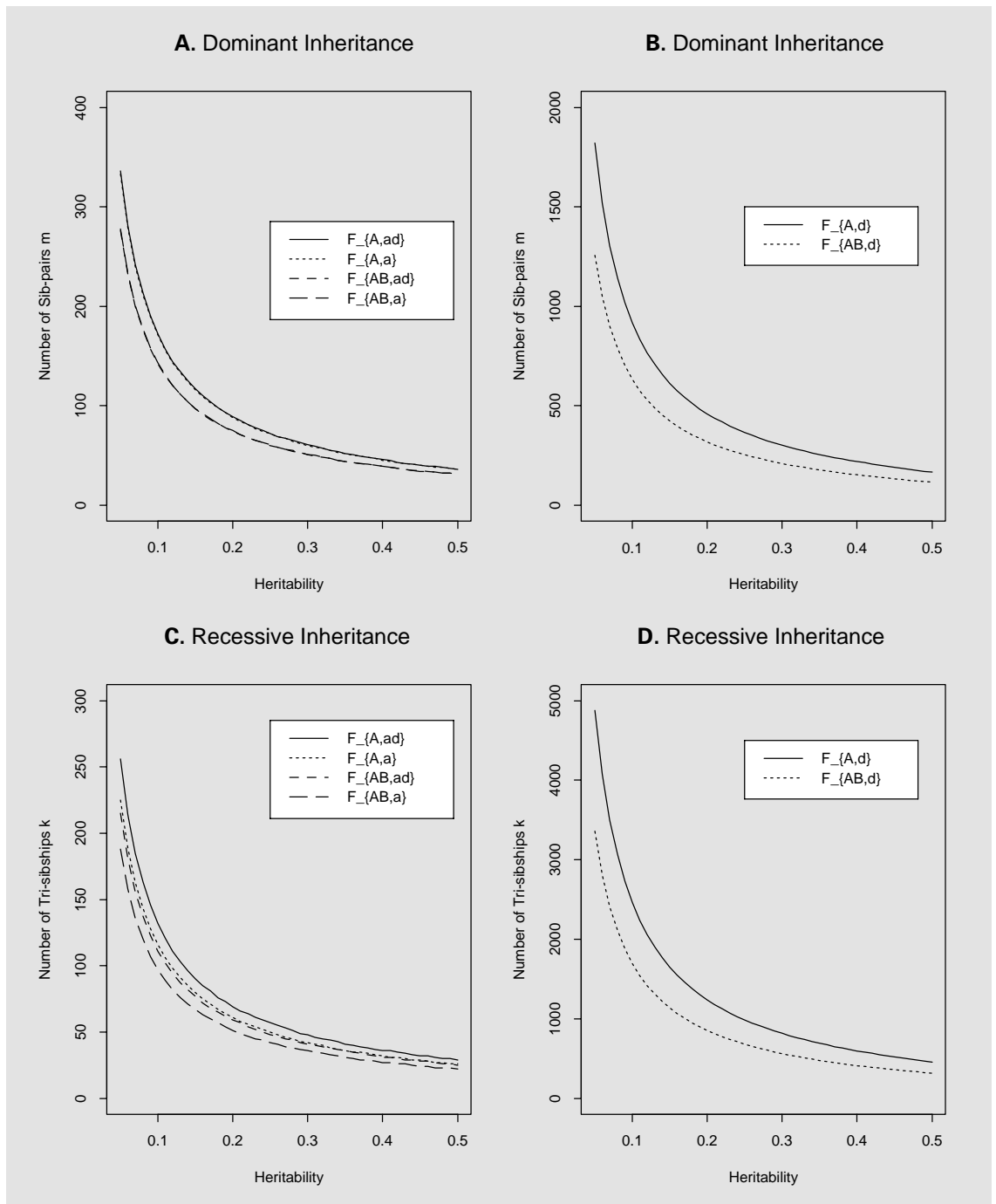


Fig. 1. Number of sibpairs (**A, B**) or tri-sibships (**C, D**) of test statistics $F_{AB, ad}$, $F_{AB, a}$, $F_{AB, d}$, $F_{A, ad}$, $F_{A, a}$, and $F_{A, d}$ against the heritability h^2 at 0.01 significant level and 0.80 power, when $q_1 = P_A = P_B = 0.50$, $D_{AB} = 0.10$, $D_{AQ} = D_{QB} = 0.15$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, $\sigma_{G_a}^2 = 0.10$, $\sigma_{G_d}^2 = 0.05$, $\sigma_s^2 = 0$, for a mode of dominant inheritance $a = d = 1.0$ (**A, B**), and a mode of recessive inheritance $a = 1.0$, $d = -0.5$ (**C, D**), respectively. For the sibpairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For the tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sibpair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5$, $\Delta_A = 0.25$ for sibpair 1 and 3; and $\pi_A = \pi_B = 0.5$, $\Delta_A = \Delta_B = 0.25$ for sibpair 2 and 3.

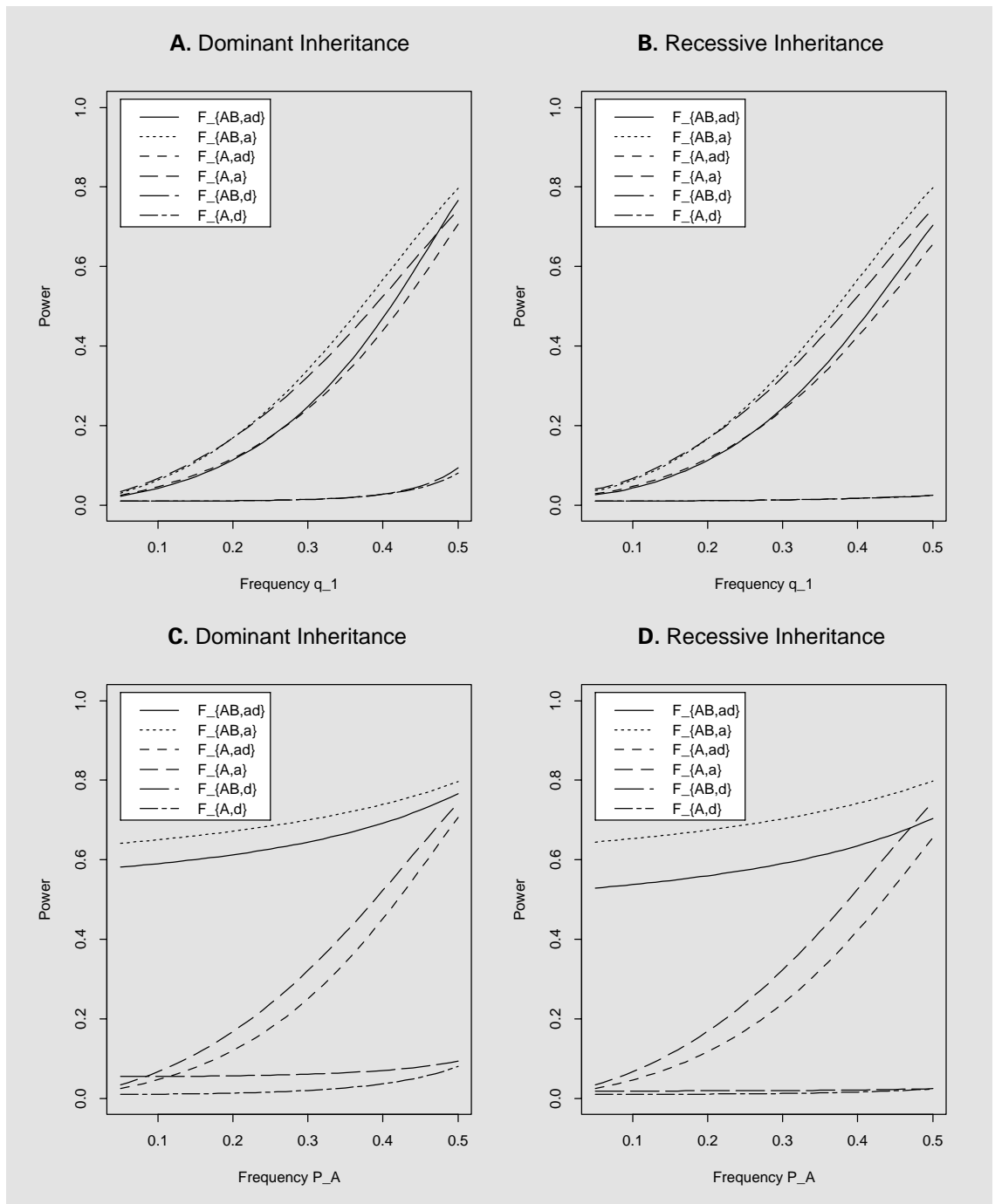


Fig. 2. Power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against trait frequency q_1 or marker allele frequency P_A at 0.01 significant level, when $P_A = 0.5$ (**A, B**), $q_1 = 0.5$ (**C, D**), $P_B = 0.50$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, $\sigma_{G_a}^2 = 0.10$, $\sigma_{G_d}^2 = 0.05$, $\sigma_s^2 = 0$, and $h^2 = 0.25$, for a mode of dominant inheritance $a = d = 1.0$, and a mode of recessive inheritance $a = 1.0$, $d = -0.5$, respectively. The linkage disequilibrium coefficients are $D_{AB} = (\min(P_A, P_B) - P_A P_B)/2$, $D_{AQ} = (\min(P_A, q_1) - P_A q_1)/2$ and $D_{QB} = (\min(P_B, q_1) - P_B q_1)/2$. For the $m = 30$ sibpairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For the $k = 20$ tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sibpair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5$, $\Delta_A = 0.25$ for sibpair 1 and 3; and $\pi_A = \pi_B = 0.5$, $\Delta_A = \Delta_B = 0.25$ for sibpair 2 and 3.

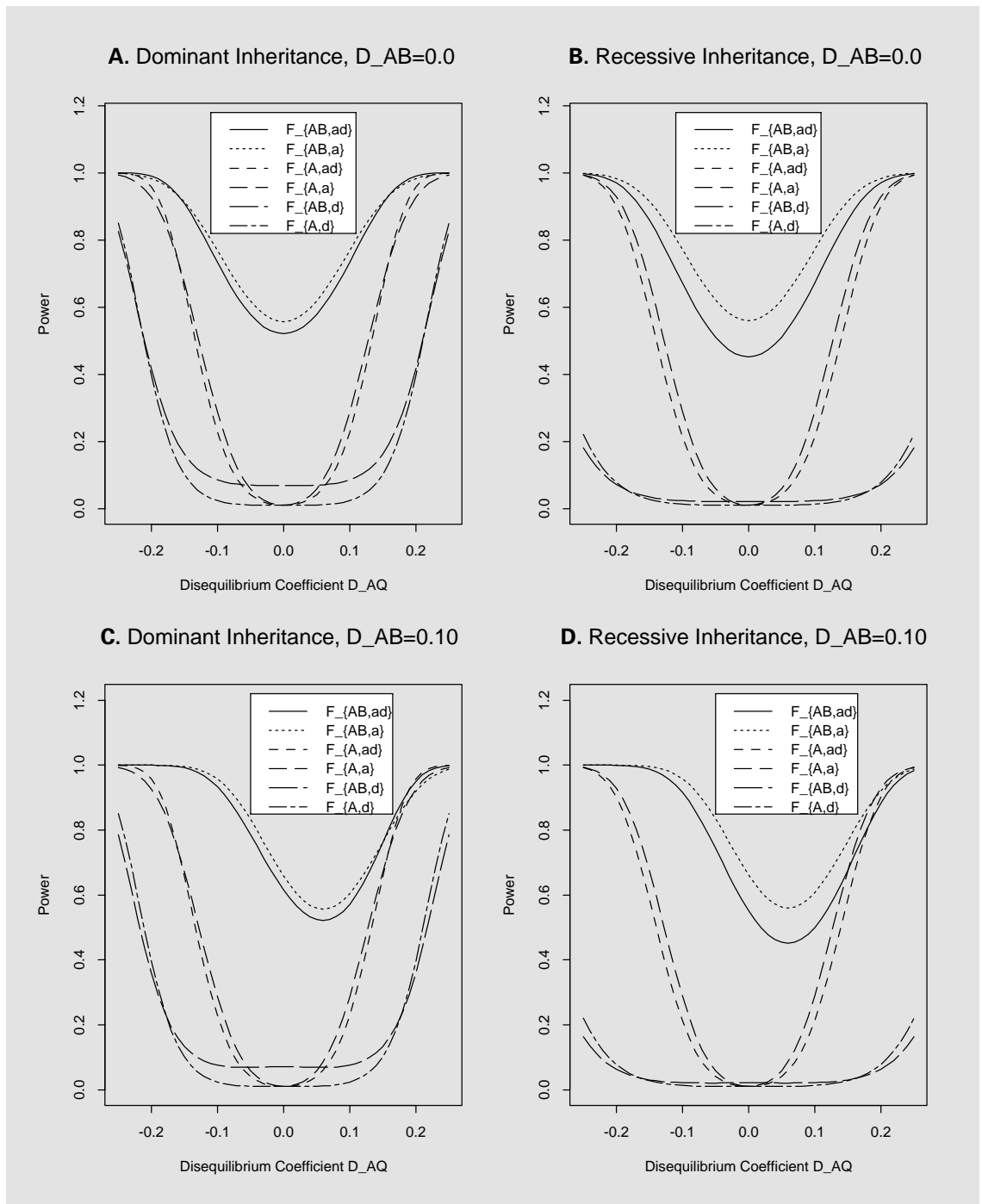


Fig. 3. Power of test statistics $F_{AB, ad}$, $F_{AB, a}$, $F_{AB, d}$, $F_{A, ad}$, $F_{A, a}$, and $F_{A, d}$ against LD coefficient D_{AQ} at 0.01 significant level, when $q_1 = P_A = P_B = 0.50$, $D_{QB} = 0.15$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, $\sigma_{Ga}^2 = 0.10$, $\sigma_{Gd}^2 = 0.05$, $\sigma_s^2 = 0$, and $h^2 = 0.15$, for a mode of dominant inheritance $a = d = 1.0$ (**A** and **C**), and a mode of recessive inheritance $a = 1.0$, $d = -0.5$ (**B** and **D**), respectively. For the $m = 30$ sibpairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For the $k = 20$ tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sibpair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5$, $\Delta_A = 0.25$ for sibpair 1 and 3; and $\pi_A = \pi_B = 0.5$, $\Delta_A = \Delta_B = 0.25$ for sibpair 2 and 3.

$q_1 = P_A = P_B = 0.50$, LD coefficients $D_{AB} = 0.10$, $D_{AQ} = D_{QB} = 0.15$, and map distances $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM. We can see the following: (1) For both dominant and recessive traits, the required number of sibpairs or tri-sibships is reasonable for test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ if the heritability h^2 is larger than 0.1 (fig. 1A, C); (2) For dominant traits, the required number of sibpairs is less than 150 for each of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ if the heritability h^2 is larger than 0.1 (fig. 1A); the required number of sibpairs of test statistic $F_{AB,ad}$ is similar to that of $F_{AB,a}$, and the required number of sibpairs of test statistic $F_{A,ad}$ is similar to that of $F_{A,a}$; (3) For recessive traits, the required number of tri-sibships is less than 100 for each of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ if the heritability h^2 is larger than 0.15 (fig. 1C); (4) The required number of sibpairs or tri-sibships of test statistics $F_{AB,d}$ and $F_{A,d}$ is much bigger, especially for recessive trait (fig. 1B, D).

Figure 2 shows power curves for the test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against trait frequency allele q_1 and marker allele frequency P_A at 0.01 significant level, when $P_A = 0.5$ (fig. 2A, B), $q_1 = 0.5$ (fig. 2C, D), $P_B = 0.50$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, and $h^2 = 0.25$, for a mode of dominant inheritance $a = d = 1.0$, and a mode of recessive inheritance $a = 1.0$, $d = -0.5$, respectively. The LD coefficients are $D_{AB} = (\min(P_A, P_B) - P_A P_B)/2$, $D_{AQ} = (\min(P_A, q_1) - P_A q_1)/2$ and $D_{QB} = (\min(P_B, q_1) - P_B q_1)/2$. The power of the statistic $F_{AB,ad}$ is lower than that of $F_{AB,a}$, and the power of $F_{A,ad}$ is slightly lower than that of $F_{A,a}$; this is due to the larger degrees of freedom of $F_{AB,ad}$ and $F_{A,ad}$. The power of the statistics $F_{AB,d}$ and $F_{A,d}$ are very low, which confirms the findings in figure 1. Interestingly, the power of statistics $F_{AB,ad}$ and $F_{AB,a}$ depends heavily on the trait allele frequency q_1 (fig. 2A, B), but not so much on the marker allele frequency P_A (fig. 2C, D). The power of the statistics $F_{A,ad}$ and $F_{A,a}$ depends heavily on both the trait allele frequency q_1 and the marker allele frequency P_A .

Figure 3 shows the power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against LD coefficient D_{AQ} at 0.01 significant level, when $q_1 = P_A = P_B = 0.50$, $D_{QB} = 0.15$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, and $h^2 = 0.15$, for a mode of dominant inheritance $a = d = 1.0$, and a mode of recessive inheritance $a = 1.0$, $d = -0.5$, respectively. We can see that the power of $F_{AB,ad}$ and $F_{AB,a}$ is high. In the absence of LD between two markers A and B , the power of $F_{AB,ad}$ and $F_{AB,a}$ is symmetric with $D_{AQ} = 0$ (fig. 3A, B). If LD measure D_{AB} is highly positive (fig. 3C, D; $D_{AB} = 0.10$), the power of $F_{AB,ad}$ and

$F_{AB,a}$ is high for large negative D_{AQ} . If the LD between trait locus Q and marker A is weak ($|D_{AQ}| < 0.10$), the power of $F_{A,ad}$ and $F_{A,a}$ is minimal. Hence, two-marker analysis is advantageous over one-marker analysis. For dominant traits, the power of $F_{AB,d}$ and $F_{A,d}$ is low except for the presence of high LD between trait locus Q and marker A ($|D_{AQ}| > 0.20$, fig. 3A, C). For recessive traits, the power of $F_{AB,d}$ and $F_{A,d}$ is very low (fig. 3B, D).

Figure 4 shows the power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against heritability h^2 at 0.01 significant level, when $q_1 = P_A = P_B = 0.50$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, for a mode of dominant inheritance $a = d = 1.0$, and a mode of recessive inheritance $a = 1.0$, $d = -0.5$, respectively. In the presence of high LD (fig. 4A, B; $D_{AB} = 0.10$, $D_{AQ} = D_{QB} = 0.15$), the power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$, and $F_{A,a}$ is high if the heritability $h^2 \geq 0.15$. If the LD are lower (fig. 4C, D; $D_{AB} = 0.05$, $D_{AQ} = D_{QB} = 0.08$), the power is lower as expected.

Assume that the LD is due to historical mutations at QTL Q which occurred T generations ago. Denote the frequency of haplotype AQ at the generation when the mutations occurred by $P(AQ)(0)$. Then the LD coefficient is $D_{AQ}(0) = P(AQ)(0) - q_1 P_A$ for the generation when the mutations occurred. For the following generations, the disequilibrium coefficient is reduced by a factor $1 - \theta_{AQ}$ in each generation [Hartl and Clark, 1989]. Then the LD coefficient is $D_{AQ}(T) = D_{AQ}(0)(1 - \theta_{AQ})^T$. Similarly, the other LD coefficients are $D_{AB}(T) = D_{AB}(0)(1 - \theta_{AB})^T$ and $D_{QB}(T) = D_{QB}(0)(1 - \theta_{QB})^T$. Figure 5A and B show the power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against position of trait locus Q at 0.01 significant level, when $q_1 = P_A = P_B = 0.50$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 4.5$ cM, and $h^2 = 0.15$, for a mode of dominant inheritance $a = d = 1.0$, and a mode of recessive inheritance $a = 1.0$, $d = -0.5$, respectively. The initial LD coefficients are $D_{AB}(0) = 0.20$, $D_{AQ}(0) = D_{QB}(0) = 0.25$, and the mutation age is $T = 45$. Marker A is located at 0 cM, and marker B is located at 4.5 cM. The power of $F_{AB,ad}$ and $F_{AB,a}$ is similar to the power of $F_{A,ad}$ and $F_{A,a}$, when the trait locus Q is close to marker A (i.e., trait locus Q locates in the region which is less than 1.5 cM from marker A). When trait locus Q locates in the region which is larger than 1.5 cM from marker A , the power of $F_{A,ad}$ and $F_{A,a}$ decrease as the recombination fraction θ_{AQ} increases. The power of $F_{AB,ad}$ and $F_{AB,a}$ is high as long as the trait locus is close to either marker A or marker B . Hence, multiple marker LD mappings have advantages in performing fine gene mappings. Figure 5C and D show the power of test statistics $F_{AB,ad}$ for different mutation ages against the

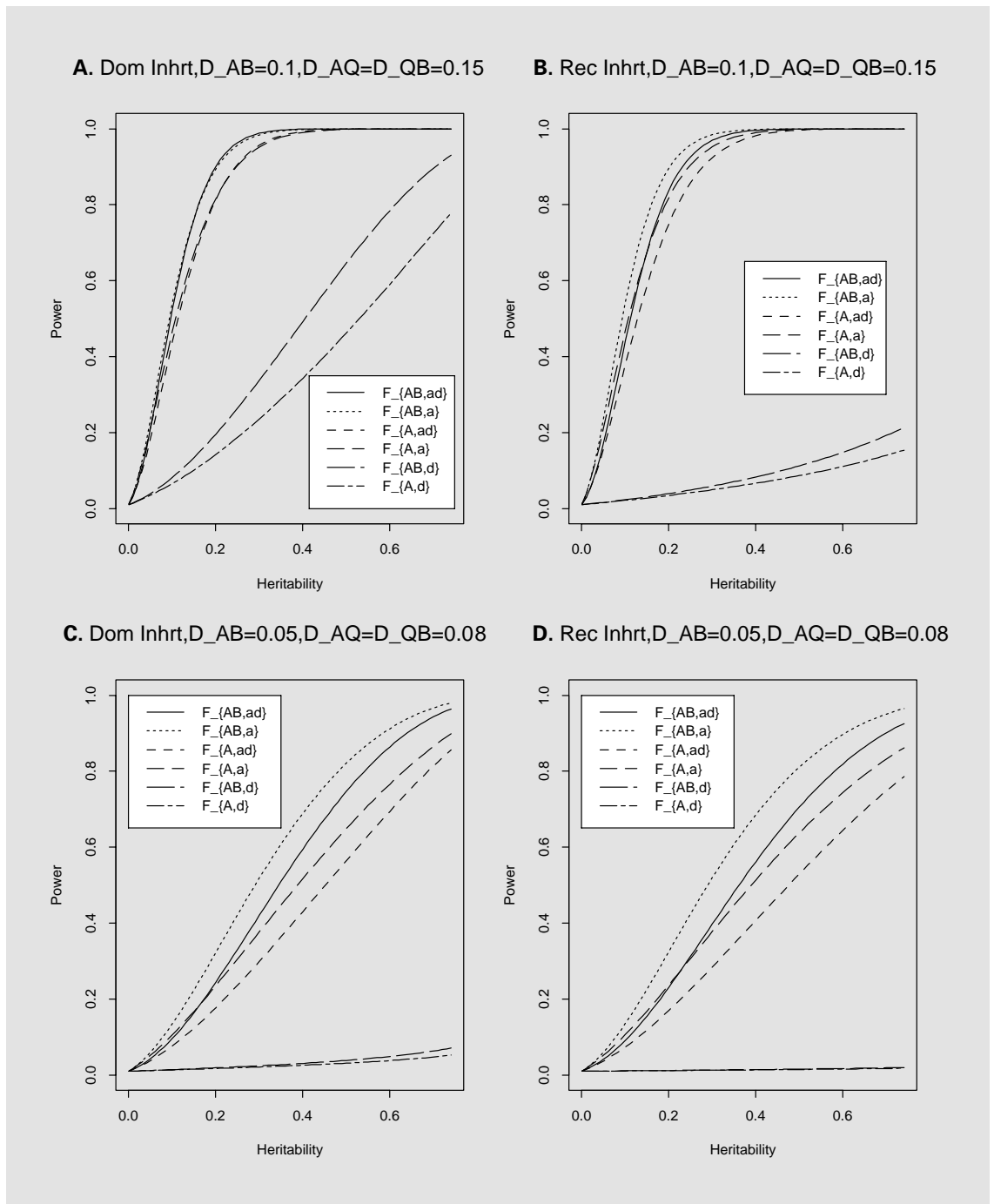


Fig. 4. Power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against heritability h_2 at 0.01 significant level, when $q_1 = P_A = P_B = 0.50$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 5$ cM, $\lambda_{AQ} = \lambda_{QB} = 2.5$ cM, $\sigma_{Ga}^2 = 0.10$, $\sigma_{Gd}^2 = 0.05$, $\sigma_s^2 = 0$, for a mode of dominant inheritance $a = d = 1.0$ (**A** and **C**), and a mode of recessive inheritance $a = 1.0$, $d = -0.5$ (**B** and **D**), respectively. For the $m = 30$ sibpairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For the $k = 20$ tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sibpair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5$, $\Delta_A = 0.25$ for sibpair 1 and 3; and $\pi_A = \pi_B = 0.5$, $\Delta_A = \Delta_B = 0.25$ for sibpair 2 and 3. Dom = dominant, inhrt = inheritance, Rec = recessive.

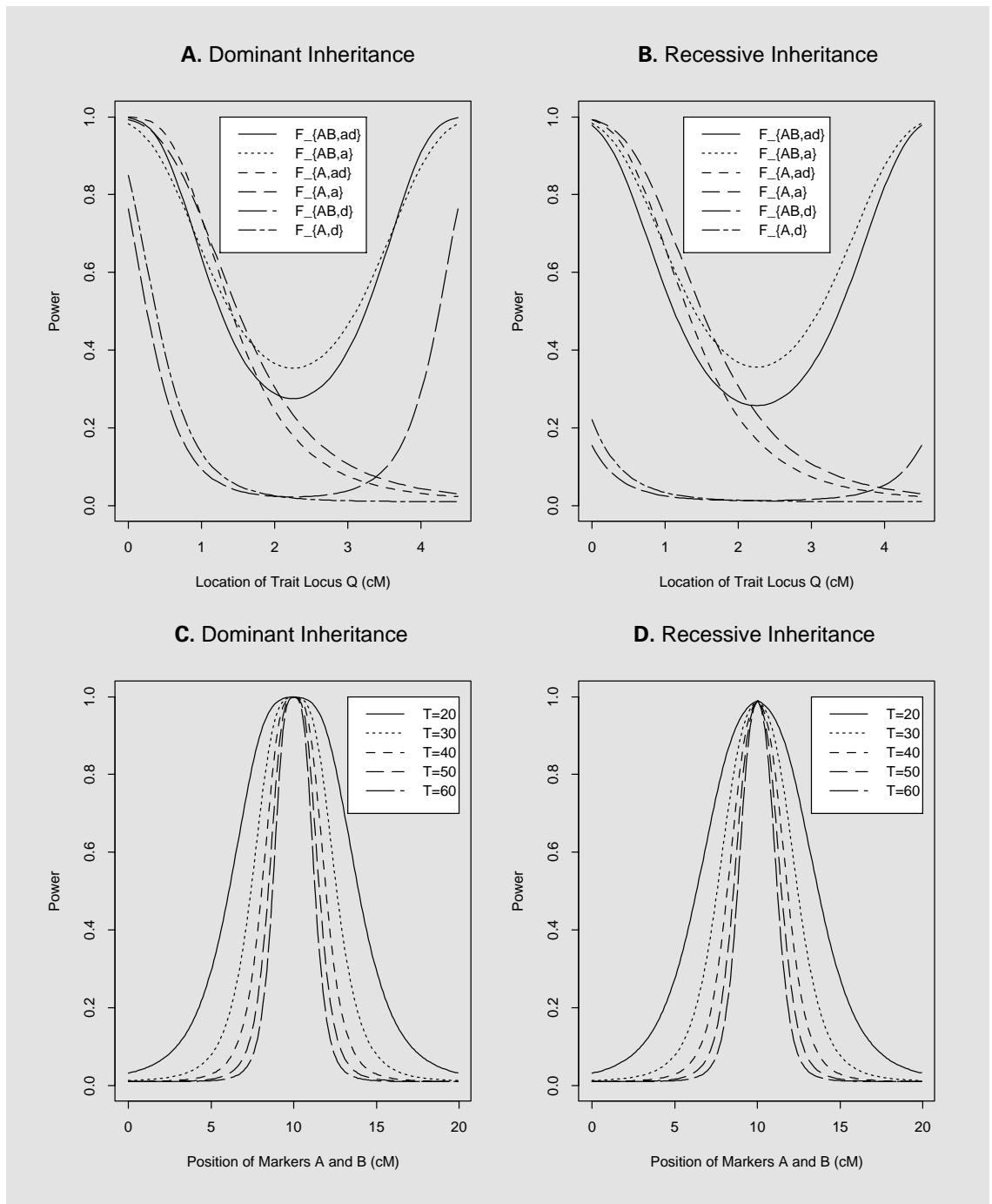


Fig. 5. A, B Power of test statistics $F_{AB, ad}$, $F_{AB, a}$, $F_{AB, d}$, $F_{A, ad}$, $F_{A, a}$, and $F_{A, d}$ against position of trait locus Q at 0.01 significant level, when $q_1 = P_A = P_B = 0.50$, $n = 60$, $m = 30$, $k = 20$, $\lambda_{AB} = 4.5$ cM, $\sigma_{Ga}^2 = 0.10$, $\sigma_{Gd}^2 = 0.05$, $\sigma_s^2 = 0$, and $h_2 = 0.15$, for a mode of dominant inheritance $a = d = 1.0$, and a mode of recessive inheritance $a = 1.0$, $d = -0.5$, respectively. The initial LD coefficients are $D_{AB}(0) = 0.20$, $D_{AQ}(0) = D_{QB}(0) = 0.25$, and the mutation age is $T = 45$. Marker A locates at 0 cM, and marker B locates at 4.5 cM. For the $m = 30$ sibpairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For the $k = 20$ tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sibpair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5$, $\Delta_A = 0.25$ for sibpair 1 and 3; and $\pi_A = \pi_B = 0.5$, $\Delta_A = \Delta_B = 0.25$ for sibpair 2 and 3. **C, D** Power of test statistics $F_{AB, ad}$ of different mutation ages against position of markers A and B at 0.01 significant level. The trait locus Q locates at 10 cM. The two markers A and B flank the trait locus Q ; one marker is on each side of the QTL with equal distance to the QTL. The other parameters are the same as **A** and **B**.

position of markers A and B at 0.01 significant level. In the two graphs, the trait locus Q locates at 10 cM; markers A and B flank the trait locus Q . One marker is on each side of the QTL with equal distance to the QTL. The power decreases quickly when the age of the mutation increases. For a mutation which is 30 generations old, one should expect very low power if the markers locate 2.5 cM away from the QTL.

Power Calculations and Comparisons for Linkage Analysis

To explore the linkage interval mapping and investigate the influence of the dominant variance of the quantitative trait, we take a sample of $m = 250$ sibpairs. Multiplying $\lambda_{linkage, AB}$ of (15) given in Appendix D by m , we calculate the non-centrality parameters for the linkage interval mapping using markers A and B . Assume that the heritability is $h^2 = 0.35$ and the genetic distance is $\lambda_{AB} = 10$ cM. Marker A locates at 0 cM, and marker B locates at 10 cM. Figure 6 gives the power curves of the linkage interval mapping by markers A and B with or without dominant variance against the location of trait locus Q . For a mode of dominant inheritance in figure 6A, we assume $a = d = 1.0$. For a mode of recessive inheritance in figure 6B, we assume $a = 1.0, d = -0.9$. By assuming there is no dominance variance at the putative trait locus Q , we include σ_{ga}^2 but not σ_{gd}^2 in calculating the correlation of sibpairs. The power without dominant variance is apparently less than that with dominant variance. Hence, including both additive and dominant variances in the model has an advantage in linkage mapping. In the presence of dominant variance, one may lose power by excluding it.

An Example

We apply our method to the Genetic Analysis Workshop 12 German asthma data [Meyers et al., 2001]. The data consist of 97 nuclear families, including 415 persons. Seventy-four families have two children, 19 have three children, and 4 have four children. In Wjst et al. [1999], linkage to total serum IgE was tested by the nonparametric statistic of MAPMAKER/SIBS 2.1. On chromosome 1, marker D1S221 at position 146.7 cM and marker D1S502 at position 151.2 cM are shown to be linked with IGE level. By the method proposed in this paper, we find that dominant variance of log(IgE) is significantly higher than 0 at position 149.85 cM ($p = 0.01$). On this basis, we treat allele 8 at marker D1S221 as allele A , and collapse other alleles as allele a . At marker D1S502, we collapse alleles 7, 8, and 13 as allele B , and others as allele b . Then,

we find that covariate Z_A is significantly different from 0 at position 149.85 cM ($\hat{\delta}_A = 1.16$, with $p = 0.0475$ by LRT and $p = 0.0484$ by F test). Hence, we are able to confirm the result of Wjst et al. [1999], and find that marker D1S221 is associated with log(IgE).

Discussion

In this paper, we explore models in high resolution joint LD and linkage mapping of QTL based on sibship data. Variance component models are proposed to perform joint analysis of LD and linkage mapping of QTL by taking into account the sibship variance-covariance structure. The models simultaneously incorporate both LD and linkage information. The LD information is contained in the mean coefficients of sibship data which include population data. The linkage information is contained in the variance-covariance matrix of trait values of sibships with at least two siblings. We generalize the linear model of high resolution LD mapping method of Fan and Xiong [2002] for population data to sibship data. The motivation is to build valid models for late-onset complex disease gene mapping when parental data are not available. Using nuclear family data and population data, Fan and Xiong [2003] have built models in combined high resolution LD and linkage mapping of QTL. Population data and any type of pedigree data can be combined together to perform a joint analysis of high resolution LD and linkage mapping of QTL. Here pedigree data may consist of nuclear families, sibships, father/mother-child pairs, uncle/aunt-niece/nephew pairs, and multi-generation pedigrees.

For qualitative traits, transmission disequilibrium tests (TDT) can be used in fine disease gene association studies in the presence of linkage. In this paper, we show that our models can be used in fine association studies between QTL and markers in the presence of linkage for complex diseases. This is because the mean coefficients (4) simultaneously account for both the LD and the effects of the putative QTL Q . The parameters of LD (i.e., D_{AQ} and D_{QB}) and gene effect (i.e., α_Q and δ_Q) are contained in the mean coefficients. In the presence of linkage to a particular chromosome region, testing association between trait locus and markers in the region is based on equations (4). By power and sample size comparisons, we show the merit of 6 test statistics of LD mapping. From the non-centrality parameter formulas of these 6 statistics, we see that the parameters to test additive effects α_A and α_B depend on a factor of R_{AQ}^2 and R_{QB}^2 . The parameters to

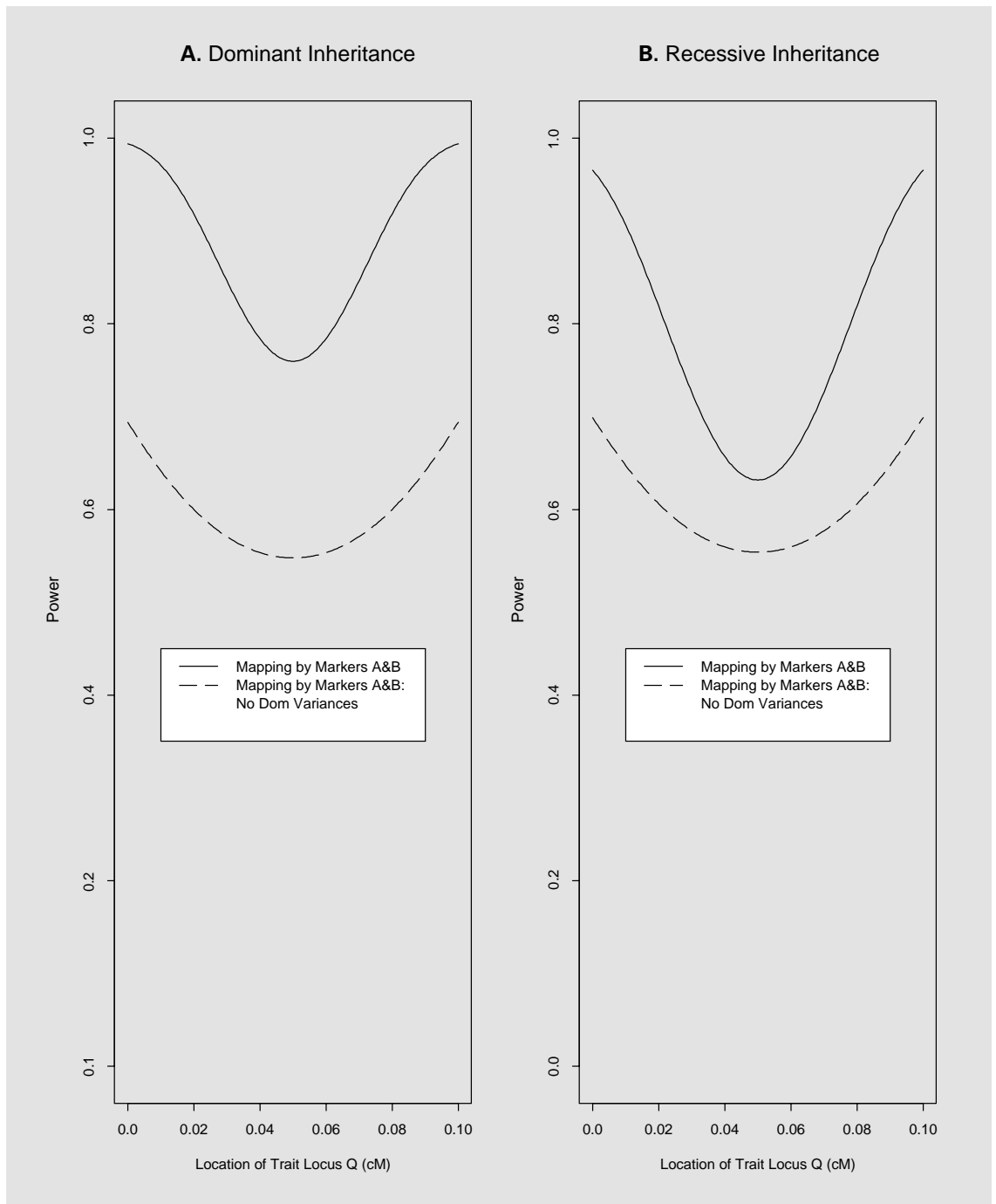


Fig. 6. Power curves of the interval mapping by markers *A* and *B* with or without dominant variances against the recombination fraction θ_{AQ} at 0.05 significant level, when $h^2 = 0.35$, $\lambda_{AB} = 10$ cM, $m = 250$, $\sigma_{Ga}^2 = 0.10$, $\sigma_{Gd}^2 = 0.05$, $\sigma_s^2 = 0$, for a dominant trait $a = d = 1.0$, $q_1 = 0.60$ (graph **A**); and a recessive trait $a = 1.0$, $d = -0.9$, $q_1 = 0.40$ (graph **B**). Marker *A* locates at 0 cM, and marker *B* locates at 10 cM.

test dominant effects δ_A and δ_B depend on a factor of R_{AQ}^4 and R_{QB}^4 . Therefore, the power of test statistics $F_{AB,ad}$, $F_{AB,a}$, $F_{A,ad}$ and $F_{A,a}$ is generally higher than that of $F_{AB,d}$ and $F_{A,d}$. Moreover, the power of $F_{AB,ad}$ can be lower than that of $F_{AB,a}$, and the power of $F_{A,ad}$ can be lower than that of $F_{A,a}$. This is due to the increase in degrees of freedom of the statistics $F_{AB,ad}$ and $F_{A,ad}$ with minimal contribution from the dominant effects to the non-centrality parameters. By theoretical and simulation study, we show that either $F_{A,a}$ or LTR of current paper is more powerful than any of between-pairs and within-pairs approaches of Fulker and Abecasis et al. 'AbAw' approach, if only one marker A is used in analysis. Moreover, the method is applied to Gaw12 German asthma data.

In recent years, there has been great interest in the LD mapping of QTL. Abecasis et al. [2000, 2001], Fulker et al. [1999] and Sham et al. [2000] explore linkage analysis and association studies of quantitative traits by variance-component procedures allowing a simultaneous test of allelic association for family data. Allison [1997] proposes test statistics for association study of QTL. George et al. [1999] propose a TDT for quantitative traits in pedigree data by multiple regression. Monks and Kaplan [2000] discuss removing sampling restrictions from association tests for QTL. Rabinowitz [1997] performs a simulation association study of QTL. Zhang and Zhao [2001] propose a quantitative similarity-based test to identify association between a bi-allelic marker and a quantitative trait. Zhao et al. [2001] apply a regression approach of LD mapping to localize QTL in humans. In these studies, most investigators use only one bi-allelic marker in their analysis. However, very dense maps such as SNPs and high resolution microsatellite markers are available for the human genome. These exciting developments allow us to explore models and methodologies of simultaneously using two or more markers in high resolution LD mapping of QTL. It is urgent to work out powerful statistical analyses and methods for high resolution joint LD and linkage mapping of QTL based on multiple markers. It is our hope that the current research may shed some light on this very important field and stimulate more interest for future investigation.

To date, linkage analysis of QTL is well studied. Using one marker, Haseman and Elston [1972] propose using sibpairs to perform a linkage analysis of quantitative traits by regressing the trait difference onto the proportion of marker alleles shared IBD by a sibpair. To improve the performance of sibpair analysis by using all inheritance information in general pedigrees, variance component linkage analysis has been developed [Amos, 1994; Gold-

gar, 1990; Goldgar and Oniki, 1992; Schork, 1993]. Using two flanking markers, Fulker and Cardon [1994] extend the method to a simple and neat interval mapping approach of sibpair analysis. Interval mapping is an important method to locate QTL; it has an advantage in determining the exact location of QTL [Fulker and Cardon, 1994; Haley and Knott; Jansen, 1993; Lander and Botstein, 1989; Xu and Atchley, 1995]. Fulker et al. [1995] further develop the approach to perform multipoint interval mapping of QTL using sibpairs. Almasy and Blangero [1998] develop a multipoint interval mapping method for general pedigrees using the framework of variance component linkage analysis. Pratt et al. [2000] propose to include both additive and dominant variances to calculate the trait covariance of relatives in exact multipoint quantitative trait linkage analysis. We calculate the probability of sharing two trait alleles IBD for sibpairs conditional on the information of flanking markers for interval mapping of QTL. Using the formulas in both Fulker and Cardon [1994] and in this paper, we can calculate the trait covariance of sibpairs which include both additive and dominant genetic variances. Due to the simplicity of the methods, it is easy to put the formulas in genetic software. By numerical calculations, we show that the formulas for calculating the probability of sharing two trait alleles IBD for sibs, conditional on the information of two flanking markers, give satisfactory results. By power calculations and comparisons, we notice that the powers of models excluding dominant variances are smaller than those for models including dominant variances. Hence, including both additive and dominant variances in the models has an advantage in linkage interval mapping. In the presence of dominant variances, power may be lost by excluding them from the models. Since linkage interval mapping is well-studied in the literature, we only include a figure in the results.

In this paper, two biallelic markers are used in joint LD and linkage mapping of QTL. It would be interesting to generalize the method toward in several ways. First, we consider using multiple bi-allelic markers such as SNPs in the analysis. Since the number of LD measures and degrees of freedom of test statistics can be large when using multiple markers in analysis, selecting important ones in analysis is critical. Moreover, using higher order LD coefficients for multiple loci quickly becomes very cumbersome and only considering pairwise LD coefficients discards information and ignores dependencies among multiple pairwise marker comparisons due to shared recombination events. Hence, one needs to be careful in dealing with the potential problems. We believe

that it is worthwhile to generalize the method to use 3 or 4 relevant markers. Second, multiallelic markers such as microsatellites or haplotype blocks can be used in building models. Since LD mapping is influenced very heavily by population stratification, more investigations are needed to gain a better understanding of the effect of the population subdivisions and admixtures on the analysis of joint LD and linkage mapping [Zhao and Xiong, 2002]. Although the proposed methods can be applied to general pedigrees, the details of these applications have not been worked out. For large pedigrees, computation can be very challenging, especially when population data and pedigree data are combined in one analysis. Finally, selection of markers in analysis and genotyping errors can raise problems, which deserve more research.

Acknowledgments

We would like to thank the two reviewers for their very insight comments to improve the paper. We are grateful to Dr. Gordon and Dr. Montague for their patience in handling the manuscript. We would also like to thank Ms. Christie Spinka for reading the manuscript carefully and providing numerous suggestions in improving the writings and the grammar; and Dr. Abecasis for kindly providing the simulation program Ldsimul to generate simulated datasets.

Appendix A

To simplify notations, we omit subscripts ij from Δ_{ijQ} , π_{ijA} , π_{ijB} , Δ_{ijA} , Δ_{ijB} in the following appendices A, B, and C. Taking the variance-covariance for equation (6), we have the following matrix equation to calculate the coefficients

$$\begin{aligned} & \text{Cov} \begin{pmatrix} (\pi_A, \pi_A) & (\pi_B, \pi_A) & (\Delta_A, \pi_A) & (\Delta_B, \pi_A) \\ (\pi_A, \pi_B) & (\pi_B, \pi_B) & (\Delta_A, \pi_B) & (\Delta_B, \pi_B) \\ (\pi_A, \Delta_A) & (\pi_B, \Delta_A) & (\Delta_A, \Delta_A) & (\Delta_B, \Delta_A) \\ (\pi_A, \Delta_B) & (\pi_B, \Delta_B) & (\Delta_A, \Delta_B) & (\Delta_B, \Delta_B) \end{pmatrix} \begin{pmatrix} \beta_A \\ \beta_B \\ r_A \\ r_B \end{pmatrix} \\ &= \text{Cov} \begin{pmatrix} (\Delta_Q, \pi_A) \\ (\Delta_Q, \pi_B) \\ (\Delta_Q, \Delta_A) \\ (\Delta_Q, \Delta_B) \end{pmatrix}. \end{aligned} \quad (9)$$

From Elston and Keats [1985] and Almasy and Blangero [1998], we have the following

$$\begin{aligned} \text{Cov}(\pi_A, \pi_A) &= \text{Cov}(\pi_B, \pi_B) = 1/8, \text{Cov}(\pi_B, \pi_A) = (1 - 2\theta_{AB})^2/8, \\ \text{Cov}(\Delta_A, \Delta_A) &= \text{Cov}(\Delta_B, \Delta_B) = \frac{3}{16}, \text{Cov}(\Delta_B, \Delta_A) = \frac{3}{16} \rho(\Delta_A, \Delta_B), \\ \text{Cov}(\Delta_A, \Delta_Q) &= \frac{3}{16} \rho(\Delta_A, \Delta_Q), \text{Cov}(\Delta_Q, \Delta_B) = \frac{3}{16} \rho(\Delta_Q, \Delta_B), \end{aligned}$$

where

$$\rho(\Delta_i, \Delta_j) = 1 - \frac{16}{3} \theta_{ij} + \frac{32}{3} \theta_{ij}^2 - \frac{32}{3} \theta_{ij}^3 + \frac{16}{3} \theta_{ij}^4.$$

In Appendix B, we will show that

$$\text{Cov}(\Delta_A, \pi_A) = \text{Cov}(\Delta_B, \pi_B) = 1/8,$$

$$\text{Cov}(\Delta_B, \pi_A) = \text{Cov}(\Delta_A, \pi_B) = (1 - 2\theta_{AB})^2/8,$$

$$\text{Cov}(\Delta_Q, \pi_A) = (1 - 2\theta_{AQ})^2/8, \text{Cov}(\Delta_Q, \pi_B) = (1 - 2\theta_{QB})^2/8. \quad (10)$$

Plugging the above results into the equation (9), we have a sub-matrix block equation

$$\begin{pmatrix} A & A \\ A & B \end{pmatrix} \begin{pmatrix} \beta_A \\ \beta_B \\ r_A \\ r_B \end{pmatrix} = \begin{pmatrix} (1 - 2\theta_{AQ})^2 \\ (1 - 2\theta_{QB})^2 \\ 3\rho(\Delta_A, \Delta_Q)/2 \\ 3\rho(\Delta_Q, \Delta_B)/2 \end{pmatrix}$$

where

$$\begin{aligned} A &= \begin{pmatrix} 1 & (1 - 2\theta_{AB})^2 \\ (1 - 2\theta_{AB})^2 & 1 \end{pmatrix}, \\ B &= \frac{3}{2} \begin{pmatrix} 1 & \rho(\Delta_A, \Delta_B) \\ \rho(\Delta_A, \Delta_B) & 1 \end{pmatrix}. \end{aligned}$$

Therefore, we have from Harville [1997]

$$\begin{aligned} \begin{pmatrix} \beta_A \\ \beta_B \\ r_A \\ r_B \end{pmatrix} &= \begin{pmatrix} A & A \\ A & B \end{pmatrix}^{-1} \begin{pmatrix} (1 - 2\theta_{AQ})^2 \\ (1 - 2\theta_{QB})^2 \\ 3\rho(\Delta_A, \Delta_Q)/2 \\ 3\rho(\Delta_Q, \Delta_B)/2 \end{pmatrix} \\ &= \begin{pmatrix} A^{-1} + (B - A)^{-1} & -(B - A)^{-1} \\ -(B - A)^{-1} & (B - A)^{-1} \end{pmatrix} \begin{pmatrix} (1 - 2\theta_{AQ})^2 \\ (1 - 2\theta_{QB})^2 \\ 3\rho(\Delta_A, \Delta_Q)/2 \\ 3\rho(\Delta_Q, \Delta_B)/2 \end{pmatrix}. \end{aligned}$$

The equation

$$\begin{aligned} 3\rho(\Delta_i, \Delta_j)/2 - (1 - 2\theta_{ij})^2 &= (1 - 8\theta_{ij} + 24\theta_{ij}^2 - 32\theta_{ij}^3 + 16\theta_{ij}^4)/2 \\ &= (1 - 2\theta_{ij})^4/2 \end{aligned}$$

leads to

$$\begin{aligned} \begin{pmatrix} r_A \\ r_B \end{pmatrix} &= (B - A)^{-1} \begin{pmatrix} 3\rho(\Delta_A, \Delta_Q)/2 - (1 - 2\theta_{AQ})^2 \\ 3\rho(\Delta_Q, \Delta_B)/2 - (1 - 2\theta_{QB})^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{(1 - 2\theta_{AB})^4}{2} \\ \frac{(1 - 2\theta_{AB})^4}{2} & \frac{1}{2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{(1 - 2\theta_{AQ})^4}{2} \\ \frac{(1 - 2\theta_{QB})^4}{2} \end{pmatrix} \\ &= \frac{1}{1 - (1 - 2\theta_{AB})^8} \begin{pmatrix} (1 - 2\theta_{AQ})^4 - (1 - 2\theta_{QB})^4 (1 - 2\theta_{AB})^4 \\ (1 - 2\theta_{QB})^4 - (1 - 2\theta_{AQ})^4 (1 - 2\theta_{AB})^4 \end{pmatrix}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} &= A^{-1} \begin{pmatrix} (1 - 2\theta_{AQ})^2 \\ (1 - 2\theta_{QB})^2 \end{pmatrix} - (B - A)^{-1} \begin{pmatrix} 3\rho(\Delta_A, \Delta_Q)/2 - (1 - 2\theta_{AQ})^2 \\ 3\rho(\Delta_Q, \Delta_B)/2 - (1 - 2\theta_{QB})^2 \end{pmatrix} \\ &= \begin{pmatrix} \beta_{\pi A} \\ \beta_{\pi B} \end{pmatrix} - \begin{pmatrix} r_A \\ r_B \end{pmatrix}. \end{aligned}$$

Hence, we have shown the first four coefficients in (7) are valid.

Appendix B

Consider a sibpair with trait values y_i and y_j . First, we have the following equation from Haseman [1970] [also see Amos, 1994, equation (5) on p 537 or Amos et al., 1989, p 437]:

$$\begin{aligned} \text{Cov}(y_i, y_j | \pi_A, \Delta_A) = & \\ \frac{1}{2} \sigma_{Ga}^2 + \frac{1}{4} \sigma_{Gd}^2 + \sigma_s^2 + (1 - \psi_A) \sigma_g^2 + \psi_A (\psi_A - 1) \sigma_{gd}^2 & \\ + [-(1 - 2\psi_A) \sigma_g^2 - (1 - 2\psi_A)^2 \sigma_{gd}^2] \pi_A + (1 - 2\psi_A)^2 \sigma_{gd}^2 \Delta_A. & \end{aligned}$$

Comparing the above equation with

$$\text{Cov}(y_i, y_j | \pi_A, \Delta_A) = \pi_Q \sigma_{ga}^2 + \Delta_Q \sigma_{gd}^2 + 1/2 \sigma_{Ga}^2 + 1/4 \sigma_{Gd}^2 + \sigma_s^2,$$

we find

$$\Delta_Q = (1 - \psi_A)^2 - [(1 - 2\psi_A) + (1 - 2\psi_A)^2] \pi_A + (1 - 2\psi_A)^2 \Delta_A. \quad (11)$$

Taking covariances on both sides of above equation with Δ_A , we get

$$\begin{aligned} \text{Cov}(\Delta_Q, \Delta_A) = & -[(1 - 2\psi_A) + (1 - 2\psi_A)^2] \text{Cov}(\pi_A, \Delta_A) \\ & + (1 - 2\psi_A)^2 \text{Cov}(\Delta_A, \Delta_A). \end{aligned}$$

Replacing $\text{Cov}(\Delta_Q, \Delta_A) = 3/16 \rho(\Delta_A, \Delta_Q)$ and $\text{Cov}(\Delta_A, \Delta_A) = 3/16$ in the above equation [Almasy and Blangero, 1998], we find that $\text{Cov}(\Delta_A, \pi_A) = 1/8$. Then taking covariance of both sides of equation (11) with π_A , we find

$$\begin{aligned} \text{Cov}(\Delta_Q, \pi_A) = & -[(1 - 2\psi_A) + (1 - 2\psi_A)^2] \text{Var}(\pi_A) \\ & + (1 - 2\psi_A)^2 \text{Cov}(\Delta_A, \pi_A) = -(1 - 2\psi_A)/8 = (1 - 2\theta_{AQ})^2/8. \end{aligned}$$

Similarly, we can show the other equations in (10).

Appendix C

To calculate the intercept α in (7), we consider the joint distribution of π_Q , π_A and π_B for a sibpair. Assume that there is no interference for disjoint chromosome regions. Then

$$\begin{aligned} P(\pi_{iA} = i_A, \pi_{iQ} = i_Q, \pi_{iB} = i_B) & \\ = P(\pi_{iA} = i_A, \pi_{iQ} = i_Q) P(\pi_{iB} = i_B | \pi_{iA} = i_A, \pi_{iQ} = i_Q) & \\ = P(\pi_{iA} = i_A | \pi_{iQ} = i_Q) P(\pi_{iQ} = i_Q) P(\pi_{iB} = i_B | \pi_{iQ} = i_Q). & \quad (12) \end{aligned}$$

From Haseman and Elston [1972], table 4, we construct the joint distribution of π_{iQ} , π_{iA} and π_{iB} by equation (12); the results are presented in table 1. Consider a sibpair with trait values y_i and y_j . Then from table 1 we have

$$\begin{aligned} \text{Cov}(y_i, y_j | \pi_A = 0, \pi_B = 0) = & \left[\frac{1}{2} \sigma_{Ga}^2 + \frac{1}{4} \sigma_{Gd}^2 + \sigma_s^2 \right] \\ = (\sigma_{ga}^2 + \sigma_{gd}^2) P(\pi_Q = 1 | \pi_A = 0, \pi_B = 0) & \\ + \frac{\sigma_{ga}^2}{2} P(\pi_Q = 1/2 | \pi_A = 0, \pi_B = 0) & \\ = \frac{(1 - \psi_A)(1 - \psi_B)}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} \sigma_{ga}^2 + \frac{(1 - \psi_A)^2(1 - \psi_B)^2 \sigma_{gd}^2}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)]^2}. & \end{aligned}$$

Therefore, we have the intercept α in (7) since it is the coefficient of σ_{gd}^2 in above equation.

Appendix D

Consider a sibship of l children. Under the null hypothesis of no linkage between the trait locus and the markers, the correlation of each sibpair is

$$\rho = \frac{\sigma_{ga}^2 \sigma_{gd}^2}{2\sigma^2 4\sigma^2} + \frac{\sigma_{Ga}^2}{2\sigma^2} + \frac{\sigma_{Gd}^2}{4\sigma^2} + \frac{\sigma_s^2}{\sigma^2}.$$

Hence, we have twice the expected log-likelihood

$$\begin{aligned} E(2L_{Null}) = & -l - l \log[2\pi\sigma^2] - \log \det \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \dots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix} \\ = & -l - l \log[2\pi\sigma^2] - \log[(1 + (l-1)\rho)(1-\rho)^{l-1}] \end{aligned}$$

Under the alternative hypothesis of linkage between the trait locus and marker A , the correlation between the sibpair i and j is $C_{2\pi_{ijA}}$ given by

$$\begin{aligned} C_k = & \text{Cov}(y_i, y_j | \pi_{iA} = k/2) / \sigma^2 \\ = & (\sigma_{ga}^2 + \sigma_{gd}^2) P(\pi_{iQ} = 1 | \pi_{iA} = k/2) / \sigma^2 \\ & + \frac{\sigma_{ga}^2}{2} P(\pi_{iQ} = 1/2 | \pi_{iA} = k/2) / \sigma^2 \\ & + [\sigma_{Ga}^2/2 + \sigma_{Gd}^2/4 + \sigma_s^2] / \sigma^2, \quad k = 0, 1, 2. \end{aligned}$$

From Haseman and Elston [1972], table 4, or Sham et al. [2000], table 1, we have

$$\begin{aligned} C_2 = & [(\sigma_{ga}^2 + \sigma_{gd}^2) \psi_A^2 + \sigma_{ga}^2 \psi_A (1 - \psi_A) + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4 + \sigma_s^2] / \sigma^2 \\ C_1 = & \{(\sigma_{ga}^2 + \sigma_{gd}^2) \psi_A (1 - \psi_A) + \sigma_{ga}^2 [1 - 2\psi_A (1 - \psi_A)] / 2 \\ & + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4 + \sigma_s^2\} / \sigma^2 \\ C_0 = & [(\sigma_{ga}^2 + \sigma_{gd}^2) (1 - \psi_A)^2 + \sigma_{ga}^2 \psi_A (1 - \psi_A) + \sigma_{Ga}^2/2 + \sigma_{Gd}^2/4 + \sigma_s^2] / \sigma^2. \end{aligned}$$

We have twice the expected log-likelihood under the alternative hypothesis of linkage

$$\begin{aligned} E(2L_{random, A}) = & -l - l \log[2\pi\sigma^2] \\ - \sum_{\pi_{12A}} \dots \sum_{\pi_{l-1, lA}} & P(\pi_{12A}) \dots P(\pi_{l-1, lA}) \log \det \begin{pmatrix} 1 & C_{2\pi_{12A}} & \dots & C_{2\pi_{1lA}} \\ C_{2\pi_{21A}} & 1 & \dots & C_{2\pi_{2lA}} \\ \vdots & \vdots & \dots & \vdots \\ C_{2\pi_{l1A}} & C_{2\pi_{l2A}} & \dots & 1 \end{pmatrix} \end{aligned}$$

where $P(\pi_{iA} = 0) = P(\pi_{iA} = 1) = 1/4$ and $P(\pi_{iA} = 1/2) = 1/2$. From Stuart and Ord [1991], the non-centrality parameter for linkage of the family is equal to $\lambda_{linkage, A} = E(2L_{random, A}) - E(2L_{Null})$. If the sibship consists of two offspring, then

$$\lambda_{linkage, A} = \log[1 - \rho^2] - \sum_{K=0}^2 P(\pi_{12A} = k/2) \log[1 - C_k^2]. \quad (13)$$

Under the alternative hypothesis of linkage between the trait locus and markers A and B , the correlation between the sibpair i and j is $C_{2\pi_{ijA}, 2\pi_{ijB}}$ given by

$$\begin{aligned}
C_{k_1 k_2} &= \text{Cov}(y_i, y_j | \pi_{ijA} = k_1/2, \pi_{ijB} = k_2/2) / \sigma^2 \\
&= \left[(\sigma_{ga}^2 + \sigma_{gd}^2) P(\pi_{ijQ} = 1 | \pi_{ijA} = k_1/2, \pi_{ijB} = k_2/2) \right. \\
&\quad + \frac{\sigma_{ga}^2}{2} P(\pi_{ijQ} = 1/2 | \pi_{ijA} = k_1/2, \pi_{ijB} = k_2/2) + \sigma_{Gd}^2/2 \\
&\quad \left. + \sigma_{Gs}^2/4 + \sigma_s^2 \right] / \sigma^2. \tag{14}
\end{aligned}$$

To calculate the quantities $C_{k_1 k_2}$, we need the joint distribution of π_{ijA} , π_{ijQ} and π_{ijB} of a sibpair i and j under the alternative hypothesis of linkage. Based on table 1, we can calculate C_{ij} , $i, j = 0, 1, 2$, which are given in Appendix E. We have twice the expected log-likelihood under the alternative hypothesis of linkage

$$\begin{aligned}
E(2L_{random, AB}) &= -l - l \log[2\pi\sigma^2] \\
&- \sum_{\pi_{12A}} \sum_{\pi_{12B}} \dots \sum_{\pi_{l-1, lA}} \sum_{\pi_{l-1, lB}} P(\pi_{12A}) P(\pi_{12B}) \dots P(\pi_{l-1, lA}) P(\pi_{l-1, lB}) \\
\log \det &\begin{pmatrix} 1 & C_{2\pi_{12A}, 2\pi_{12B}} & \dots & C_{2\pi_{1lA}, 2\pi_{1lB}} \\ C_{2\pi_{21A}, 2\pi_{21B}} & 1 & \dots & C_{2\pi_{2lA}, 2\pi_{2lB}} \\ \vdots & \vdots & \dots & \vdots \\ C_{2\pi_{l1A}, 2\pi_{l1B}} & C_{2\pi_{l2A}, 2\pi_{l2B}} & \dots & 1 \end{pmatrix}
\end{aligned}$$

where $P(\pi_{ijB} = 0) = P(\pi_{ijB} = 1) = 1/4$ and $P(\pi_{ijB} = 1/2) = 1/2$. From Stuart and Ord [1991], the non-centrality parameter for linkage of the sibship is equal to $\lambda_{linkage, AB} = E(2L_{random, AB}) - E(2L_{Null})$. If the sibship consists of two offspring, then

$$\begin{aligned}
\lambda_{linkage, AB} \\
&= \log[1 - \rho^2] - \sum_{i, j=0}^2 P(\pi_{12A} = i/2) P(\pi_{12B} = j/2) \log[1 - C_{ij}^2]. \tag{15}
\end{aligned}$$

Appendix E

For simplicity, let us assume $\sigma^2 = 1$. From table 1 and equation (14), we may calculate

$$\begin{aligned}
C_{22} &= \sigma_{ga}^2 \frac{\psi_A \psi_B}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} + \sigma_{gd}^2 \frac{\psi_A^2 \psi_B^2}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)]^2} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{21} &= \frac{\sigma_{ga}^2}{2} \left[\frac{\psi_A \psi_B}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} + \frac{\psi_A(1 - \psi_B)}{\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B} \right] \\
&\quad + \sigma_{gd}^2 \frac{\psi_A^2 \psi_B(1 - \psi_B)}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)] [\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{20} &= \sigma_{ga}^2 \frac{\psi_A(1 - \psi_B)}{\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B} + \sigma_{gd}^2 \frac{\psi_A^2(1 - \psi_B)^2}{[\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]^2} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2
\end{aligned}$$

$$\begin{aligned}
C_{12} &= \frac{\sigma_{ga}^2}{2} \left[\frac{\psi_A \psi_B}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} + \frac{(1 - \psi_A)\psi_B}{\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B} \right] \\
&\quad + \sigma_{gd}^2 \frac{\psi_A(1 - \psi_A)\psi_B^2}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)] [\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{11} &= \frac{\sigma_{ga}^2}{2} + \sigma_{gd}^2 \frac{2\psi_A(1 - \psi_A)\psi_B(1 - \psi_B)}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)]^2 + [\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]^2} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{10} &= \frac{\sigma_{ga}^2}{2} \left[\frac{(1 - \psi_A)(1 - \psi_B)}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} + \frac{\psi_A(1 - \psi_B)}{\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B} \right] \\
&\quad + \sigma_{gd}^2 \frac{\psi_A(1 - \psi_A)(1 - \psi_B)^2}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)] [\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{02} &= \sigma_{ga}^2 \frac{(1 - \psi_A)\psi_B}{\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B} + \sigma_{gd}^2 \frac{(1 - \psi_A)^2 \psi_B^2}{[\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]^2} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{01} &= \frac{\sigma_{ga}^2}{2} \left[\frac{(1 - \psi_A)(1 - \psi_B)}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} + \frac{(1 - \psi_A)\psi_B}{\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B} \right] \\
&\quad + \sigma_{gd}^2 \frac{(1 - \psi_A)^2 \psi_B(1 - \psi_B)}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)] [\psi_A(1 - \psi_B) + (1 - \psi_A)\psi_B]} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2 \\
C_{00} &= \sigma_{ga}^2 \frac{(1 - \psi_A)(1 - \psi_B)}{\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)} + \sigma_{gd}^2 \frac{(1 - \psi_A)^2(1 - \psi_B)^2}{[\psi_A \psi_B + (1 - \psi_A)(1 - \psi_B)]^2} \\
&\quad + \frac{\sigma_{Ga}^2}{2} + \frac{\sigma_{Gd}^2}{4} + \sigma_s^2.
\end{aligned}$$

Appendix F

For each y_i of the n individuals, $\Sigma_i = \sigma^2$ and

$$X_i = (1 \quad x_{Ai} \quad x_{Bi} \quad z_{Ai} \quad z_{Bi}), \quad i = 1, 2, \dots, n.$$

From formulas in Fan and Xiong [2002], Appendix A, we show that

$$\frac{1}{n} \sum_{i=1}^n X_i^T \Sigma_i^{-1} X_i = \frac{1}{n\sigma^2} \sum_{i=1}^n X_i^T X_i \approx \frac{1}{\sigma^2} \text{diag}(1, V_A, V_D), \tag{16}$$

where V_A and V_D are additive and dominant variance-covariance matrices of (3). For each of the m sibpairs, the variance-covariance matrix

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}$$

and the model matrix

$$X_i = \begin{pmatrix} 1 & x_{A1}^{(i)} & x_{B1}^{(i)} & z_{A1}^{(i)} & z_{B1}^{(i)} \\ 1 & x_{A2}^{(i)} & x_{B2}^{(i)} & z_{A2}^{(i)} & z_{B2}^{(i)} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}$$

$i = n + 1, 2, \dots, n + m$.

Notice

$$\Sigma_i^{-1} = [\sigma^{-2}/(1 - \rho_{12}^2)] \begin{pmatrix} 1 & -\rho_{12} \\ -\rho_{12} & 1 \end{pmatrix}.$$

From Fan and Xiong [2003], Appendix C, we have $E[X_{i1}^T X_{i2}] = E[X_{i2}^T X_{i1}] = \text{diag}(1, V_A/2, V_D/4)$. By above formulas and the formulas in Fan and Xiong [2002], Appendix A, we have the following

$$\frac{1}{m} \sum_{i=n+1}^{n+m} X_i^T \Sigma_i^{-1} X_i \approx \frac{2}{(1 - \rho_{12}^2)\sigma^2} [\text{diag}(1, V_A, V_D) - \rho_{12} \text{diag}(1, V_A/2, V_D/4)]. \quad (17)$$

For each of the k tri-sibships, the variance-covariance matrix

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

and the model matrix

$$X_i = \begin{pmatrix} 1 & x_{A1}^{(i)} & x_{B1}^{(i)} & z_{A1}^{(i)} & z_{B1}^{(i)} \\ 1 & x_{A2}^{(i)} & x_{B2}^{(i)} & z_{A2}^{(i)} & z_{B2}^{(i)} \\ 1 & x_{A3}^{(i)} & x_{B3}^{(i)} & z_{A3}^{(i)} & z_{B3}^{(i)} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \end{pmatrix}$$

$i = n + m + 1, 2, \dots, n + m + k$.

Notice

$$\Sigma_i^{-1} = [\sigma^{-2}/C_3] \begin{pmatrix} 1 - \rho_{23}^2 & \rho_{13}\rho_{23} - \rho_{12} & \rho_{12}\rho_{23} - \rho_{13} \\ \rho_{13}\rho_{23} - \rho_{12} & 1 - \rho_{13}^2 & \rho_{12}\rho_{13} - \rho_{23} \\ \rho_{12}\rho_{23} - \rho_{13} & \rho_{12}\rho_{13} - \rho_{23} & 1 - \rho_{12}^2 \end{pmatrix}$$

where $C_3 = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}$. From Fan and Xiong [2003], Appendix C, we have $E[X_{ij}^T X_{ik}] = E[X_{ik}^T X_{ij}] = \text{diag}(1, V_A/2, V_D/4)$, $j, k = 1, 2, 3, j \neq k$. Denote $C_{31} = 3 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2$, and $C_{32} = 2[\rho_{12}\rho_{13} + \rho_{12}\rho_{23} + \rho_{13}\rho_{23} - \rho_{12} - \rho_{13} - \rho_{23}]$. By the above formulas, constants, and the formulas in Fan and Xiong [2002], Appendix A, we have

$$\frac{1}{k} \sum_{i=n+m+1}^{n+m+k} X_i^T \Sigma_i^{-1} X_i \approx \frac{1}{C_3\sigma^2} [C_{31} \text{diag}(1, V_A, V_D) + C_{32} \text{diag}(1, V_A/2, V_D/4)]. \quad (18)$$

Combine the n individuals, m sibpairs, and k tri-sibships. Denote

$$\begin{aligned} a_1 &= n + 2m(1 - \rho_{12}^2)^{-1}(1 - \rho_{12}) + k[C_{31} + C_{32}]/C_3, \\ a_2 &= n + 2m(1 - \rho_{12}^2)^{-1}(1 - \rho_{12}^2/2) + k[C_{31} + C_{32}]/C_3 \\ a_3 &= n + 2m(1 - \rho_{12}^2)^{-1}(1 - \rho_{12}/4) + k[C_{31} + C_{32}]/C_3. \end{aligned} \quad (19)$$

Then equations (16), (17) and (18) lead to equation (8).

References

- Abecasis GR, Cardon LR, Cookson WOC (2000) A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66: 279–292.
- Abecasis GR, Cookson WOC, Cardon LR (2000) Pedigree tests of linkage disequilibrium. *Eur J Hum Genet* 8:545–551.
- Abecasis GR, Cookson WOC, Cardon LR (2001) The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet* 68:1463–1474.
- Allison DB (1997) Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 60:676–690.
- Almasy L, Blangero J (1998) Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198–1211.
- Almasy L, Williams JT, Dyer TD, Blangero J (1999) Quantitative trait locus detection using combined linkage/disequilibrium analysis. *Genet Epidemiol* 17 (Suppl 1):S31–S36.
- Amos CI (1994) Robust variance-components approach for assessing linkage in pedigrees. *Am J Hum Genet* 54:534–543.
- Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE (1989) A more powerful robust sibpair test of linkage for quantitative traits. *Genet Epidemiol* 6:435–449.
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic map: Individual and sex-specific variation in recombination. *Am J Hum Genet* 63: 861–869.
- Cardon LR (2000) A sibpair regression model of linkage disequilibrium for quantitative traits. *Hum Hered* 50:350–358.
- Elston RC, Keats BJB (1985) Genetic analysis workshop III: sib pair analyses to determine linkage groups and to order loci. *Genet Epidemiol* 2:211–213.
- Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th ed. Longman, London.
- Fan R, Xiong M (2002) High resolution mapping of quantitative trait loci by linkage disequilibrium analysis. *Eur J Hum Genet* 10:607–615.
- Fan R, Xiong M (2003) Combined high resolution linkage and association mapping of quantitative trait loci. *Eur J Hum Genet* 11:125–137.
- Fulker DW, Cardon LR (1994) A sibpair approach to interval mapping of quantitative trait loci. *Am J Hum Genet* 54:1092–1103.
- Fulker DW, Cherny SS, Cardon LR (1995) Multiple interval mapping of quantitative trait loci, using sibpairs. *Am J Hum Genet* 56:1224–1233.
- Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sibpair analysis for quantitative traits. *Am J Hum Genet* 64:259–267.
- George V, Tiwari HK, Zhu XF, Elston RC (1999) A test of transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression. *Am J Hum Genet* 65:236–245.
- Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. *Am J Hum Genet* 47:957–967.
- Goldgar DE, Oniki RS (1992) Comparison of a multipoint identity-by-descent method with parametric multipoint linkage analysis for mapping quantitative traits. *Am J Hum Genet* 50:598–606.

- Göring HHH, Terwillinger JD (2000) Linkage analysis in the presence of error IV: Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 66:1310–1327.
- Graybill FA (1976) *Theory and Application of the Linear Model*. Pacific Grove, California.
- Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69:315–324.
- Hartl DL, Clark AG (1989) *Principles of Population Genetics*, 2nd ed. Sunderland, Sinauer.
- Harville DA (1997) *Matrix Algebra from a Statistician's Perspective*. Springer.
- Haseman JK (1970) *The Genetic Analysis of Quantitative Traits Using Twin and Sib Data*. Diss University of North Carolina, Chapel Hill.
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3–19.
- Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* 117:331–341.
- 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.
- The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933.
- Jansen RC (1993) Interval mapping of multiple quantitative trait loci. *Genetics* 135:205–211.
- Kong A, Gudbjartsson DF, Sainz J et al. (2002) A high resolution recombination map of the human genome. *Nat Genet* 31:241–247.
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199.
- Lange K (1997) *Mathematical and Statistical Methods for Genetic Analysis*. Springer.
- Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49:67.
- Meyers DA, Wjst M, Ober C (2001) Description of three data sets: Collaborative Study on the Genetics of Asthma (CSGA), the German Affected Sib Pair Study, and the Hutterites of South Dakota. *Genet Epidemiol* 21(suppl 1): S4–S8.
- 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.
- Pratt SC, Daly M, Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 66:1153–1157.
- Rabinowitz D (1997) A transmission disequilibrium test for quantitative trait loci. *Hum Hered* 47:342–350.
- Schaid DJ, Li H (1997) Genotype relative-risks and association tests for nuclear families with missing parents. *Genet Epidemiol* 14:1113–1118.
- Schaid DJ, Rowland C (1998) Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. *Am J Hum Genet* 63:1492–1506.
- Schork N (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: Efficiency, power, and modeling considerations. *Am J Hum Genet* 53:1306–1319.
- Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82:605–610.
- Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 66:1616–1630.
- 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.
- Kendall MG, Stuart A, Ord JK (1991) *Kendall's Advanced Theory of Statistics, Vol. 2: Classical Inference and Relationships*, 5th ed. New York, Oxford University Press.
- Wjst M, Fischer G, Immervoll T, et al. (1999) A genome-wide search for linkage to asthma. *Genomics* 58:1–8.
- Wu RL, Ma CX, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 160:779–792.
- Xiong MM, Jin L (2000) Combined linkage and linkage disequilibrium mapping for genome screens. *Genet Epidemiol* 19:211–234.
- Xu SZ, Atchley WR (1995) A random model approach to interval mapping of quantitative trait loci. *Genetics* 141:1189–1197.
- Zhang SL, Zhao HY (2001) Quantitative similarity-based association tests using population samples. *Am J Hum Genet* 69:601–614.
- Zhao J, Li W, Xiong M (2001) Population based linkage disequilibrium mapping of QTL: an application to simulated data in an isolated population. *Genet Epidemiol* 21 (S1):S655–659.
- Zhao J, Xiong M (2002) Unbiased quantitative population association test. *Am J Hum Genet* 71(suppl):568, Poster 2336.
- Zhu XF, Elston RC (2000) Power comparison of regression methods to test quantitative traits for association and linkage. *Genet Epidemiol* 18:322–330.