European Journal of Human Genetics (2002) 10, 607–615

www.nature.com/ejhg

ARTICLE

High resolution mapping of quantitative trait loci by linkage disequilibrium analysis

Ruzong Fan*[,1,2] and Momiao Xiong[3]

[1]*Department of Statistics, The Texas A&M University, 447 Blocker Building, College Station, Texas, TX 77843-3143, USA;* [2]*Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Sigmund Freud Strasse 25, D-53105 Bonn, Germany;* [3]*Human Genetics Center, University of Texas–Houston, P.O. Box 20334, Houston, Texas, TX 77225, USA*

Two methods, linkage analysis and linkage disequilibrium (LD) mapping or association study, are usually utilised for mapping quantitative trait loci (QTL). Linkage mapping is appropriate for low resolution mapping to localise trait loci to broad chromosome regions within a few cM (<10 cM), and is based on family data. Linkage disequilibrium mapping, on the other hand, is useful in high resolution or fine mapping, and is based on both population and family data. Using only one marker, one may carry out single-point linkage analysis and linkage disequilibrium mapping. Using two or more markers, it is possible to flank the QTL by multipoint analysis. The development and thus availability of dense marker maps, such as single nucleotide polymorphisms (SNP) in human genome, presents a tremendous opportunity for multipoint fine mapping. In this article, we propose a regression approach of mapping QTL by linkage disequilibrium mapping based on population data. Assuming that two marker loci flank one quantitative trait locus, a two-point linear regression is proposed to analyse population data. We derive analytical formulas of parameter estimations, and non-centrality parameters of appropriate tests of genetic effects and linkage disequilibrium coefficients. The merit of the method is shown by the power calculation and comparison. The two-point regression model can capture much more linkage and linkage disequilibrium information than that derived when only one marker is used. For a complex disease with heritability $h^2 \geqslant 0.15$, a study with sample size of 250 can provide high power for QTL detection under moderate linkage disequilibria.
European Journal of Human Genetics (2002) **10,** 607–615. doi:10.1038/sj.ejhg.5200843

Keywords: high resolution mapping; linkage mapping; linkage disequilibrium mapping; QTL

Introduction

Much research has been done on linkage mapping of qualitative or quantitative trait loci (QTL). Schork investigated multipoint identity-by-descent analysis of human quantitative traits.[1] Fulker and Cardon worked out a sib-pair approach of a two-point interval mapping for QTL.[2] Researchers have been extending the available methods toward several directions: (1) Fulker *et al*.[3] extended the method of Fulker and Cardon[2] for usage in multipoint interval mapping; (2) Almasy and Blangero worked on multipoint mapping for general pedigrees;[4] (3) Liang *et al*. proposed a unified sampling method for both qualitative and quantitative traits;[5] (4) Pratt *et al*. used an exact multipoint algorithm to analyse family data by variance component models.[6] The focus of the above studies was on linkage mapping, which is based on family data. Linkage analysis is appropriate for low resolution genetic mapping to localise trait loci to broad chromosome regions within a few cM (<10 cM).

Linkage disequilibrium (LD) mapping or association study, on the other hand, is based on both family and population data, and is useful in high resolution of genetic mapping, ie, fine disease gene mapping. The reason for the

*Correspondence: Dr R Fan, Department of Statistics, The Texas A&M University, 447 Blocker Building, College Station, TX 77843-3143, USA. Tel: 979 845 3156; Fax: 979 845 3144; E-mail: rfan@stat.tamu.edu
Received 18 October 2001; revised 12 March 2002; accepted 16 May 2002

high resolution of linkage disequilibrium mapping is that the allelic association due to linkage disequilibrium usually operates over short genetic distances. Linkage analysis and linkage disequilibrium mapping are complementary in disease gene mapping. To localise genetic traits, one may carry out linkage analysis as the first step on a sparse map to get suggestive linkage between genetic traits and markers. Then linkage disequilibrium mapping can be used as a follow-up in high resolution mapping of the genetic traits on a more dense map.

Abecasis *et al*.,[7] Fulker *et al*.[8] and Sham *et al*.[9] have explored linkage and association studies of quantitative traits by variance–component procedures, allowing a simultaneous test of allelic association for family data. Zhao *et al*.[10] applied a regression approach of linkage disequilibrium mapping to localise QTL in humans. In these studies the investigators used only one marker in their analysis. However, very dense maps such as single nucleotide polymorphisms (SNPs) in human genome (The International SNP Map Work Group) are available now.[11] These exciting developments allow us to explore models and methodologies of simultaneously using two or more markers in high resolution linkage disequilibrium mapping of QTL.

In this article, we propose a linear regression method of high resolution mapping for QTL by using linkage disequilibrium analysis which is based on population data. Assuming that two marker loci flank one genetic trait locus, a linear regression is introduced based on an intuitive rationale. Then we derive analytical formulas of parameter estimations, and non-centrality parameters of appropriate tests of genetic effects and linkage disequilibrium coefficients. The merit of the regression method is shown by the power calculation and comparison.

Models

Consider a quantitative trait which is influenced by a quantitative trait locus Q, which is flanked by two markers A and B in an order of AQB. Suppose that there are two alleles Q_1 and Q_2 at the trait locus with frequencies q_1 and q_2. 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, respectively. Suppose that markers A and B are in Hardy–Weinberg equilibrium, ie, $P(AA)=P_A^2,\ P(Aa)=2P_aP_A,\ P(aa)=P_a^2$ and $P(BB)=P_B^2,\ P(Bb)=2P_BP_b,\ P(bb)=P_b^2$. However, they may be in linkage disequilibrium. Let us denote the measure of linkage disequilibrium between trait locus Q and marker A by $D_{AQ}=P(AQ_1)-q_1P_A$, the measure of linkage disequilibrium between trait locus Q and marker B by $D_{QB}=P(BQ_1)-q_1P_B$, and the measure of linkage disequilibrium between marker A and marker B by $D_{AB}=P(AB)-P_AP_B$.[12–14] In addition to the major QTL Q, assume that there is an error effect that influences the trait. Then the total variance can be decomposed as $\sigma^2=\sigma_g^2+\sigma_e^2$, σ_g^2 is variance explained by the putative QTL Q, and σ_e^2 is error variance. The genetic variances $\sigma_g^2=\sigma_{ga}^2+\sigma_{gd}^2$ is decomposed into additive and dominant components, respectively. Assume that there are n independent individuals from a population with trait values y_i, genotype A_i at marker A and genotype B_i at marker B. Consider the following regression equation

$$y_i=\beta+w_i\gamma+x_{Ai}\alpha_A+x_{Bi}\alpha_B+z_{Ai}\delta_A+z_{Bi}\delta_B+e_i, \qquad (1)$$

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

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

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

α_A, α_B, δ_A and δ_B are regression coefficients of the dummy variables x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi}. Let us denote an experimental design matrix X by

$$X=\begin{pmatrix}1 & w_1 & x_{A1} & x_{B1} & z_{A1} & z_{B1}\\ 1 & w_2 & x_{A2} & x_{B2} & z_{A2} & z_{B2}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & w_n & x_{An} & x_{Bn} & z_{An} & z_{Bn}\end{pmatrix}=\begin{pmatrix}X_1^\tau\\ X_2^\tau\\ \vdots\\ X_n^\tau\end{pmatrix},$$

a vector of regression coefficients by $\mu=(\beta,\gamma^\tau,\alpha_A,\alpha_B,\delta_A,\delta_B)^\tau$, the quantitative traits by a vector $Y=(y_1,y_2,\ldots,y_n)^\tau$, and errors terms by $e=(e_1,e_2,\ldots,e_n)^\tau$. Then we may write the model (1) as $Y=X\mu+e$. By standard regression theory, we may estimate the coefficients by $\hat{\mu}=(X^\tau X)^{-1}X^\tau Y$.

To give an intuitive rationale of model (1), let μ_{ij} be the effect of genotype Q_iQ_j,i,j=1,2,$\mu_{12}=\mu_{21}$. Let the genic effect of allele Q_i be α_i,i=1,2. Then genotypic effects can be expressed as $\mu_{11}=\mu_0+2\alpha_1+d_1$,$\mu_{12}=\mu_0+\alpha_1+\alpha_2+d_2$,$\mu_{22}=\mu_0+2\alpha_2+d_3$, where μ_0 is the overall population mean, d_i is the deviation of the related genotypic value from that of an additive effect model. Minimising $F(\mu_0,\alpha_1,\alpha_2)=\sum_{i=1}^{2}\sum_{j=1}^{2}q_iq_j\ (\mu_{ij}-\mu_0-\alpha_i-\alpha_j)^2$, the estimates of μ_0,α_1,α_2 are $\mu=\mu_{11}q_1^2+2\mu_{12}q_1q_2+\mu_{22}q_2^2$, $\hat{\alpha}_1=q_1\mu_{11}+q_2\mu_{12}-\mu$ and $\hat{\alpha}_2=q_1\mu_{21}+q_2\mu_{22}-\mu$ (Jacquard,[15] Chapter 5). Plugging these estimates into μ_{ij}, we can obtain that $\mu_{11}=\mu+2q_2\alpha_Q-q_2^2\delta_Q$, $\mu_{12}=\mu+(q_2-q_1)\alpha_Q+q_1q_2\delta_Q$, $\mu_{22}=\mu-2q_1\alpha_Q-q_1^2\delta_Q$. Here $\alpha_Q=q_1\mu_{11}+(q_2-q_1)\mu_{12}-q_2\mu_{22}$ is the average effect of gene substitution, and $\delta_Q=2\mu_{12}-\mu_{11}-\mu_{22}$ is the dominant deviation. Assume that marker A coincides with the trait locus Q, and marker allele A is trait allele Q_1

and marker allele a is trait allele Q_2. Then the trait value can be expressed as $y_i=\mu+x_{Qi}\alpha_Q+z_{Qi}\delta_Q+e_i$. In practice, information of trait locus Q is unknown, but the information at marker loci is available. This prompts us to propose regression model (1) to map QTLs.

Assume that there are no covariates. Suppose that the markers A and B are in Hardy–Weinberg equilibrium. Then $Ex_{Ai}=Ex_{Bi}=Ez_{Ai}=Ez_{Bi}=0$. When the sample size n is large enough, we show in Appendix A that the coefficients are approximately given by

$$\hat{\beta}\approx\sum_{i=1}^{n}y_i/n,\ \begin{pmatrix}\hat{\alpha}_A\\ \hat{\alpha}_B\end{pmatrix}\approx\begin{pmatrix}\sum_{i=1}^n x_{Ai}^2 & \sum_{i=1}^n x_{Ai}x_{Bi}\\ \sum_{i=1}^n x_{Ai}x_{Bi} & \sum_{i=1}^n x_{Bi}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum_{i=1}^n x_{Ai}y_i\\ \sum_{i=1}^n B_iy_i\end{pmatrix},\ \text{and}$$

$$\begin{pmatrix}\hat{\delta}_A\\ \hat{\delta}_B\end{pmatrix}\approx\begin{pmatrix}\sum_{i=1}^n z_{Ai}^2 & \sum_{i=1}^n z_{Ai}z_{Bi}\\ \sum_{i=1}^n z_{Ai}z_{Bi} & \sum_{i=1}^n z_{Bi}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum_{i=1}^n z_{Ai}y_i\\ \sum_{i=1}^n z_{Bi}y_i\end{pmatrix}.$$

If the markers A and B are in linkage equilibrium, ie, $D_{AB}=0$, then the above equations simplify to the following (Appendix A)

$$\hat{\alpha}_A\approx\frac{\sum_{i=1}^n x_{Ai}y_i}{\sum_{i=1}^n x_{Ai}^2},\hat{\alpha}_B\approx\frac{\sum_{i=1}^n x_{Bi}y_i}{\sum_{i=1}^n x_{Bi}^2},\hat{\delta}_A\approx\frac{\sum_{i=1}^n z_{Ai}y_i}{\sum_{i=1}^n z_{Ai}^2},\hat{\delta}_B\approx\frac{\sum_{i=1}^n z_{Bi}y_i}{\sum_{i=1}^n z_{Bi}^2}.\quad(2)$$

Property of regression coefficients

As in the previous section, let μ_{ij} be the effect of genotype $Q_iQ_j,i,j=1,2$. If μ_{11}=a, $\mu_{12}=d$, and $\mu_{22}=-a$ as in the traditional quantitative genetics (Falconer and Mackay[16]), $\alpha_Q=a+(q_2-q_1)d$ and $\delta_Q=2d$. For general case, one may form the above relations by letting $a=\mu_{11}-(\mu_{11}+\mu_{22})/2$ and $d=\mu_{12}-(\mu_{11}-\mu_{22})/2$. It is well known that the additive variance $\sigma_{ga}^2=2q_1q_2\alpha_Q^2$ and the dominant variance $\sigma_{gd}^2=(q_1q_2)^2\sigma_Q^2$. A true random effect model describing the trait value is

$$y_i=\beta+w_i\gamma+g_i+e_i,\quad(3)$$

where

$$g_i=\begin{cases}\mu_{11} & \text{for genotype } Q_1Q_1\\ \mu_{12} & \text{for genotype } Q_1Q_1\\ \mu_{22} & \text{for genotype } Q_2Q_2.\end{cases}$$

Let us denote three ratios $D_{AB}^2/(P_aP_AP_bP_B)=R_{AB}^2, D_{AQ}^2/(P_aP_Aq_1q_2)=R_{AQ}^2$, and $D_{QB}^2/(q_1q_2P_bP_B=R_{QB}^2)$. In Appendix B, we will show that the coefficients of regression equation (1) are given by

$$\begin{pmatrix}\alpha_A\\ \alpha_B\end{pmatrix}=\begin{pmatrix}P_aP_A & D_{AB}\\ D_{AB} & P_bP_B\end{pmatrix}^{-1}\begin{pmatrix}D_{AQ}\\ D_{QB}\end{pmatrix}\alpha_Q=\begin{pmatrix}\frac{R_{AQ}-R_{AB}R_{QB}}{\sqrt{P_AP_a}}\\ \frac{R_{QB}-R_{AB}R_{AQ}}{\sqrt{P_BP_b}}\end{pmatrix}\frac{\sqrt{q_1q_2\alpha_Q}}{1-R_{AB}^2},\quad(4)$$

$$\begin{pmatrix}\delta_A\\ \delta_B\end{pmatrix}=\begin{pmatrix}P_a^2P_A^2 & D_{AB}^2\\ D_{AB}^2 & P_b^2P_B^2\end{pmatrix}^{-1}\begin{pmatrix}D_{AQ}^2\\ D_{QB}^2\end{pmatrix}\delta_Q=\begin{pmatrix}\frac{R_{AQ}^2-R_{AB}^2R_{QB}^2}{P_AP_a}\\ \frac{R_{QB}^2-R_{AB}^2R_{AQ}^2}{P_BP_b}\end{pmatrix}\frac{q_1q_2\delta_Q}{1-R_{AB}^4}.\quad(5)$$

Assume that the two markers A and B are not in linkage disequilibrium, ie, $D_{AB}=0$. Then $\alpha_A=D_{AQ}\alpha_Q/(P_aP_A),\alpha_B=D_{QB}\alpha_Q/(P_bP_B)$, $\delta_A=D_{AQ}^2\delta_Q/(P_a^2P_A^2)$, and $\delta_B=D_{QB}^2\delta_Q/(P_b^2P_B^2)$. Hence, marker A and marker B independently contribute to the analysis of the trait values. Furthermore, assume the trait locus Q is in linkage disequilibrium with marker A but not with marker B. Then $D_{QB}=0$ and so $\alpha_A=D_{AQ}\alpha_Q/(P_aP_A),\delta_A=D_{AQ}^2\delta_Q/(P_a^2P_A^2),\alpha_B=\delta_B=0$. Hence, only marker A contributes to the analysis and marker B has no effect on the result. This is equivalent to using one marker for the analysis.

If one marker coincides with the trait locus, for instance locus Q is marker A, we can show that the other marker B does not contribute to estimations of the substitution and dominant effects of the trait locus. Actually, assume that allele $A=Q_1$ and allele $a=Q_2$. Then $D_{AB}=D_{QB}$ and $D_{AQ}=q_1q_2$. This leads to $\begin{pmatrix}\alpha_A\\ \alpha_B\end{pmatrix}=\begin{pmatrix}\alpha_Q\\ 0\end{pmatrix}$ and $\begin{pmatrix}\delta_A\\ \delta_B\end{pmatrix}=\begin{pmatrix}\delta_Q\\ 0\end{pmatrix}$. Hence, marker A can fully estimate the substitution and dominant effects of the trait locus Q.

In general, assume that marker A and marker B are in linkage disequilibrium. Then model (1) simultaneously takes care of the linkage disequilibrium and the effects of the putative trait locus Q. The parameters of linkage disequilibrium (ie, D_{AQ} and D_{QB}) and gene effect (ie, α_Q and δ_Q) are contained in the mean coefficients. We may simultaneously test linkage disequilibrium of marker A and marker B with trait locus Q, the gene substitution and dominant effects by testing $\alpha_A=\alpha_B=\delta_A=\delta_B=0$. From equation (4), we may test the linkage disequilibrium of markers A and B with the trait locus Q and the gene substitution effect α_Q by testing $\alpha_A=\alpha_B=0$. From equation (5), we may test the linkage disequilibrium of markers A and B with the trait locus Q and the dominant effect by testing $\delta_A=\delta_B=0$.

Non-centrality parameters

Assume that there are no covariates. Then $\mu=(\beta,\alpha_A,\alpha_B,\delta_A,\delta_B)^\tau$. Let H be a $q\times5$ matrix of rank q. By Graybill,[17] Chapter 6, the test statistic of a hypothesis $H\mu=0$ is non-central $F(q,n-5)$ defined by $F=\frac{(H\hat{\mu})^\tau[H(X^\tau X)^{-1}H^\tau]^{-1}(H\hat{\mu})}{Y^\tau[I_n-(X^\tau X)^{-1}]Y}\frac{n-5}{q}$, where I_n is the $n\times n$ identity matrix. The non-centrality parameter of the test statistic F can be calculated by $\lambda=[1/(2\sigma^2)](H\mu)^\tau[H[X^\tau X]^{-1}H^2]^{-1}H\mu$. To test if there are additive and dominant effects, we may test the hypothesis $H_{AB,ad}$: $\alpha_A=\alpha_B=\delta_A=\delta_B=0$. Then the test matrix H is defined by

$$H=\begin{pmatrix}0&1&0&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&0&1\end{pmatrix}.\quad(6)$$

Let us denote the corresponding F-test statistic by $F_{AB,ad}$. In Appendix C, we show

$$\lambda_{AB,ad} \approx \frac{n}{2\sigma^2}[2\alpha_Q^2[P_bP_BD_{AQ}^2 \quad 2D_{AQ}D_{AB}D_{QB} + P_aP_AD_{QB}^2]/(P_aP_AP_bP_B \quad D_{AB}^2) + \delta_Q^2[P_b^2P_B^2D_{AQ}^4 \quad 2D_{AQ}^2D_{AB}^2D_{QB}^2 + P_a^2P_A^2D_{QB}^4]/(P_a^2P_A^2P_b^2P_B^2 \quad D_{AB}^4)] = \frac{n}{2\sigma^2}[\sigma_{ga}^2[R_{AQ}^2 \quad 2R_{AQ}R_{AB}R_{QB} + R_{QB}^2]/(1 \quad R_{AB}^2) + \sigma_{gd}^2[R_{AQ}^4 \quad 2R_{AQ}^2R_{AB}^2R_{QB}^2 + R_{QB}^4]/(1 \quad R_{AB}^4)]. \tag{7}$$

If one assumes that (a) the two markers A and B are not in linkage disequilibrium, then D_{AB}=0; (b) the trait locus Q is in linkage disequilibrium with marker A but not with marker B, then D_{QB}=0 and $D_{AQ} \neq 0$. Then $\lambda_{AB,ad} \approx [n/(2\sigma^2)][\sigma_{ga}^2R_{AQ}^2 + \sigma_{gd}^2R_{AQ}^4]$, which only involves marker A and can be written as $\lambda_{A,ad}$. Correspondingly, we denote the related F-test statistic by $F_{A,ad}$. Furthermore, assume (c) there is no dominant effect, ie, $\sigma_{gd}^2 = 0$. Then $\lambda_{A,a} \approx [n/(2\sigma^2)]\sigma_{ga}^2R_{AQ}^2$ is the non-centrality parameter of the related F-test statistic $F_{A,a}$.

To test other hypotheses, we may get the non-centrality parameters in a similar way by taking appropriate test matrices H. To test if there is dominant effect, we may test the hypothesis $H_{A,B,d}: \delta_A=\delta_B=0$. The non-centrality parameter is $\lambda_{AB,d} \approx [n/(2\sigma^2)]\sigma_{gd}^2[R_{AQ}^4 \quad 2R_{AQ}^2R_{AB}^2R_{QB}^2 + R_{QB}^4]/(1 \quad R_{AB}^4)$. The related F-test statistic is denoted by $F_{AB,d}$. To test if there is additive or substitution effect, we may test the hypothesis $H_{AB,a}: \alpha_A=\alpha_B=0$. The non-centrality parameter is $\lambda_{AB,a} \approx [n/(2\sigma^2)]\sigma_{ga}^2[R_{AQ}^2 \quad 2R_{AQ}R_{AB}R_{QB} + R_{QB}^2]/(1 \quad R_{AB}^2)$. The related F-test statistic is denoted by $F_{AB,a}$. To test if there are additive and dominant effects at marker locus A given that there are effects at marker locus B, we may test the hypothesis $H_{A|B,ad}: \alpha_A=\delta_A=0$. The non-centrality parameter is

$$\lambda_{A|B,ad} \approx \frac{n}{2\sigma^2}[\frac{2P_aP_AP_bP_B \quad 2D_{AB}^2}{P_bP_B}\alpha_A^2 + \frac{P_a^2P_A^2P_b^2P_B^2 \quad D_{AB}^4}{P_b^2P_B^2}\delta_A^2] = \frac{n}{2\sigma^2}[\sigma_{ga}^2[P_bP_BD_{AQ} \quad D_{AB}D_{QB}]^2/[P_bP_Bq_1q_2(P_aP_AP_bP_B \quad D_{AB}^2)] + \sigma_{gd}^2[P_b^2P_B^2D_{AQ}^2 \quad D_{AB}^2D_{QB}^2]^2/[P_b^2P_B^2q_1^2q_2^2(P_a^2P_A^2P_b^2P_B^2 \quad D_{AB}^4)]]. = \frac{n}{2\sigma^2}[\sigma_{ga}^2[R_{AQ} \quad R_{AB}R_{QB}]^2/[1 \quad R_{AB}^2] + \sigma_{gd}^2[R_{AQ}^2 \quad R_{AB}^2R_{QB}^2]^2/[1 \quad R_{AB}^4]].$$

To test if there is dominant effect at marker locus A given that there are effects at marker locus B, we may test the hypothesis $H_{A|B,d}: \delta_A=0$. The non-centrality parameter is $\lambda_{A|B,d} \approx [n/(2\sigma^2)]\sigma_{gd}^2[R_{AQ}^2 \quad R_{AB}^2R_{QB}^2]^2/[1 \quad R_{AB}^4]$.

Power calculation and comparison

To investigate the usefulness of the methods proposed in this article, we performed power and sample size calculations. As usual, we denote the heritability by h^2 which is defined by $h^2 = \sigma_{ga}^2/\sigma^2$. In the power calculations, we first take the equal allele frequencies P_A=q_1=P_B=0.5 at the two markers A and B, and the trait locus Q. Moreover, suppose that μ_{11}=a,μ_{12}=μ_{21}=d and μ_{22}=−a. Assume that marker A and marker B are in linkage equilibrium, ie, D_{AB}=0, the heritability h^2=0.25, and a sample size n=120. Figures 1 and 2 show the power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the disequilibrium coefficient D_{AQ} when D_{QB}=0.15 for a mode of dominant inheritance with a=d=1.0 and a mode of recessive inheritance with a=1.0, d=−0.5, respectively. The statistic $F_{AB,ad}$ has the highest power, and $F_{A,ad}$ has higher power than that of $F_{A,a}$. Hence, the regression approach that uses two markers A and B is advantageous over the one marker mapping that uses only one marker A or B.

Assume that the markers A and B are in moderate linkage disequilibrium, ie, D_{AB}=0.1, and that the linkage disequilibrium coefficients D_{AQ}=D_{QB}=0.15. Figures 3 and 4 show the power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$ and $F_{A,a}$ against the heritability h^2 for a mode of dominant inheritance with a=d=1.0 and a mode of recessive inheritance with a=1.0, d=−0.5, respectively. For a population with sample size n=250, the regression approach can achieve a high power for a trait with heritability $h^2 \geqslant 0.15$. Hence, the high resolution linkage disequilibrium mapping is a promising tool in mapping complex traits.

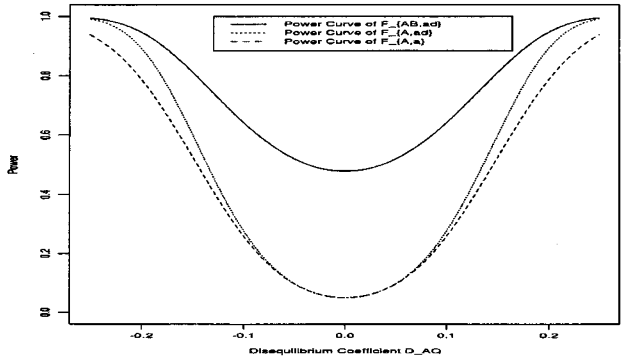

Figure 1 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the disequilibrium coefficient D_{AQ} when q_1=P_A=P_B=0.50, D_{AB}=0.0, D_{QB}=0.15, h_2=0.25, n=120 for a dominant trait a=d=1.0.

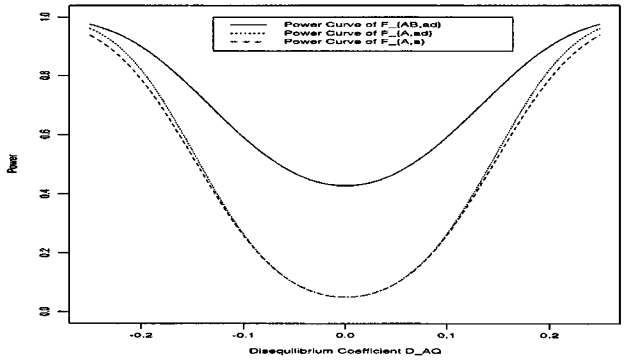

Figure 2 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the disequilibrium coefficient D_{AQ} when q_1=P_A=P_B=0.50, D_{AB}=0.0, D_{QB}=0.15, h^2=0.25, n=120 for a recessive trait a=1.0 and d=−0.5.

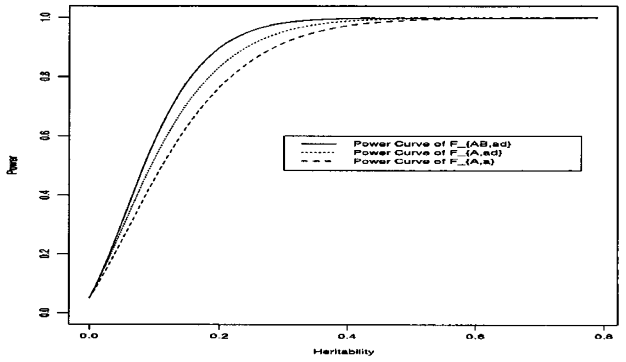

Figure 3 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the heritability h^2 when $q_1=P_A=P_B=0.50$, $D_{AB}=0.10$, $D_{AQ}=D_{QB}=0.15$, $n=250$ for a dominant trait $a=d=1.0$.

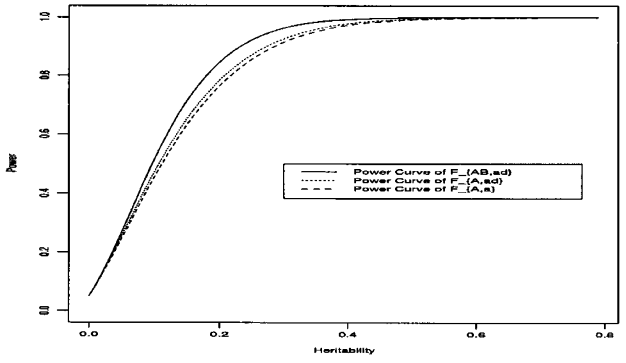

Figure 4 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the heritability h^2 when $q_1=P_A=P_B=0.50$, $D_{AB}=0.10$, $D_{AQ}=D_{QB}=0.15$, $n=250$ for a recessive trait $a=1.0$ and $d=-0.5$.

In a population, the linkage disequilibrium exists if mutations at the trait locus occur. Once the mutations occur, the recombination between a marker locus and the trait locus can dissipate the disequilibrium from generation to generation. Let us denote the frequency of haplotype AQ at the generation when the mutations occur by $P(AQ)(0)$. Then the linkage disequilibrium coefficient is $D_{AQ}(0)=P(AQ)(0)-q_1P_A$ at the generation when the mutations occur. For the following generations, the disequilibrium coefficient is reduced by a factor $1-\theta_{AQ}$ in each generation,[12] where θ_{AQ} is the recombination fraction between trait locus Q and marker A. Suppose that the mutation is already T generations old. Then the disequilibrium coefficient is $D_{AQ}(T)=D_{AQ}(0)(1-\theta_{AQ})^T$. Similarly, we may calculate the disequilibrium coefficients by $D_{AB}(T)=D_{AB}(0)(1-\theta_{AB})^T$ and $D_{QB}(T)=D_{QB}(0)(1-\theta_{QB})^T$, where θ_{QB} is the recombination fraction between trait locus Q and marker B, and θ_{AB} is the recombination fraction between marker A and marker B.

Suppose that we know the map distance λ_{AB} between marker A and marker B. Under the assumption of no interference, we may calculate the recombination fraction $\theta_{AB}=[1-\exp(-2\lambda_{AB}]/2$ by Haldane's map function. Similarly, we may calculate the recombination fractions θ_{AQ} and θ_{QB} by the map distances λ_{AQ} and λ_{QB}. Assume that the map distance between marker A and marker B is $\lambda_{AB}=5cM$, and the other parameters are given by $D_{AB}(0)=0.20$, $D_{AQ}(0)=D_{QB}(0)=0.25$, $h^2=0.25$, $n=120$, $T=20$. Figures 5 and 6 show the power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the recombination fraction θ_{AQ}, for a mode of dominant inheritance with $a=d=1.0$ and a mode of recessive inheritance with $a=1.0$, $d=-0.5$, respectively. We can see that the power of $F_{AB,ad}$ is very high, although the power of $F_{A,ad}$ and $F_{A,a}$ decreases very rapidly as the recombination fraction θ_{AQ} increases. Hence, the regressions using two markers are advantageous for fine gene mapping, and appropriate for the dense marker map such as SNPs in human genome.

To investigate the less favourable case other than the equal allele frequencies of trait locus and marker loci,

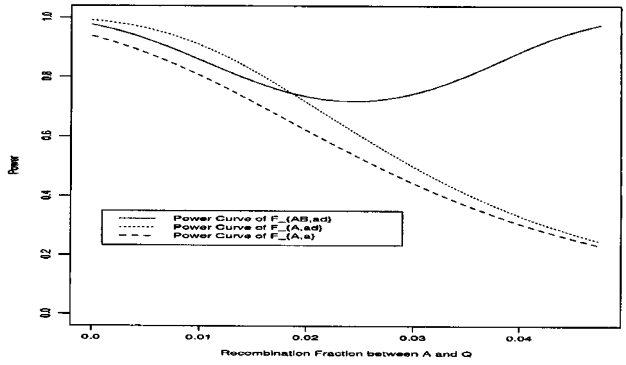

Figure 5 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the recombination fraction θ_{AQ} when $q_1=P_A=P_B=0.50$, $D_{AB}(0)=0.20$, $D_{AQ}(0)=D_{QB}(0)=0.25$, $h^2=0.25$, $\lambda_{AB}=5$ cM, $n=120$, $T=20$ for a dominant trait $a=d=1.0$.

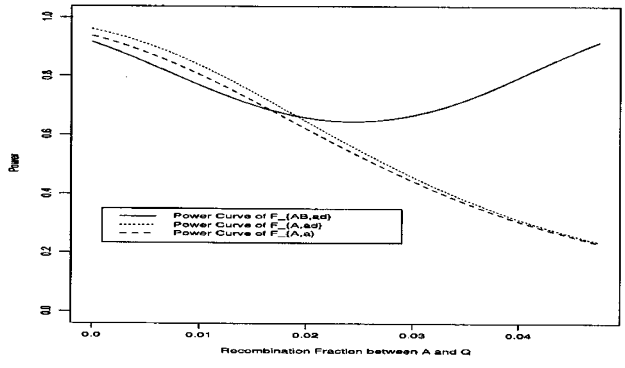

Figure 6 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the recombination fraction θ_{AQ} when $q_1=P_A=P_B=0.50$, $D_{AB}(0)=0.20$, $D_{AQ}(0)=D_{QB}(0)=0.25$, $h^2=0.25$, $\lambda_{AB}=5$ cM, $n=120$, $T=20$ for a recessive trait $a=1.0$ and $d=-0.5$.

Figure 7 shows the power curves of $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the linkage disequilibrium coefficient D_{AQ} when q_1=0.20, P_A=P_B=0.80, D_{AB}=0.0, D_{QB}=0.04, h^2=0.25, n=120 for a dominant trait a=1.0 and d=0.8. The three power curves are very close. Moreover, the power decreases rapidly when the linkage disequilibrium between trait locus Q and marker A decreases. For a recessive trait a=1.0 and d=−0.5, Figure 8 shows the power curves against the recombination fraction when q_1=0.20, P_A=P_B=0.80, D_{AB}(0)=0.10, D_{AQ}(0)=D_{QB}(0)=−0.15, h^2=0.25, λ_{AB}=5cM, n=120, T=20.

Figures 9 and 10 show two plots of the sample size against the heritability h^2 at a significant level 0.05 for a given power 0.80. In a favourable case when q_1=P_A=P_B=0.50, D_{AB}=0.10, D_{AQ}=D_{QB}=0.15 for a dominant trait a=1.0 and d=0.80, the required sample size is lower than 400, if the heritability is not lower than 0.1 (Figure 9). However, for an extremely less favourable case when q_1=0.20, P_A=P_B=0.80, D_{AB}=0.0, D_{AQ}=0.03, D_{QB}=0.04 for a recessive trait a=1.0 and d=−0.5, the required sample size is huge (Figure 10). Unfortunately, the true QTL frequency is rarely, if ever, known. Hence, linkage disequilibrium mapping works only when the linkage disequilibria are reasonably high, at least one needs moderate linkage disequilibria.

Discussion

With the development of dense marker maps, such as SNPs in human genome (The International SNP Map Work Group[11]), fine disease gene mapping is getting more and more important for the study of complex diseases. Association study is a simple and useful method in fine disease gene mapping (Cardon and Bell[18]; Risch and Merikangas[19]). In this article, we proposed a linear regression method to perform high resolution linkage disequilibrium mapping of QTLs. In the regression, we used information of two flanking markers to model the additive and dominant effects of a QTL, and also the linkage disequilibria between the markers and the trait locus. In addition to the additive and dominant effects, we may add the covariates to model

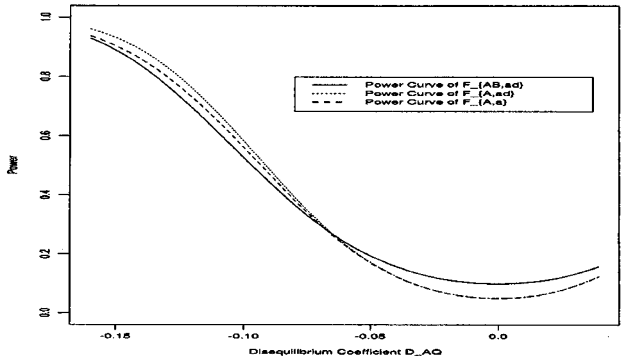

Figure 7 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the disequilibrium coefficient D_{AQ} when q_1=0.20, P_A=P_B=0.80, D_{AB}=0.0, D_{QB}=0.04, h^2=0.25, n=120 for a dominant trait a=1.0 and d=0.8.

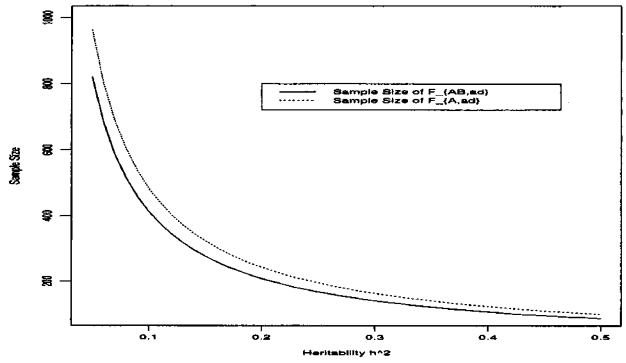

Figure 9 Sample sizes of the test statistics $F_{AB,ad}$ and $F_{A,ad}$ against the heritability h^2 at a significant level 0.05 for a given power 0.80, when q_1=P_A=P_B=0.50, D_{AB}=0.10, D_{AQ}=D_{QB}=0.15 for a dominant trait a=1.0 and d=0.80.

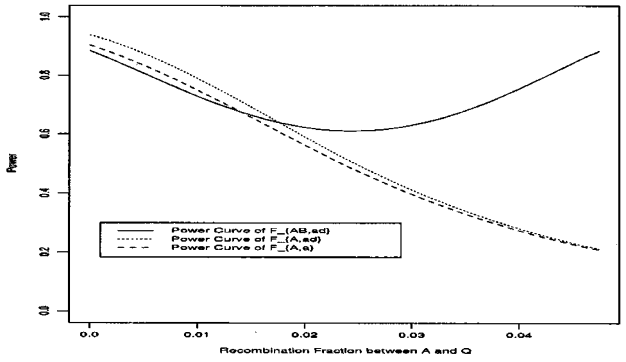

Figure 8 Power curves of the test statistics $F_{AB,ad}$, $F_{A,ad}$, and $F_{A,a}$ against the recombination fraction θ_{AQ} when q_1=0.20, P_A=P_B=0.80, D_{AB}(0)=0.10, D_{AQ}(0)=D_{QB}(0)=−0.15, h^2=0.25, λ_{AB}=5cM, n=120, T=20 for a recessive trait a=1.0 and d=−0.5.

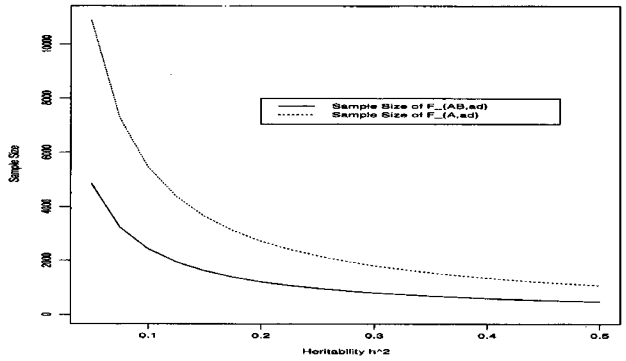

Figure 10 Sample sizes of the test statistics $F_{AB,ad}$ and $F_{A,ad}$ against the heritability h^2 at a significant level 0.05 for a given power 0.80, when q_1=0.20, P_A=P_B=0.80, D_{AB}=0.0, D_{AQ}=0.03, D_{QB}=0.04 for a recessive trait a=1.0 and d=−0.5.

their effects. Due to the simplicity, the method can be easily performed by routine statistical analysis softwares such as SAS and Splus.

After studying the merits of the method of using two markers as proposed in this article, we concluded that this method is well suited for mapping complex diseases. It provides higher power than that of using only one marker approach. The advantages of high resolution mapping have been explored by many authors by using linkage analysis of family data or plant/animal data[20–25]). However, there is not sufficient statistical analysis regarding the high resolution mapping by linkage disequilibrium mapping method. Using population data, Zhao *et al.*[10] applied an approach of linkage disequilibrium mapping based on regression to map QTL in humans. Abecasis *et al.*,[7] Allison *et al.*,[26] Fulker *et al.*,[3] Göring and Terwillinger,[27] and Sham *et al.*[9] have explored linkage and association studies of quantitative traits by variance–component procedures allowing a simultaneous test of allelic association for family data. One interesting approach is to combine both family and population data, and perform combined linkage analysis and linkage disequilibrium high resolution mapping.

The power of linkage disequilibrium mapping depends on the existence of disequilibrium between a trait locus and a marker. In a population, linkage disequilibrium exists if mutations at the trait locus occur. In the absence of tight linkage, the degree of linkage disequilibrium decreases very rapidly after a few generations due to the recombination between the trait locus and the markers. Hence, linkage disequilibrium mapping is appropriate for the analysis of dense marker maps to do high resolution fine gene mapping. In practice, one can perform linkage disequilibrium mapping following prior evidence of linkage. Linkage analysis is less sensitive to population stratification, population history, or environmental effects. Moreover, linkage mapping is appropriate for low resolution mapping to localise trait loci to broad chromosome regions (<10 cM). The two methods, linkage mapping and linkage disequilibrium mapping, are complementary for disease gene mapping.

Potential problems of linkage disequilibrium mapping include population stratification, population history, or environmental effects. It is well understood that for the same number of individuals, family based linkage disequilibrium methods are less powerful than the population based methods. However, utilising family based linkage disequilibrium approaches may avoid false positives due to the sources of linkage disequilibrium such as population admixtures rather than linkage. One research area is to combine the population and pedigree data to do linkage disequilibrium mapping, and use the pedigree data alone to perform linkage mapping (Fulker *et al.*[8]).

As in Sham *et al.*,[9] we notice that the non-centrality parameter is reduced by a factor equal to R^2_{AQ} for additive variance, and a factor of R^4_{AQ} for dominant variance, if we use only one marker A to perform analysis. Hence, the power decreases rapidly when the linkage disequilibrium between the trait locus and the marker is reduced. The degree of linkage disequilibrium depends heavily on the map distance between the trait locus and the marker locus, and most likely maintains high linkage disequilibrium when the two loci are very close. Hence, the high resolution mapping method proposed in this article has a good potential for being used in fine disease gene mapping. As mentioned in Sham *et al.*,[9] the property of the measurements R^2_{AQ}, R^2_{QB} and R^2_{AB} needs more investigation, and their roles in different scenarios should be studied more thoroughly.

Acknowledgements

We thank two reviewers, Section Editors, and Dr Gert-Jan B von Ommen for their helpful comments to improve the article. Dr Joanna Floros read the manuscript and provided helpful suggestions in improving the grammar. R Fan was supported partially by a research fellowship from Alexander von Humboldt Foundation, Germany, and an International Research Travel Assistance Grant of the Texas A&M University. M Xiong was supported by NIH grant R01-GM56515, and MH59518.

References

1 Schork NJ: Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am J Hum Genet* 1993; **53**: 1306–1319.
2 Fulker DW, Cardon LR: A sib-pair approach to interval mapping of quantitative trait loci. *Am J Hum Genet* 1994; **54**: 1092–1103.
3 Fulker DW, Cherny SS, Cardon LR: Multipoint interval mapping of quantitative trait loci, using sibpairs. *Am J Hum Genet* 1995; **56**: 1224–1233.
4 Almasy L, Blangero J: Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**: 1198–1211.
5 Liang KY, Huang CY, Beaty TH: A unified sampling approach for multipoint analysis of qualitative and quantitative traits in sibs. *Am J Hum Genet* 2000; **66**: 1631–1641.
6 Pratt SC, Daly M, Kruglyak L: Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 2000; **66**: 1153–1157.
7 Abecasis GR, Cardon LR, Cookson WOC: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.
8 Fulker DW, Cherny SS, Sham PC, Hewitt JK: Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet* 1999; **64**: 259–267.
9 Sham PC, Cherny SS, Purcell S, Hewitt JK: Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 2000; **66**: 1616–1630.
10 Zhao J, Li W, Xiong M: Population based linkage disequilibrium mapping of QTL: an application to simulated data in an isolated population. *Genetic Epidemiology* 2001; **21** (S1): S655–S659.
11 The International SNP Map Working Group: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; **409**: 928–933.
12 Hartl DL, Clark AG: *Principles of Population Genetics*. 2nd edn. Sunderland, MA: Sinauer Associates, 1989.

13 Hedrick PW: Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.
14 Lewontin RC: The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 1964; **49**: 67.
15 Jacquard A: *The Genetic Structure of Populations*. New York: Springer-Verlag, 1974.
16 Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics*. 4th edn. London: Longman, 1996.
17 Graybill FA: *Theory and Application of the Linear Model*. California: Pacific Grove, 1976.
18 Cardon LR, Bell J: Association study designs for complex diseases. *Nature Reviews Genetics* 2001; **2**: 91–99.
19 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
20 Amos C, Andrade MD: Genetic linkage methods for quantitative traits. *Statistical Methods in Medical Research* 2001; **10**: 3–25.
21 Goldgar DE: Multipoint analysis of human quantitative genetic variation. *Amr J Hum Genet* 1990; **47**: 957–967.
22 Haley CS, Knott SA: A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 1992; **69**: 315–324.
23 Jansen RC: Interval mapping of multiple quantitative trait loci. *Genetics* 1993; **135**: 205–211.
24 Xu SZ, Atchley WR: A random model approach to interval mapping of quantitative trait loci. *Genetics* 1995; **141**: 1189–1197.
25 Zeng ZB: Precision mapping of quantitative trait loci. *Genetics* 1994; **136**: 1457–1468.
26 Allison DB, Heo M, Kaplan N, Martin ER: Sibling-based tests of linkage and association for quantitative traits. *Am J Hum Genet* 1999; **64**: 1754–1764.
27 Göring HHH, Terwillinger JD: 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* 2000; **66**: 1310–1327.

Appendix A

Suppose that the markers A and B are in Hardy–Weinberg equilibrium. Then $Ex_{Ai}=Ex_{Bi}=Ez_{Ai}=Ez_{Bi}=0$. We first can show the following equations

$$\begin{aligned}
&E(x_{Ai}^2)=\mathrm{Cov}(x_{Ai},x_{Ai})=2P_aP_A, E(x_{Bi}x_{Ai})=\mathrm{Cov}(x_{Bi},x_{Ai})=2D_{AB},\\
&E(z_{Ai}x_{Ai})=\mathrm{Cov}(z_{Ai},x_{Ai})=0, E(z_{Bi}x_{Ai})=\mathrm{Cov}(z_{Bi},x_{Ai})=0,\\
&E(x_{Bi}^2)=\mathrm{Cov}(x_{Bi},x_{Bi})=2P_bP_B,\ E(z_{Ai}x_{Bi})=\mathrm{Cov}(z_{Ai},x_{Bi})=0,\\
&E(z_{Bi}x_{Bi})=\mathrm{Cov}(z_{Bi}x_{Bi})=0, E(z_{Ai}^2)=\mathrm{Cov}(z_{Ai},z_{Ai})=P_a^2P_A^2,\\
&E(z_{Bi}z_{Ai})=\mathrm{Cov}(z_{Bi},z_{Ai})=D_{AB}^2, E(z_{Bi}^2)=Cov(z_{Bi},z_{Bi})=P_b^2P_B^2.
\end{aligned} \tag{8}$$

In the following, we are going to show the first two of the above equations. The other equations can be shown by similar calculations. Actually, we have

$$\begin{aligned}
&E(x_{Ai}^2)=4P_a^2P_A^2+(P_a\quad P_A)^2 2P_aP_A+4P_A^2P_a^2=2P_aP_A,\\
&E(x_{Bi}x_{Ai})=2P_a[2P_bP(AA,BB)+(P_b\quad P_B)P(AA,Bb)+(\ \ 2P_B)P(AA,bb)]\\
&+(P_a\quad P_A)[2P_bP(Aa,BB)+(P_b\quad P_B)P(Aa,Bb)+(\ \ 2P_B)P(Aa,bb)]\\
&+(\ \ 2P_A)[2P_bP(aa,BB)+(P_b\quad P_B)P(aa,Bb)+(\ \ 2P_B)P(aa,bb)]\\
&=2P_a[2P_bP(AB)^2+(P_b\quad P_B)2P(AB)P(Ab)\quad 2P_BP(Ab)^2]+\\
&(P_a\quad P_A)[2P_b\cdot 2P(AB)P(aB)+(P_b\quad P_B)[2P(AB)P(ab)+2P(Ab)P(aB)]\\
&2P_B\cdot 2P(Ab)P(ab)]\quad 2P_A[2P_bP(aB)^2+(P_b\quad P_B)2P(aB)P(ab)\quad 2P_BP(ab)^2]\\
&=4P_aP_A[P_bP(AB)\quad P_BP(Ab)]+2(P_a\quad P_A)[P_b[P(AB)P_a+P_AP(aB)]\\
&\quad P_B[P_AP(ab)+P_aP(Ab)]]\quad 4P_aP_A[P_bP(aB)\quad P_BP(ab)]\\
&=2D_{AB}.
\end{aligned}$$

When the sample size n is large enough, the large number law leads to

$$\frac{1}{n}X^{\tau}X=\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} n & x_{Ai} & x_{Bi} & z_{Ai} & z_{Bi}\\ x_{Ai} & x_{Ai}^2 & x_{Bi}x_{Ai} & z_{Ai}x_{Ai} & z_{Bi}x_{Ai}\\ x_{Bi} & x_{Ai}x_{Bi} & x_{Bi}^2 & z_{Ai}x_{Bi} & z_{Bi}x_{Bi}\\ z_{Ai} & x_{Ai}z_{Ai} & x_{Bi}z_{Ai} & z_{Ai}^2 & z_{Bi}z_{Ai}\\ z_{Bi} & x_{Ai}z_{Bi} & x_{Bi}z_{Bi} & z_{Ai}z_{Bi} & z_{Bi}^2 \end{pmatrix}$$

$$\approx\begin{pmatrix} 1 & Ex_{A1} & Ex_{B1} & Ez_{A1} & Ez_{B1}\\ Ex_{A1} & Ex_{A1}^2 & Ex_{B1}x_{A1} & Ez_{A1}x_{A1} & Ez_{B1}x_{A1}\\ Ex_{B1} & Ex_{A1}x_{B1} & Ex_{B1}^2 & Ez_{A1}x_{B1} & Ez_{B1}x_{B1}\\ Ez_{A1} & Ex_{A1}z_{A1} & Ex_{B1}z_{A1} & Ez_{A1}^2 & Ez_{B1}z_{A1}\\ Ez_{B1} & Ex_{A1}z_{B1} & Ex_{B1}z_{B1} & Ez_{A1}z_{B1} & Ez_{B1}^2 \end{pmatrix}$$

$$=\begin{pmatrix} 1 & 0 & 0 & 0 & 0\\ 0 & 2P_aP_A & 2D_{AB} & 0 & 0\\ 0 & 2D_{AB} & 2P_bP_B & 0 & 0\\ 0 & 0 & 0 & P_a^2P_A^2 & D_{AB}^2\\ 0 & 0 & 0 & D_{AB}^2 & P_b^2P_B^2 \end{pmatrix} \tag{9}$$

This implies that the coefficients are approximately given by $\hat{\boldsymbol{\beta}}\approx\sum_{i=1}^{n}y_i/n$, and

$$\begin{pmatrix}\hat{\alpha}_A\\ \hat{\alpha}_B\end{pmatrix}\approx\frac{1}{2n}\begin{pmatrix}P_aP_A & D_{AB}\\ D_{AB} & P_bP_B\end{pmatrix}^{1}\begin{pmatrix}\sum_{i=1}^{n}x_{Ai}y_i\\ \sum_{i=1}^{n}x_{Bi}y_i\end{pmatrix}$$

$$\approx\begin{pmatrix}\sum_{i=1}^{n}x_{Ai}^2 & \sum_{i=1}^{n}x_{Ai}x_{Bi}\\ \sum_{i=1}^{n}x_{Ai}x_{Bi} & \sum_{i=1}^{n}x_{Bi}^2\end{pmatrix}^{1}\begin{pmatrix}\sum_{i=1}^{n}x_{Ai}y_i\\ \sum_{i=1}^{n}x_{Bi}y_i\end{pmatrix}$$

$$\begin{pmatrix}\hat{\delta}_A\\ \hat{\delta}_B\end{pmatrix}\approx\frac{1}{n}\begin{pmatrix}P_a^2P_A^2 & D_{AB}^2\\ D_{AB}^2 & P_b^2P_B^2\end{pmatrix}^{1}\begin{pmatrix}\sum_{i=1}^{n}z_{Ai}y_i\\ \sum_{i=1}^{n}z_{Bi}y_i\end{pmatrix}$$

$$\approx\begin{pmatrix}\sum_{i=1}^{n}z_{Ai}^2 & \sum_{i=1}^{n}z_{Ai}z_{Bi}\\ \sum_{i=1}^{n}z_{Ai}z_{Bi} & \sum_{i=1}^{n}z_{Bi}^2\end{pmatrix}^{1}\begin{pmatrix}\sum_{i=1}^{n}z_{Ai}y_i\\ \sum_{i=1}^{n}z_{Bi}y_i\end{pmatrix}.$$

If the marker A and marker B are in linkage equilibrium, i.e., $D_{AB}=0$, then $\sum_{i=1}^{n}x_{Ai}x_{Bi}/n\approx 0$ and $\sum_{i=1}^{n}z_{Ai}z_{Bi}/n\approx 0$. This will lead to equations in (2).

Appendix B

Notice that we have the following variance-covariance equations from model (1)

$$\mathrm{Cov}\begin{pmatrix}(x_{Ai},x_{Ai}) & (x_{Bi},x_{Ai}) & (z_{Ai},x_{Ai}) & (z_{Bi},x_{Ai})\\ (x_{Ai},x_{Bi}) & (x_{Bi},x_{Bi}) & (z_{Ai},x_{Bi}) & (z_{Bi},x_{Bi})\\ (x_{Ai},z_{Ai}) & (x_{Bi},z_{Ai}) & (z_{Ai},z_{Ai}) & (z_{Bi},z_{Ai})\\ (x_{Ai},z_{Bi}) & (x_{Bi},z_{Bi}) & (z_{Ai},z_{Bi}) & (z_{Bi},z_{Bi})\end{pmatrix}\begin{pmatrix}\alpha_A\\ \alpha_B\\ \delta_A\\ \delta_B\end{pmatrix}=\mathrm{Cov}\begin{pmatrix}(y_i,x_{Ai})\\ (y_i,x_{Bi})\\ (y_i,z_{Ai})\\ (y_i,z_{Bi})\end{pmatrix}. \tag{10}$$

The elements of the variance-covariance matrix on the left-hand side of the above equation are given in equations (8). For the elements on the right-hand side, we can show that

$$\begin{aligned}&\mathrm{Cov}(y_i,x_{Ai})=2D_{AQ}\alpha_Q,\mathrm{Cov}(y_i,x_{Bi})=2D_{QB}\alpha_Q,\\ &\mathrm{Cov}(y_i,z_{Ai})=D^2_{AQ}\delta_Q,\mathrm{Cov}(y_i,z_{Bi})=D^2_{QB}\delta_Q.\end{aligned} \tag{11}$$

In the following, we are going to show the first one of the above equations. The rest can be shown in the same way.

Appendix C

Using equations (4), (5), (6) and (9), the non-centrality parameter is

$$\begin{aligned}\lambda_{AB,ad}&\approx\frac{1}{2\sigma^2}(H\mu)^{\tau}[H[X^{\tau}X]^{-1}H^{\tau}]^{-1}H\mu,\\ &=\frac{n}{2\sigma^2}[2(\alpha_A,\alpha_B)\begin{pmatrix}P_aP_A & D_{AB}\\ D_{AB} & P_bP_B\end{pmatrix}\begin{pmatrix}\alpha_A\\ \alpha_B\end{pmatrix}+(\delta_A,\delta_B)\begin{pmatrix}P_a^2P_A^2 & D^2_{AB}\\ D^2_{AB} & P_b^2P_B^2\end{pmatrix}\begin{pmatrix}\delta_A\\ \delta_B\end{pmatrix}]\\ &=\frac{n}{2\sigma^2}[2\alpha_Q^2(D_{AQ},D_{QB})\begin{pmatrix}P_aP_A & D_{AB}\\ D_{AB} & P_bP_B\end{pmatrix}^{-1}\begin{pmatrix}D_{AQ}\\ D_{QB}\end{pmatrix}\\ &+\delta_Q^2(D^2_{AQ},D^2_{QB})\begin{pmatrix}P_a^2P_A^2 & D^2_{AB}\\ D^2_{AB} & P_b^2P_B^2\end{pmatrix}^{-1}\begin{pmatrix}D^2_{AQ}\\ D^2_{QB}\end{pmatrix}],\end{aligned}$$

which is equal to that in (7) by using equations $\sigma^2_{ga}=2q_1q_2\alpha_Q^2$ and $\sigma^2_{gd}=q_1^2q_2^2\delta_Q^2$.

For the first equation, we have

$$\begin{aligned}\mathrm{Cov}(y_i,x_{Ai})&=\mu_{11}[2P_aP(AQ_1)^2+2(P_a\quad P_A)P(AQ_1)P(aQ_1)\quad 2P_AP(aQ_1)^2]\\ &+\mu_{12}[2P_a\cdot 2P(AQ_1)P(AQ_2)+(P_a\quad P_A)[2P(AQ_1)P(aQ_2)\\ &+2P(AQ_2)P(aQ_1)\quad 2P_A\cdot 2P(aQ_1)P(aQ_2)]\\ &+\mu_{22}[2P_aP(AQ_2)^2+2(P_a\quad P_A)P(AQ_2)P(aQ_2)\quad 2P_AP(aQ_2)^2]\\ &=2\mu_{11}[P_aP(AQ_1)q_1\quad P_AP(aQ_1)q_1]+2\mu_{12}[P_a[P(AQ_1)q_2+P(AQ_2)q_1]\\ &\quad P_A[P(aQ_2)q_1+P(aQ_1)q_2]]+2\mu_{22}[P_aP(AQ_2)q_2\quad P_AP(aQ_2)q_2]\\ &=2D_{AQ}\alpha_Q.\end{aligned}$$

Plugging equations (8) and (11) into equation (10), we have

$$\begin{pmatrix}2P_aP_A & 2D_{AB} & 0 & 0\\ 2D_{AB} & 2P_bP_B & 0 & 0\\ 0 & 0 & P_a^2P_A^2 & D^2_{AB}\\ 0 & 0 & D^2_{AB} & P_b^2P_B^2\end{pmatrix}\begin{pmatrix}\alpha_A\\ \alpha_B\\ \delta_A\\ \delta_B\end{pmatrix}=\begin{pmatrix}2D_{AQ}\alpha_Q\\ 2D_{QB}\alpha_Q\\ D^2_{AQ}\delta_Q\\ D^2_{QB}\delta_Q\end{pmatrix}.$$

Hence, one may get equations (4) and (5).