HIGH RESOLUTION LINKAGE AND ASSOCIATION STUDY OF QUANTITATIVE TRAIT LOCI

A Dissertation

by

JEESUN JUNG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Statistics

HIGH RESOLUTION LINKAGE AND ASSOCIATION STUDY OF QUANTITATIVE TRAIT LOCI

A Dissertation

by

JEESUN JUNG

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Ruzong Fan (Chair of Committee)	
P. Fred Dahm (Member)	Naisyin Wang (Member)
Sing-Hoi Sze (Member)	Michael T. Longnecker (Head of Department)

August 2004

Major Subject: Statistics

ABSTRACT

High Resolution Linkage and Association Study
of Quantitative Trait Loci. (August 2004)
Jeesun Jung , B.S., Inje University, Korea;
M.A., Yonsei University, Korea
Chair of Advisory Committee: Dr. Ruzong Fan

As a large number of single nucleotide polymorphisms (SNPs) and microsatellite markers are available, high resolution mapping employing multiple markers or multiple allele markers is an important step to identify quantitative trait locus (QTL) of complex human disease. For many complex diseases, quantitative phenotype values contain more information than dichotomous traits do.

Much research has been done on conducting high resolution mapping using information of linkage and linkage disequilibrium. The most commonly employed approaches for mapping QTL are pedigree-based linkage analysis and population-based association analysis. As one of the methods dealing with multiple alleles markers, mixed models are developed to work out family-based association study with the information of transmitted allele and nontransmitted allele from one parent to offspring.

For multiple markers, variance component models are proposed to perform association study and linkage analysis simultaneously. Linkage analysis provides suggestive linkage based on a broad chromosome region and is robust to population admixtures. One the other hand, allelic association due to linkage disequilibrium (LD) usually operates over very short genetic distance, but is affected by population stratification. Combining both approaches plays a synergistic role in overcoming their limitations and in increasing the efficiency and effectiveness of gene mapping.

To My parents

ACKNOWLEDGMENTS

As I have completed my Ph.D. in statistics at Texas A&M University, I would like to express my gratitude to many people. I would never have been able to complete my Ph.D. without their encouragement. My special appreciation goes to my advisor, Dr. Ruzong Fan, who has led me to the new field of statistical genetics. I met him almost three years ago when he came to Texas A&M University as an assistant professor. As my mentor, he gave me the unique opportunity to gain much knowledge on genetic mapping. I think that being his student was the best decision I have ever made.

I would like to thank Dr. P. Fred Dahm, Dr. Naisyin Wang and Dr. Sing-Hoi Sze for serving on my dissertation committee and posing many insightful questions. I appreciate that Dr. Michael T. Longnecker always gave me good advice about my work.

I have lots of colleagues to express my thanks to, especially Kyeong Eun Lee, Ho-jin Lee and Joon Jin Song for their friendship, and Gosia Leyk Williams and Iryna Lobach for being my office mates and sharing such great times. There is one person, Sunghoon Chung who deserves to get my sincere gratitude for endless support and encouragement.

Finally, it would be impossible to have my research career without my parents' love and support, as well as help from my sister and brother. This dissertation is dedicated to them.

To all of you, thank you.

TABLE OF CONTENTS

CHAPTER Page

Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

1.1. General Description of Genetic Mapping

There have been lots of efforts to develop methodologies in order to find locations of Quantitative Trait Loci (QTL). For many human complex diseases, quantitative phenotypic values contain more information than dichotomous traits do. They can provide effective descriptions of diseases such as asthma, type II diabetes, learning difficulties, and osteoporosis. Quantitative trait value is affected by more than one gene as well as by environment effect. With this reason, it is not easy to localize QTL on chromosome. The most commonly employed approaches for mapping QTL of human complex diseases are pedigree-based linkage analysis and population-based association study.

1.1.1. Transmission Disequilibrium Test

Transmission Disequilibrium Test (TDT) was first introduced by Spielman et al. (1993) to test the presence of both linkage and linkage disequilibrium (LD) between a marker and putative disease locus when the marker locus and the hypothetical disease locus are linked or are in linkage disequilibrium. The TDT, as a model free method, is based on the unequal probability of transmission of different marker allele from parents to the affected offspring. This unequal pattern of transmission gives the evidence that the marker and disease locus are tightly linked or in LD. With the concept, lots of methods have been developed to test whether a marker allele exhibits

---

The format and style of this dissertation follows that of *Biometrics*.

transmission disequilibrium with a disease. But there are several possible drawbacks of TDT. It is positive only if both linkage and linkage disequilibrium are present. When the sibship observed are related, it is difficult to find out if there is evidence for linkage disequilibrium in addition to linkage.

1.1.2. Linkage Analysis

The most widely used method, linkage analysis, is developed from the methodology of Haseman and Elston (1972), as a family-based method. Linkage analysis exploits sharing allele identical-by-descent (IBD) which is a measure of genetic similarity between pairs of relatives. IBD is a function of recombination fraction which is a measure of genetic distance. The idea of linkage analysis is that the smaller the amount of recombinations observed between genes, i.e. the more tightly linked they are, the more possible they lie on a chromosome. Using the idea, lots of models such as variance component model, Haseman and Elston method, have been proposed to conduct linkage analysis. However, it is difficult to detect recombination events between closely spaced(< 2.5cM) loci since there is a limited number of meiosis occurring. Therefore, linkage analysis is usually proper for broad chromosome region mapping($\leq$ 10cM), but is not appropriate for high resolution mapping(≤ 2.5cM).

1.1.3. Linkage Disequilibrium Analysis

The other popular mapping tool is association analysis due to linkage disequilibrium that is a tendency of alleles to be inherited together more often than would be expected under random segregation. It is also called linkage disequilibrium mapping (LDM). LD mapping is based on both population data and pedigree data; it uses historical recombination events between genetic loci when non-random association of alleles at genetic loci was introduced into a population. LD can work over short

map distances, and can increase mapping precision in high resolution mapping. However the LD mapping largely depends on the level of LD, and its power to detect the putative QTL decays rapidly as the distance between the marker and putative QTL increases. Therefore, the allelic association study is useful to operate only over very short distance of loci. The most serious disadvantage is that the level of LD is sensitive to population stratification, although LD mapping can increase resolution in dissecting genetic traits when the association between markers and trait loci is introduced by events such as mutations at trait.

1.2. Literature Review

The Transmission Disequilibrium Test (TDT) developed by Spielman et al, (1993) is a powerful family-based test of linkage and a test of association. Sham and Curtis (1995) derived transmission probabilities for a logistic regression model with a multi-allele marker locus linked to a single disease locus. Allison (1997) extended the TDT method to quantitative traits by investigating the difference between average quantitative trait values of offspring with different alleles transmitted from heterozygous parents. Rabinowitz (1997) developed the TDT without parametric assumptions on the distribution of the quantitative traits. Xiong et al. (1998) generalized TDT which is allowed for multi-allelic loci. A disadvantage of all TDT methods is that they can detect linkage between the marker locus and the disease trait only if there is an association between the disease locus and alleles at the linked locus. George et al. (1999) proposed a regression-based TDT method which is based on regressing the trait on the parental transmission of a marker allele with no restriction on either the family structure sampled or the affected status of individuals in the pedigree. Zhu and Elston (2000,2001) also developed a TDT method for quantitative traits by defining a

linear transformation. Fan et al. (2002) explored linear regression models to detect linkage in the presence of association between a multi-allele locus and a disease locus for trio families. The methods are not valid for general nuclear families with more than one offspring, because they do not consider the correlation of offspring's trait values which are not independent. Fan and Xiong (2003) proposed mixed model to perform linkage and association studies for nuclear families with any number of offspring. The mean structure and variance-covariance structures in the mixed model are applied for bi-allele markers. Fan and Jung (2003) extended the mixed model to use a multiple alleles marker.

One of the best known approaches of sib-pair analysis is Haseman and Elston method (1972) which was developed to detect linkage between a quantitative trait and a marker. Linkage approach of Haseman and Elston (1972) exploited sharing allele identical-by-decent (IBD) to carry out regression of the squared trait differences of trait values between sib pairs. Haseman and Elston method (1972) was extended to allow all pedigree members (Amos et al., 1989). Amos (1994) developed a mixed-effects variance components approach for evaluating covariate effects, as well as evidence for genetic linkage to a single trait-affecting locus from pedigrees.

A simple interval-mapping approach to linkage analysis of quantitative traits, based on the sib-pair method of Haseman and Elston (1972), was proposed by Fulker and Cardon (1994). This approach provided not only useful information regarding the location of QTL, but also the valuable improvement in power over that of Haseman and Elston. The sib-pair interval-mapping procedure of Fulker and Cardon (1994) is extended to take account of all available markers information simultaneously on a chromosome (Fulker et al., 1995). The multipoint interval mapping increases power in dense mapping and is more accurate under conditions of variable marker information. Almasy and Blangero (1998) carried out multipoint mapping based on general

pedigrees. The variance component model proposed by Almasy and Blangero (1998) is more powerful than Haseman-Elston regression. Pratt et al. (2000) proposed variance component model that accounts for both additive and dominant variances to calculate covariance of trait between relatives in an exact multipoint quantitative trait linkage analysis.

Linkage disequilibrium mapping was also suggested for genome-wide screens (Xiong and Jin, 1997). Cardon (2000) proposed a multiple regression model to analyze very large number of SNPs. The International SNP Map Working Group (2001) has led to a novel approach of linkage disequilibrium (LD) mapping. Xiong et al. (1998) presented multiple regression for LD mapping and proposed two strategies to increase the probability of detecting LD. Fan and Xiong (2002) proposed a linear regression method based on population data in order to conduct LD analysis with two flanking markers.

Recently, the interests in joint LD and linkage mapping have been occurring. Almasy et al. (1999) proposed variance component models in QTL detection using combined linkage and LD analysis. Fulker et al. (1999) also combined both approaches based on sib pairs using variance component methods. Sham et al. (2000) performed analytical analyses of linkage versus association mapping of quantitative traits for sibship data in terms of power. Abecasis et al. (2000, 2001) generalized the method of Fulker et al. (1999) to apply for nuclear families and general pedigrees. Wu et al. (2002) made use of mixture models in joint linkage and LD mapping. Almost all research has employed only one marker. Since dense marker maps such as single nucleotide polymorphisms (SNPs) have been available, high resolution multiple markers mapping is needed. Fan and Xiong (2002) used two flanking markers to perform high resolution LD mapping with linear model, which applies to only data of population. Variance component models are proposed to combine linkage and LD

mapping based on both population and pedigree data. (Fan and Xiong, 2003; Fan and Jung, 2003).

1.3. Motivation and Overview of Dissertation

As large numbers of dense markers such as single-nucleotide polymorphisms (SNPs) and high resolution micro-satellite markers have been available, there is an urgent need to develop methodologies which deal with dense markers.

In certain situation, one may have data of multiple allele markers to be analyzed. One may collapse a multiple allele marker to be a bi-allelic marker in his/her study. However, this may not be a good idea since much information may be lost by combining different alleles that may have different roles. Moreover, different ways of collapsing multiple alleles can lead to different results which may cause different interpretation. With these reasons, it is necessary to build multi-allele markers mapping. In chapter II, mixed model is utilized to fit multi-allele markers for association study based on nuclear families with any number of offspring. Two types of nuclear families are considered in terms of genotype of parents. Using the information of the allele transmitted from parents to offspring for each type of nuclear families, mixed models are presented.

In views of statistics, the more information available, the better the results. A combined linkage and linkage disequilibrium analysis may give increased information and potentially more power to detect QTL. Separate method of either linkage analysis or LD mapping makes use of only one part of the available information and also have its own drawbacks. As we put both approaches together, the combination plays a synergistic role in overcoming their limitations and in increasing the efficiency and effectiveness of gene mapping. In chapter III, the combined mapping strategy is

introduced in the absence of parental information with two flanking markers. For late-onset disease such as Alzheimer's disease, heart disease, osteoporosis, and many forms of cancer, it is difficult to recruit parental data. In this case, one may perform sib pair or sibship analyses to study late-onset disorders. The new mapping method is the variance component model which integrates the linkage information in variance-covariance matrices and LD information in the mean coefficient of the linear model.

In chapter IV, we extend the combined mapping from two flanking markers to multiple markers. The objective is to build models which may fully use marker information for association mapping of QTL in the presence of prior linkage. Based on the information of markers, a multi-point interval mapping method is provided to build variance component model. The unified analysis in chapter IV is applied to both family with parental data and population data.

Finally, chapter V discusses the conclusions of our research with some open problems for further challenging investigation and discussion.

CHAPTER II

ASSOCIATION STUDIES FOR A MULTI-ALLELE MARKER*

2.1. Introduction

There has been a considerable interest in association study using transmission disequilibrium test (TDT) between a quantitative trait locus (QTL) and a marker locus. The TDT of Spielman et al. (1993) was originally introduced to test linkage between a qualitative trait and a marker. Allison (1997) and Xiong et al. (1998) extended the TDT procedure to quantitative traits. George et al. (1999) presented linear regression models for TDT by regressing the trait on the parental transmission of an allele of interest. This method can be applied to general pedigree structures. Zhu and Elston (2000, 2001) extended the method of George et al. (1999), and proposed better test statistics in detecting linkage and association. Fan and Xiong (2003) explored mixed models to perform linkage and association studies. The mixed model accommodated bi-allelic marker of nuclear families with any number of offspring. In certain circumstances, one may encounter the data of multiple allele markers such as micro-satellites. One may combine a multiple allele marker to be a bi-allelic marker as the purpose of analysis, but this may not be a good method because it may cause loss of much information. In addition, different ways to combine a multiple allele marker can lead to different results which make different interpretation possible. With these reasons, we need to develop methods to fit multi-allele markers in order to carry out association study.

---

*Reprinted with permission from "Association Studies of QTL for Multi-Allele Markers by Mixed Models" by Ruzong Fan, Jeesun Jung, 2002. *Human Heredity*, Vol. 54, 132–150. by S. Karger AG Basel.

Fan et al. (2003) proposed models and their test to perform association and linkage between a QTL and a multi-allele marker locus for trio families. Trio families consist of two parents and one single offspring. The methods of Fan et al. (2003) are not working for general nuclear families with more than one offspring, since the methods do not consider correlation of trait values of offspring that are not independent. To construct valid test statistics and models, one needs to consider the variance-covariance structure of trait values of offspring, as well as the mean structure under the normal assumption.

In this chapter, mixed models are introduced to investigate the association between a QTL and a multiple allele marker in terms of two types of nuclear families data. One is nuclear family with at least one heterozygous parent, the other is general nuclear family with no restriction on genotypes of parents. The conditional mean and conditional variance-covariance matrix of trait values of offspring for each type of nuclear families are derived. The theoretic basis is the difference of conditional means given information of a transmitted allele from heterozygous parents. The differences would give evidence that the allele is associated with putatitive quantitative trait locus. For a multiple allele marker, the number of parameters can be too large in data of nuclear-family with at least one heterozygous parent. Under the assumption of tight linkage between the trait locus and the interesting marker, the number of parameters can be significantly reduced by approximations. Test statistics based on the related conditional mean and conditional variance-covariance structures are derived. The non-centrality parameters of their test statistics are calculated to show the merits of the proposed methods in terms of power and sample size. The proposed models are used to analyze chromosome 4 and 16 data of the Oxford asthma data (Genetic Analysis Workshop 12)

2.2. Methods

We consider one quantitative trait locus (QTL) Q which has two alleles Q_1 and Q_2 with frequencies q_1 and q_2, respectively. Assume that the expected phenotypic trait value of a person with genotype Q_rQ_s is $\nu + \mu_{rs}, r, s = 1, 2$, where ν is overall mean and μ_{rs} is the effect of genotype Q_rQ_s, obviously $\mu_{12} = \mu_{21}$. There are m alleles $M_1, \cdots, M_m$ typed at the marker locus M, each M_i allele has frequency p_i, $i = 1, \cdots, m$. Suppose that a marker locus M is linked to the trait locus Q. Denote the recombination fraction between the marker locus M and the trait locus Q by θ. The haplotype frequency is denoted by h_{ri} for haplotype $Q_rM_i, r = 1, 2, i = 1, \cdots, m$. If $h_{ri} = q_rp_i$ for all r and i, the trait locus Q and the marker M are in linkage equilibrium. Otherwise, the trait locus Q and the marker M are in LD or association. The measure of LD between the trait allele Q_1 and the marker allele M_i is defined by $\delta_i = h_{1i} - q_1p_i, i = 1, \cdots, m$. Since $\sum_{i=1}^{m} \delta_i = 0$, one of $\delta_1, \cdots, \delta_m$ can be expressed by others, e.g.,$\delta_m = -\sum_{i=1}^{m-1} \delta_i$.

Let Y be the phenotypic trait variable decomposed into $Y = \nu + g + G + e$, where ν is overall mean, g is random major gene effect. Polygenic effect G has normal distribution with mean 0 and variance σ_G^2, and sampling error e is distributed as normal $N(0, \sigma_e^2)$. These g, G, and e are independent. If an individual has genotype Q_sQ_r at the trait locus, then $\mathrm{E}(g|Q_sQ_r) = \mu_{rs}$. Let TQ denote the abbreviation of "transmitted quantitative trait allele". We have the conditional mean given information of transmitted allele as following $\mathrm{E}[Y|TQ = Q_r] = \nu + \sum_{s=1}^{2} \mu_{rs}q_s = \nu + \mu_r$. Let $P(M_i, M_j)$ be the probability of an offspring who receives marker allele M_i from his/her heterozygous parent but not alleles M_j. That is $P(M_i, M_j) = P(M_j, M_i) = p_ip_j$. Let $P(Q_rM_i, M_j)$ be the probability of a child who receives haplotype Q_rM_i from his/her heterozygous parent but not alleles M_j. It can

be shown as $P(Q_r M_i, M_j) = (1-\theta)h_{ri}p_j + \theta h_{rj}p_i$.

2.2.1. Heterozygous Parent Data

For a family with two parents and at least one offspring, we assume that at least one parent is heterozygous at the marker locus M. Moreover, assume we may infer clearly the transmission of parental marker alleles to the offspring. If both parents and an offspring have the same genotype $M_iM_j, i \neq j$, it is impossible to tell which parent transmits which allele to the offspring, and hence the data can not be used in analysis. Actually, this is the only type of data which needs to be excluded. For a bi-allelic marker, one needs to exclude the heterozygous offspring of a mating heterozygous $\times$ heterozygous (Fan and Xiong 2002; George at al. 1999; Zhu and Elston 2000, 2001). For a multi-allelic marker, any offspring from a mating $M_iM_i \times M_jM_k, j \neq k$ or $M_iM_j \times M_iM_k, i \neq j, i \neq k, j \neq k$ or a mating $M_iM_j \times M_lM_k$, $i \neq j, i \neq l, i \neq k, j \neq l, j \neq k, l \neq k$ can be included in analysis since one can infer clearly the transmission of parental marker alleles to the offspring. Hence, a heterozygous offspring of a mating heterozygous $\times$ heterozygous may not be necessarily excluded in case of multi-allelic marker unless both parents and offspring have exactly the same heterozygous genotype.

Let us look at a pedigree depicted in Figure 1. Assume that the genotype of the father at the marker locus is heterozygous $M_iM_j, i \neq j$. Moreover, the father transmits allele M_i to children $1, \cdots, k$, and transmits allele M_j to children $k{+}1, \cdots, n$. The quantitative trait value for offspring i is denoted by $y_i, i = 1, \cdots, n$. For the mother, we can perform similar analysis. If the mother is homozygous M_iM_i, every offspring receives an allele M_i from her and so she does not provide useful information (Spielman et al. 1993). If the mother is heterozygous, one should examine if an allele is transmitted to an offspring by the mother. Keeping all offspring with whom one

may infer clearly the transmission of allele from the mother and father, we use those data to develop following methods.

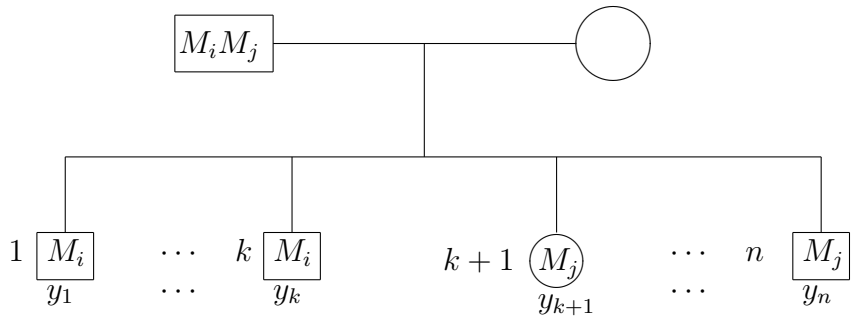

Fig. 1. A nuclear family with n offspring. Assume that the genotype of the father at the marker locus is heterozygous $M_iM_j, i \neq j$. Moreover, the father transmits allele M_i to kids $1, \cdots, k$, and transmits allele M_j to kids $k+1, \cdots, n$.

2.2.1.1. Mean and Variance-Covariance Structures

Let TM denote the abbreviation of "transmitted marker allele", and NM of "non-transmitted marker allele". Given that marker allele M_i is transmitted and allele M_j is not transmitted from the heterozygous father for children $1, \cdots, k$, the conditional expected mean can be calculated in the same way as equation (1) or (2) of Fan and Xiong (2002)

$$\alpha_{i,j} = \mathrm{E}\left[Y|TM = M_i, NM = M_j\right] = \nu + \sum_{r=1}^{2} \mu_r[(1-\theta)h_{ri}p_j + \theta h_{rj}p_i]/[p_ip_j]. \quad (2.1)$$

With the same way, the conditional expected mean of the children $k+1, \cdots, n$ in Figure 1 is

$$\alpha_{j,i} = \mathrm{E}\left[Y|TM = M_j, NM = M_i\right] = \nu + \sum_{r=1}^{2} \mu_r[(1-\theta)h_{rj}p_i + \theta h_{ri}p_j]/[p_ip_j]. \quad (2.2)$$

Using $h_{2i}p_j - h_{2j}p_i = (p_i - h_{1i})p_j - (p_j - h_{1j})p_i = -h_{1i}p_j + h_{1j}p_i$, we derive a difference between $\alpha_{i,j}$ and $\alpha_{j,i}$ as following:

$$\begin{aligned} \alpha_{i,j} - \alpha_{j,i} &= (1-2\theta)\sum_{r=1}^{2}\mu_r(h_{ri}p_j - h_{rj}p_i)/(p_ip_j) \\ &= (1-2\theta)(\mu_1-\mu_2)(h_{1i}p_j - h_{1j}p_i)/(p_ip_j) \\ &= (1-2\theta)(\mu_1-\mu_2)(\delta_ip_j - \delta_jp_i)/(p_ip_j). \end{aligned} \tag{2.3}$$

Assume that the trait locus Q is linked to the marker locus M, i.e., $0 \le \theta < 1/2$. The difference between conditional means is induced by $\delta_i p_j - \delta_j p_i \neq 0$, which implies at least one of δ_i and δ_j is not equal to 0. That shows the marker M is in LD with trait locus Q. Hence, one may construct statistics and models to test association in the presence of linkage between the marker M and the trait locus Q based on the difference (2.3).

To build valid test statistics and models, we need to calculate the variance-covariances of the trait values of offspring in nuclear families. In a similar manner as Appendix A of Fan and Xiong (2002), we may show that the conditional variance of trait value of the offspring $1, \cdots, k$ is $\sigma^2_{i,j} = \sigma^2_e + \sigma^2_G + \Sigma^2_{ij}$, where $\Sigma^2_{ij} = \sum_{r=1}^{2}\sum_{s=1}^{2}(\nu + \mu_{rs} - \alpha_{i,j})^2 q_s P(Q_r M_i, M_j)/P(M_i, M_j)$. Likewise, the conditional variance of trait values of the offspring $k+1, \cdots, n$ is $\sigma^2_{j,i} = \sigma^2_e + \sigma^2_G + \Sigma^2_{ji}$, where $\Sigma^2_{ji} = \sum_{r=1}^{2}\sum_{s=1}^{2}(\nu + \mu_{rs} - \alpha_{j,i})^2 q_s P(Q_r M_j, M_i)/P(M_j, M_i)$. For the conditional covariances, let us denote the expected conditional covariance between y_l $(l = 1, \cdots, k)$ and y_t $(t \neq l, t = 1, \cdots, k)$ by $\Sigma_{ij,ij}$, the expected conditional covariance between y_l $(l = 1, \cdots, k)$ and y_t $(t = k+1, \cdots, n)$ as $\Sigma_{ij,ji} = \Sigma_{ji,ij}$, and the expected conditional covariance between y_l $(l = k+1, \cdots, n)$ and y_t $(t \neq l, t = k+1, \cdots, n)$ as $\Sigma_{ji,ji}$. $\Sigma_{ij,ij}$ and $\Sigma_{ij,ji}$ are calculated in Appendix A.

On the other hand, we need to build model under the null hypothesis of no association in the presence of linkage. To do this, we need to calculate the mean and variance-covariance parameters. Under the assumption of linkage equilibrium between the marker locus and the trait locus, we show that $\alpha_{i,j} = \sum_{r=1}^{2}(\nu + \mu_r)q_r = \nu + \mu = \alpha, \sigma^2_{i,j} = \sigma^2, \Sigma_{ij,ij} = \Sigma_{ts}$ and $\Sigma_{ij,ji} = \Sigma_{td}$, which do not depend on subscripts i and j in Appendix B.

2.2.1.2. Parameter Reductions

In Subsection 2.2.1.1, we work out the mean and variance-covariance structures of siblings for a nuclear family. Although the structure is valid theoretically, the number of parameters can be very large for a multi-allele marker M. The number of mean parameters $\alpha_{i,j}$ is $m(m-1)$, and the number of variance-covariances $\sigma^2_{i,j}, \Sigma_{ij,ij}, \Sigma_{ij,ji}$ is $5[m(m-1)/2]$ for a marker M with m alleles. Hence, the total number of the parameters is $7m(m-1)/2$. For a marker with 3 alleles, the number of parameters is 21; for a marker with 4 alleles, the number of parameters is 42. One needs to reduce the number of parameters to build valid models and obtain their robust test statistics.

In a population, the presence of LD is usually the result of tight linkage between a trait locus and a marker locus (Falconer and Mackay 1996; Fan et al. 2002; Sham and Curtis 1995). Assume that the recombination fraction $\theta \approx 0$, i.e. there is tight linkage between the trait locus and the marker. In Appendix C, we show that approximately $\alpha_{i,j} \approx \alpha_i$, $\sigma^2_{i,j} \approx \sigma^2_i$ and $\Sigma_{ij,ij} \approx \Sigma_{i,i}$ only depend on subscript i, and the covariance $\Sigma_{ij,ji} \approx \Sigma_{i,j} = \Sigma_{j,i}$ depends on both i and j. Therefore, the expected conditional variance-covariance matrix of $y_l, l = 1, \cdots, n$, in Figure 1 can be expressed as

$$\begin{pmatrix} \sigma^2_{i,j} & \Sigma_{ij,ij} & \cdots & \Sigma_{ij,ij} & \Sigma_{ij,ji} & \cdots & \Sigma_{ij,ji} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{ij,ij} & \Sigma_{ij,ij} & \cdots & \sigma^2_{i,j} & \Sigma_{ij,ji} & \cdots & \Sigma_{ij,ji} \\ \Sigma_{ji,ij} & \Sigma_{ji,ij} & \cdots & \Sigma_{ji,ij} & \sigma^2_{j,i} & \cdots & \Sigma_{ji,ji} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{ji,ij} & \Sigma_{ji,ij} & \cdots & \Sigma_{ji,ij} & \Sigma_{ji,ji} & \cdots & \sigma^2_{j,i} \end{pmatrix} \approx \begin{pmatrix} \sigma^2_i & \Sigma_{i,i} & \cdots & \Sigma_{i,i} & \Sigma_{i,j} & \cdots & \Sigma_{i,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{i,i} & \Sigma_{i,i} & \cdots & \sigma^2_i & \Sigma_{i,j} & \cdots & \Sigma_{i,j} \\ \Sigma_{j,i} & \Sigma_{j,i} & \cdots & \Sigma_{j,i} & \sigma^2_j & \cdots & \Sigma_{j,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{j,i} & \Sigma_{j,i} & \cdots & \Sigma_{j,i} & \Sigma_{j,j} & \cdots & \sigma^2_j \end{pmatrix}.$$

With these parameter reductions, the number of mean parameters α_i is m, and the number of variance-covariance parameters $\sigma^2_i, \Sigma_{i,i}, \Sigma_{i,j}$ is $2m+m(m-1)/2$. Hence, the total number of the parameters is $3m + m(m-1)/2$. Such as in Fan and Xiong (2002), the number of parameters for a bi-allele marker is 7. For a marker with 3 alleles, the number of parameters is 12, and for a marker with 4 alleles, the number of parameters is 18. Therefore, the number of parameters can be significantly reduced under the assumption of tight linkage between the trait locus and the marker.

2.2.1.3. Mixed Model

Suppose that the data consist of nuclear families with I heterozygous parents. Each of them has at least one offspring. For each family, suppose that genotypes of both parents are typed at the marker locus M and at least one of the parents is heterozygous. For the offspring of each heterozygous parent, assume that one may clearly determine which allele at the marker locus M are transmitted from the heterozygous parent. A quantitative trait value of each offspring is observed.

For the l-th heterozygous parent, assume that the genotype at the marker locus is $M_iM_j, i \neq j$. Moreover, he/she has n_i offspring, and the offspring's trait values are listed as $y_{l1}, \cdots, y_{ln_l}$. Assume that the offspring consist of two parts: (1) k_l offspring have the fact that allele M_i is transmitted and allele M_j is not transmitted from their heterozygous parent, and their trait values are listed as $y_{l1}, \cdots, y_{lk_l}$; (2) the rest of the

offspring have the fact that allele M_i is not transmitted and allele M_j is transmitted from their heterozygous parent, and their trait values are listed as $y_{l,k_l+1}, \cdots, y_{ln_l}$.

Under the null hypothesis of no association in the presence of linkage between the trait locus Q and the marker locus M, one may use a multivariate linear model

$$y_{lu} = \nu + g_{lu} + G_{lu} + e_{lu}, u = 1, 2, \cdots, n_l, \text{ reduced model,} \tag{2.4}$$

where y_{lu} are normal variables with mean α and $n_l \times n_l$ variance-covariance matrix

$$V_l = \begin{pmatrix} \sigma^2 & \Sigma_{ts} & \cdots & \Sigma_{ts} & \Sigma_{td} & \cdots & \Sigma_{td} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{ts} & \Sigma_{ts} & \cdots & \sigma^2 & \Sigma_{td} & \cdots & \Sigma_{td} \\ \Sigma_{td} & \Sigma_{td} & \cdots & \Sigma_{td} & \sigma^2 & \cdots & \Sigma_{ts} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{td} & \Sigma_{td} & \cdots & \Sigma_{td} & \Sigma_{ts} & \cdots & \sigma^2 \end{pmatrix}.$$

Under the alternative hypothesis of association in the presence of linkage, one may use a full model

$$\begin{aligned} y_{lu} &= \nu + g_{lu}|_{(TM=M_i, NM=M_j)} + G_{lu} + e_{lu}, u = 1, 2, \cdots, k_l, \\ y_{lu} &= \nu + g_{lu}|_{(TM=M_j, NM=M_i)} + G_{lu} + e_{lu}, u = k_l + 1, \cdots, n_l. \end{aligned} \tag{2.5}$$

y_{lu} are normal variables with mean α_i for $u = 1, \cdots, k_l$ and mean α_j for $u = k_l + 1, \cdots, n_l$, and a variance-covariance matrix

$$\Gamma_l = \begin{pmatrix} \sigma_i^2 & \Sigma_{i,i} & \cdots & \Sigma_{i,i} & \Sigma_{i,j} & \cdots & \Sigma_{i,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{i,i} & \Sigma_{i,i} & \cdots & \sigma_i^2 & \Sigma_{i,j} & \cdots & \Sigma_{i,j} \\ \Sigma_{j,i} & \Sigma_{j,i} & \cdots & \Sigma_{j,i} & \sigma_j^2 & \cdots & \Sigma_{j,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{j,i} & \Sigma_{j,i} & \cdots & \Sigma_{j,i} & \Sigma_{j,j} & \cdots & \sigma_j^2 \end{pmatrix}.$$

Putting all data together, we may perform association studies based on reduced model and full model. Denote $n = \sum_{l=1}^{I} n_l, \vec{y}_l = (y_{l,1}, \cdots, y_{ln_l})^\tau, \vec{y} = (\vec{y_1}^\tau, \cdots, \vec{y_I}^\tau)^\tau$,

$V = \text{diag}(V_1, V_2, \cdots, V_I)$ and $\Gamma = \text{diag}(\Gamma_1, \Gamma_2, \cdots, \Gamma_I)$. Let I_n be the identity $n \times n$ matrix. In the reduced model, $\vec{y}$ is normal with mean αI_n and variance-covariance matrix V. In the full model, similarly $\vec{y}$ is normal with mean $X(\alpha_1, \cdots, \alpha_m)^\tau$ and variance-covariance matrix Γ, where X is an $n \times m$ design matrix based on model (2.5).

2.2.2. General Nuclear Family Data

2.2.2.1. Mean and Variance-Covariance Structures

Consider a sample of general nuclear families which consist of two parents with no restriction on parental genotype and at least one offspring each. For each parent-offspring pair, one first determines which allele is transmitted from the parent to the offspring. In the general nuclear family, we use a different approach from that in Section 2.2.1. For instance, we simply assume that an allele M_i is transmitted from a homozygous parent M_iM_i to any of his/her offspring, and ignore which one it is. If both parents and an offspring have the same genotype $M_iM_j, i \neq j$, we assume that one parent transmits M_i to the offspring and the other parent transmits M_j to the offspring. In this way for each parent-offspring pair, we may define an transmission of allele from the parent to the offspring. Putting all data together, we may arrange the trait values of offspring in a way as Table 1 in Fan et al. (2002). Hence, all data from a nuclear family can be used in analysis. Based on which marker allele is transmitted from a parent, the conditional mean $\beta_i = \text{E}\,(Y|TM = M_i)$ is calculated in Appendix D as following.

$$\begin{aligned} \beta_i &= \text{E}\,[Y|TM = M_i] \\ &= (1-\theta)\Big[(\nu+\mu_1)h_{1i} + (\nu+\mu_2)h_{2i}\Big]/p_i + \theta\alpha \end{aligned}$$

Therefore,

$$\begin{aligned}\frac{\beta_i - \alpha}{1-\theta} &= \Big[(\nu+\mu_1)h_{1i} + (\nu+\mu_2)h_{2i}\Big]/p_i - [(\nu+\mu_1)q_1 + (\nu+\mu_2)q_2] \\ &= (\mu_1 - \mu_2)\delta_i/p_i.\end{aligned}$$

The absence of association between trait locus Q and marker M, i.e., $\delta_i = 0$, means $\beta_i = \alpha$. This constitutes the basis to build models and to construct appropriate statistics to test the association between trait locus Q and marker M by comparing the estimates of parameters β_i and α. To build models, we need variance covariance structures of the trait values of offspring. In Appendix D, we calculate conditional variance $\sigma^2_{ir} = \mathrm{Var}(Y|TM = M_i)$. For two offspring of a nuclear family, let TM_1 be the abbreviation of "transmitted marker allele for child 1", and let TM_2 be the abbreviation of "transmitted marker allele for child 2". For $i \neq j$, the conditional covariance $\Sigma_{i,jr} = \mathrm{Cov}(Y_1, Y_2|TM_1 = M_i, TM_2 = M_j) = \Sigma_{ij,ji}$. The conditional covariance $\Sigma_{i,ir} = \mathrm{Cov}(Y_1, Y_2|TM_1 = M_i, TM_2 = M_i)$ is calculated.

2.2.2.2. Mixed Model

In this Subsection, we are going to build models and construct their statistics to test association between the trait locus Q and marker M to analyze general nuclear family data. We assume that there is at least one offspring for each nuclear family. For a homozygous parent with genotype M_iM_i at the marker M and n_l offspring, let the trait values of the offspring be $y_1, \cdots, y_{n_l}$. One may use a multivariate linear model for data analysis

$$y_u = \nu + g_u|_{(TM=M_i)} + G_u + e_u, u = 1, 2, \cdots, n_l, \tag{2.6}$$

where y_u are normal variables with mean β_i and $n_l \times n_l$ variance-covariance matrix

$$\begin{pmatrix} \sigma^2_{ir} & \Sigma_{i,ir} & \cdots & \Sigma_{i,ir} \\ \Sigma_{i,ir} & \sigma^2_{ir} & \cdots & \Sigma_{i,ir} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_{i,ir} & \Sigma_{i,ir} & \cdots & \sigma^2_{ir} \end{pmatrix}.$$

For a heterozygous parent with genotype $M_iM_j, i \neq j$ at the marker M and n_l offspring, let the trait values of the offspring be $y_1, \cdots, y_{n_l}$. Suppose that: (1) k_l offspring have the fact that allele M_i is transmitted and allele M_j is not transmitted from their heterozygous parent, and their trait values are listed as $y_1, \cdots, y_{k_l}$; (2) the rest of the offspring have the fact that allele M_i is not transmitted and allele M_j is transmitted from their heterozygous parent, and their trait values are listed as $y_{k_l+1}, \cdots, y_{n_l}$. One may use a model

$$\begin{aligned} y_u &= \nu + g_u|_{(TM=M_i)} + G_u + e_u, u = 1, 2, \cdots, k_l, \\ y_u &= \nu + g_u|_{(TM=M_j)} + G_u + e_u, u = k_l + 1, \cdots, n_l. \end{aligned} \tag{2.7}$$

y_u are normal variables with mean β_i for $u = 1, \cdots, k_l$ and mean β_j for $u = k_l + 1, \cdots, n_l$, and an $n_l \times n_l$ variance-covariance matrix

$$\begin{pmatrix} \sigma^2_{ir} & \Sigma_{i,ir} & \cdots & \Sigma_{i,ir} & \Sigma_{i,jr} & \cdots & \Sigma_{i,jr} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{i,ir} & \Sigma_{i,ir} & \cdots & \sigma^2_{ir} & \Sigma_{i,jr} & \cdots & \Sigma_{i,jr} \\ \Sigma_{j,ir} & \Sigma_{j,ir} & \cdots & \Sigma_{j,ir} & \sigma^2_{jr} & \cdots & \Sigma_{j,jr} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma_{j,ir} & \Sigma_{j,ir} & \cdots & \Sigma_{j,ir} & \Sigma_{j,jr} & \cdots & \sigma^2_{jr} \end{pmatrix}.$$

2.3. Test Statistics and Non-Centrality Parameter

2.3.1. Heterozygous Parent Data

Let $\hat{\alpha}_i, \hat{\sigma}^2_i, \hat{\Sigma}_{i,i}, \hat{\Sigma}_{i,j}$ be the maximum likelihood estimators of parameters $\alpha_i, \sigma^2_i, \Sigma_{i,i}, \Sigma_{i,j}$ of the full model (2.5). Then the estimate of $\gamma = (\alpha_1, \cdots, \alpha_m)^\tau$ is $\hat{\gamma} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_m)^\tau =$

$\left[X^\tau\hat{\Gamma}^{-1}X\right]^{-1}X^\tau\hat{\Gamma}^{-1}\vec{y}$. Assume that the sample size is large. In Appendix E, we show that the test statistic of the null hypothesis $H_0 : \alpha_1 = \cdots = \alpha_m$, is non-central $F(m-1, n-m)$ defined by (details are given in Appendix E)

$$F_{het} = \frac{(H\hat{\gamma})^\tau[H(X^\tau\hat{\Gamma}^{-1}X)^{-1}H^\tau]^{-1}H\hat{\gamma}/(m-1)}{\vec{y}^\tau[\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}X(X^\tau\hat{\Gamma}^{-1}X)^{-1}X^\tau\hat{\Gamma}^{-1}]\vec{y}/(n-m)}$$

, where

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

Here H is a $(m-1)\times m$ testing matrix. The non-centrality parameter of the test statistic F can be calculated by $\lambda_{het} \approx (H\gamma)^\tau[H(X^\tau\Gamma^{-1}X)^{-1}H^\tau]^{-1}H\gamma$. If $n_i = 1$ for each family, then there is only one single child in each family and the above formula can be simplified. Let $k_i, i = 1, 2, \cdots, m$ be the number of offspring who receive allele M_i from their heterozygous parents. In Appendix F, we show that the non-centrality parameter of the singleton test statistic $F_{het,singleton}$ is

$$\lambda_{het,singleton} \approx \sum_{i=2}^{m}(\alpha_1 - \alpha_i)^2 k_i/\sigma_i^2 - \Big[\sum_{i=2}^{m}(\alpha_1 - \alpha_i)k_i/\sigma_i^2\Big]^2/[\sum_{i=1}^{m} k_i/\sigma_i^2].$$

Assume that the data consist of both singleton families and sib-pair families. Suppose there are k_i singleton offspring who receive allele M_i from their heterozygous parents, k_{ii} $(i = 1, 2, \cdots, m)$ sib pairs who receive allele M_i from their heterozygous parents, and $k_{ij} = k_{ji}, i \neq j$ sib pairs whose one sib receives allele M_i from his/her heterozygous parent and the other receives allele M_j from the same heterozygous parent. In Appendix G, we obtain the matrix

$$X^\tau\Gamma^{-1}X = \mathrm{diag}\Big(\frac{k_1}{\sigma_1^2} + \frac{2k_{11}}{\sigma_1^2 + \Sigma_{1,1}}, \cdots, \frac{k_m}{\sigma_m^2} + \frac{2k_{mm}}{\sigma_m^2 + \Sigma_{m,m}}\Big) + X_3^\tau\Gamma_3^{-1}X_3,$$

where matrix X_3, sub-variance-covariance matrix Γ_3, and $X^\tau\Gamma_3^{-1}X$ are given in Ap-

pendix G. Inserting the above matrix to the formula λ_{het}, one may calculate the non-centrality parameter $\lambda_{het,singleton,sibs}$ of a test statistic $F_{het,singleton,sibs}$. For a bi-allele marker M, it is the same as that in Fan and Xiong (2002).

2.3.2. General Nuclear Family Data

For model introduced in Subsection 2.2.2.2, we may calculate the non-centrality parameter of statistic F_{Gen_Nuc} to test null hypothesis $H_0 : \beta_1 = \cdots = \beta_m$ in a similar manner. First, assume that each family has only one child. Let $k_i, i = 1, 2, \cdots, m$ be the number of offspring who receive allele M_i from their parents. We can show that the corresponding non-centrality parameter of a singleton test statistic $F_{Gen\_Nuc,singleton}$ is $\lambda_{Gen\_Nuc,singleton} \approx \sum_{i=2}^{m}(\beta_1 - \beta_i)^2 k_i/\sigma_{ir}^2 - \left[\sum_{i=2}^{m}(\beta_1 - \beta_i)k_i/\sigma_{ir}^2\right]^2 / [\sum_{i=1}^{m} k_i/\sigma_{ir}^2]$.

Second, the data consist of both singleton families and sib-pair families. Suppose there are k_i singleton offspring who receive allele M_i from their parents, k_{ii} $(i = 1, 2, \cdots, m)$ sib pairs who receive allele M_i from their parents, and $k_{ij} = k_{ji}, i \neq j$ sib pairs whose one sib receives allele M_i from his/her heterozygous parent and the other receives allele M_j from the same heterozygous parent. We may calculate the corresponding non-centrality parameter $\lambda_{Gen\_Nuc,singleton,sibs} \approx (H\beta)^\tau [H\Pi^{-1}H^\tau]^{-1} H\beta$ of a statistic $F_{Gen\_Nuc,singleton,sibs}$, where

$$\Pi = \mathrm{diag}\Big(\frac{k_1}{\sigma_{1r}^2} + \frac{2k_{11}}{\sigma_{1r}^2 + \Sigma_{1,1r}}, \cdots, \frac{k_m}{\sigma_{mr}^2} + \frac{2k_{mm}}{\sigma_{mr}^2 + \Sigma_{m,mr}}\Big) + \Pi_3, \text{ and}$$

$$\Pi_3 = \begin{pmatrix} \sum_{i\neq 1} \frac{k_{1i}\sigma_{ir}^2}{\sigma_{1r}^2\sigma_{ir}^2 - \Sigma_{1,ir}^2} & -\frac{k_{12}\Sigma_{1,2r}}{\sigma_{1r}^2\sigma_{2r}^2 - \Sigma_{1,2r}^2} & \cdots & -\frac{k_{1m}\Sigma_{1,mr}}{\sigma_{1r}^2\sigma_{mr}^2 - \Sigma_{1,mr}^2} \\ -\frac{k_{12}\Sigma_{1,2r}}{\sigma_{1r}^2\sigma_{2r}^2 - \Sigma_{1,2r}^2} & \sum_{i\neq 2} \frac{k_{2i}\sigma_{ir}^2}{\sigma_{2r}^2\sigma_{ir}^2 - \Sigma_{2,ir}^2} & \cdots & -\frac{k_{2m}\Sigma_{2,mr}}{\sigma_{2r}^2\sigma_{mr}^2 - \Sigma_{2,mr}^2} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{k_{1m}\Sigma_{1,mr}}{\sigma_{1r}^2\sigma_{mr}^2 - \Sigma_{1,mr}^2} & -\frac{k_{2m}\Sigma_{2,mr}}{\sigma_{2r}^2\sigma_{mr}^2 - \Sigma_{2,mr}^2} & \cdots & \sum_{i\neq m} \frac{k_{mi}\sigma_{ir}^2}{\sigma_{mr}^2\sigma_{ir}^2 - \Sigma_{m,ir}^2} \end{pmatrix}.$$

2.4. Power Comparison

Assume that $\nu = 0, \mu_{11} = a, \mu_{12} = \mu_{21} = d, \mu_{22} = -a$ in terms of the standard theory of quantitative genetics (Falconer and Mackay 1996). Let the additive variance be $\sigma_a^2 = 2q_1q_2(a+d(q_2-q_1))^2$, and the dominant variance be $\sigma_d^2 = (2q_1q_2d)^2$. Let the heritability be denoted by h^2, which is defined by $\sigma_a^2/(\sigma_a^2+\sigma_d^2+\sigma_e^2)$. In the history of a population, the disease genes are usually due to a mutation. Because of the evolutionary process, the haplotype frequencies h_{ri} change from generation to generation. The expected haplotype frequencies can be calculated by $\mathrm{E}[h_{ri}] = h_{ri}(0)e^{-\theta A}+p_iq_r(1-e^{-\theta A})$, where A is the age of the most recent mutation at the trait locus, $h_{ri}(0)$ is the initial haplotype frequencies of haplotypes Q_rM_i at the generation of occurrence of the mutation at the trait locus. If there is only a single mutation in the population, one may assume that $h_{11}(0) = q_1, h_{1i}(0) = 0$, and $h_{21}(0) = p_1-q_1 \geq 0, h_{2i}(0) = p_i, i = 2, \cdots, m$. Replacing h_{ri} in $P(Q_rM_i, M_j)$ by $\mathrm{E}[h_{ri}]$, we may calculate the approximations of the non-centrality parameters using the non-centrality parameters given in Section 2.3. To calculate the non-centrality parameters, we need parameter values such as the marker allele frequencies p_1 and p_2, trait allele frequencies q_1 and q_2, heritability h^2, mutation age A, haplotype frequencies h_{ri}, recombination fraction θ, additive effect a, dominant effect d, polygenic variance σ_G^2, and error variance σ_e^2.

Assume that the frequencies of marker alleles are evenly distributed. Figures 3 and 2 plot the power curves of $F_{het,singleton}$ and $F_{het,singleton,sibs}$ against the heritability at 0.05 significant level, for dominant and recessive traits for 2, 3 and 4 allele markers, respectively. In each graph of the two Figures, the total numbers of offspring for 2, 3 and 4 allele markers are the same. Hence, the comparison of the power is meaningful. It is clear from the 4 graphs of the two Figures 3 and 2 that the power of the test statistic using 4 allele marker is higher than that of the test statistic using 3 allele

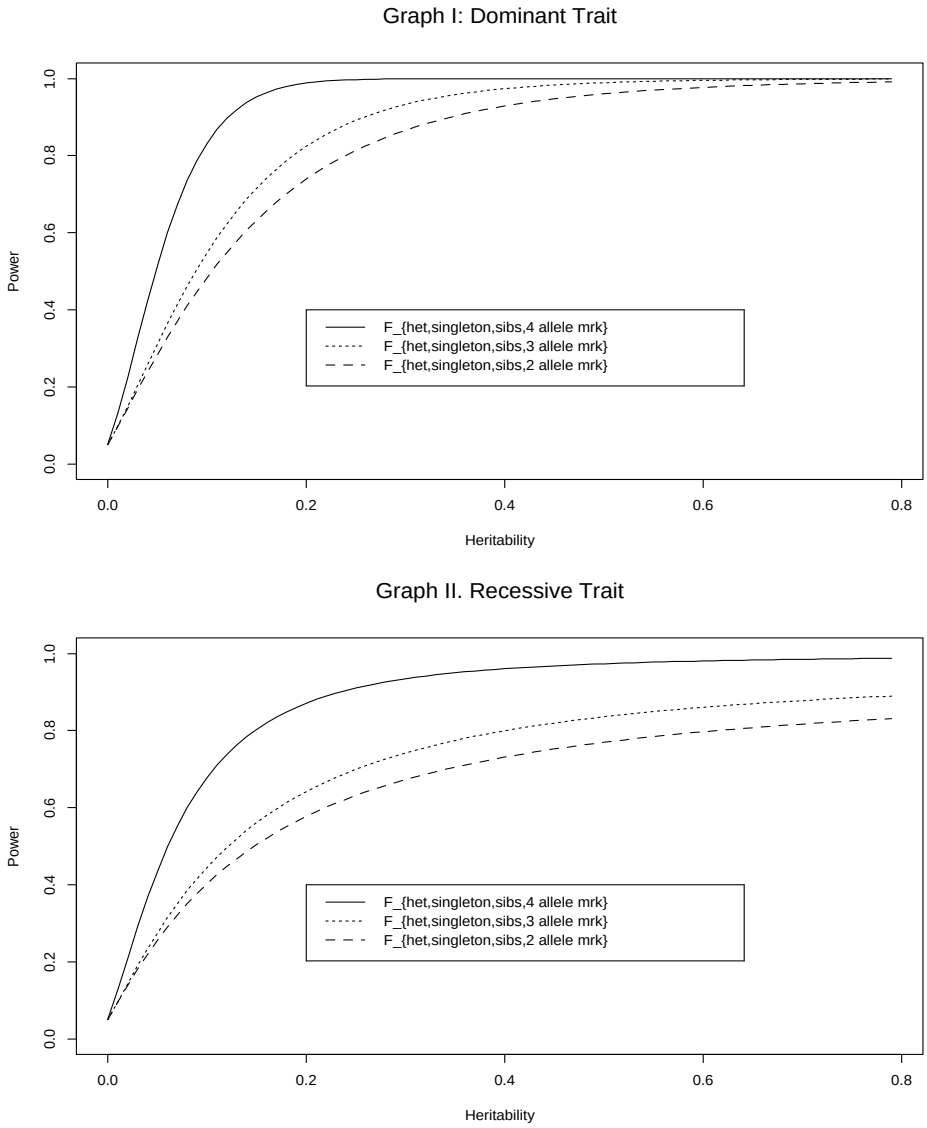

Fig. 2. Power curves of $F_{het,singleton,sibs}$ for 2, 3 and 4 allele markers against the heritability at 0.05 significant level, when $q_1 = 0.25, \sigma_G^2 = 0.75, A = 20, \theta = 0.005$ for a dominant trait $a = d = 1.0$, Graph I; and a recessive trait $a = 1.0$ and $d = -0.5$, Graph II. For a 2 allele marker, $p_1 = 0.50, k_i = 60, k_{ij} = 30, i, j = 1, 2$; For a 3 allele marker, $p_1 = 0.4, p_2 = 0.3, k_1 = 60, k_2 = k_3 = 30, k_{ij} = 15, i, j = 1, 2, 3$; For a 4 allele marker, $p_i = 0.25, k_i = 30, k_{ij} = 9, i, j = 1, \cdots, 4$.

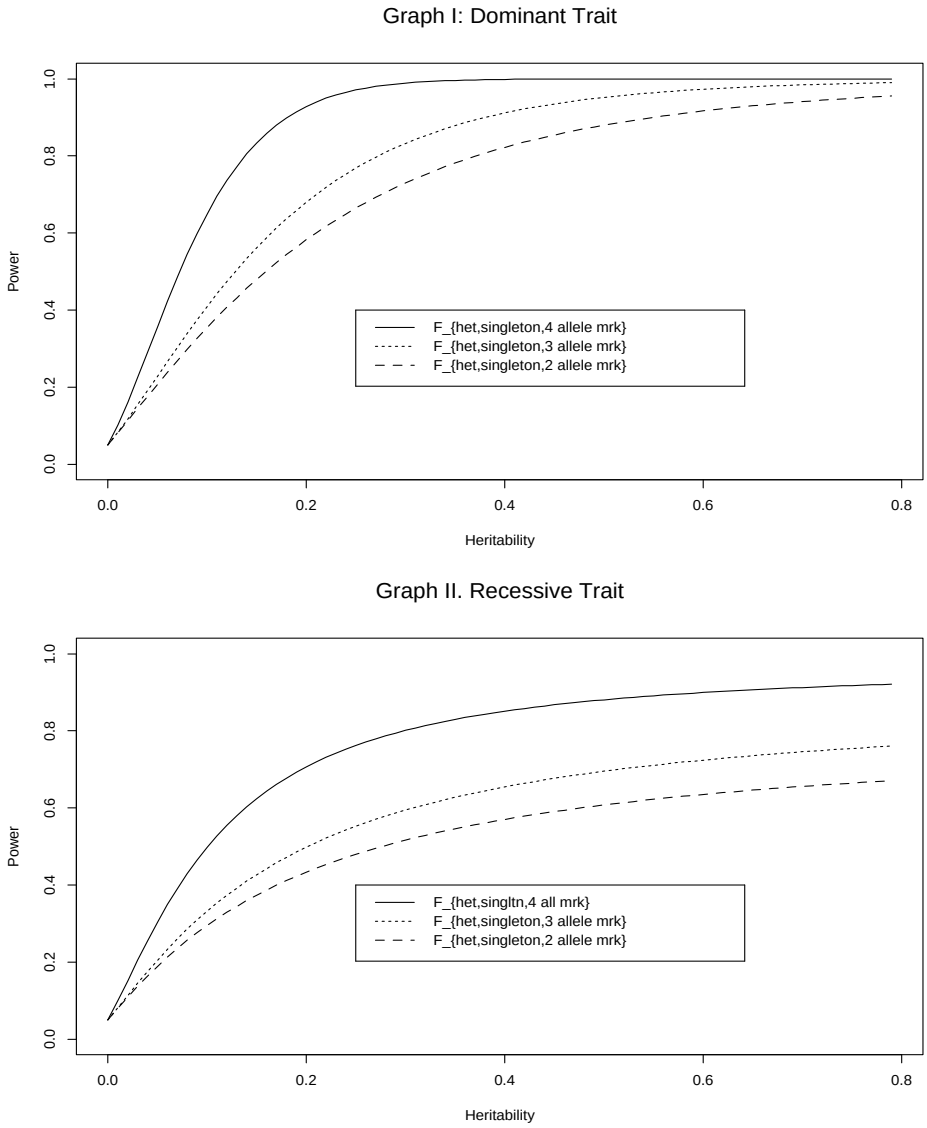

Fig. 3. Power curves of $F_{het,singleton}$ for 2, 3 and 4 allele markers against the heritability at 0.05 significant level, when $q_1 = 0.25, \sigma_G^2 = 0.75, A = 20, \theta = 0.005$ for a dominant trait $a = d = 1.0$, Graph I; and a recessive trait $a = 1.0$ and $d = -0.5$, Graph II. For a 2 allele marker, $p_1 = 0.50, k_1 = k_2 = 100$; For a 3 allele marker, $p_1 = 0.4, p_2 = 0.3, k_1 = 100, k_2 = k_3 = 50$; For a 4 allele marker, $p_i = 0.25, k_i = 50, i = 1, \cdots, 4$.

marker, which in turn is higher than that of the test statistic using 2 allele marker.

Figures 5 and 4 plot the power curves of $F_{Gen\_Nuc,singleton}$ and $F_{Gen\_Nuc,singleton,sibs}$ against the recombination fraction at 0.05 significant level, for dominant and recessive traits for 2, 3 and 4 allele markers, respectively. The four graphs in the two Figures 5 and 4 show that the power of the test statistic using 4 allele marker is higher than that of the test statistic using 3 allele marker, which in turn is higher than that of the test statistic using 2 allele marker. In addition, the power is high when the trait locus is tightly linked to the marker ($\theta < 0.01$); otherwise, the power decreases very rapidly once the trait locus is getting far away from the marker ($\theta > 0.02$).

Assume that the frequencies of marker alleles are not evenly distributed. Figure 6 plots the power curves of $F_{het,singleton,sibs}$ against the heritability at 0.05 significant level, for dominant and recessive traits for 2, 3 and 4 allele markers, respectively. In each of two graphs in the Figure, the power of the test statistic using 3 allele marker is higher than that of the test statistic using 4 allele marker, which in turn is higher than that of the test statistic using 2 allele marker in general.

2.5. Application

The methods and models are applied to analyze the chromosomes 4 and chromosome 16 data of the Oxford asthma data, Genetic Analysis Workshop 12 (Cookson and Abecasis 2001). The data consist of 80 nuclear family with a total of 203 offspring. In these 80 families, 43 have two offspring, 31 have three offspring, and 6 have four offspring. On chromosome 4, 18 markers are typed and each marker has 4 alleles. On chromosome 16, 22 markers are typed and each marker has 4 alleles. In Daniel et al. (1996), linkage to bronchial responsiveness to methacholine (slope) and other quantitative traits were tested by the Haseman-Elston sib-pair technique (Haseman and

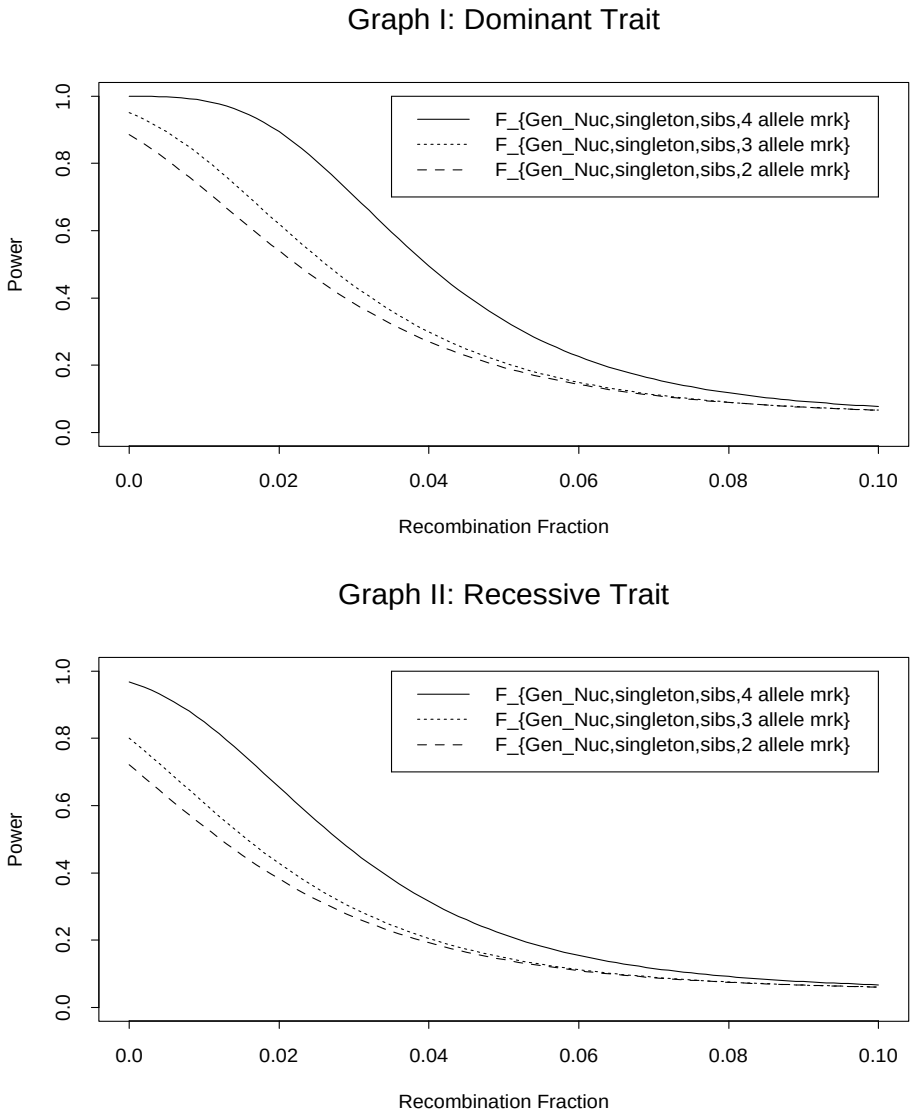

Fig. 4. Power curves of $F_{Gen\_Nuc,singleton,sibs}$ for 2, 3 and 4 allele markers against the recombination fraction at 0.05 significant level, when $q_1 = 0.25, \sigma_G^2 = 0.75, A = 20, h^2 = 0.25$ for a dominant trait $a = d = 1.0$, Graph I; and a recessive trait $a = 1.0$ and $d = -0.5$, Graph II. For a 2 allele marker, $p_1 = 0.50, k_i = 60, k_{ij} = 30, i, j = 1, 2$; For a 3 allele marker, $p_1 = 0.4, p_2 = 0.3, k_1 = 60, k_2 = k_3 = 30, k_{ij} = 15, i, j = 1, 2, 3$; For a 4 allele marker, $p_i = 0.25, k_i = 30, k_{ij} = 9, i, j = 1, \cdots, 4$.

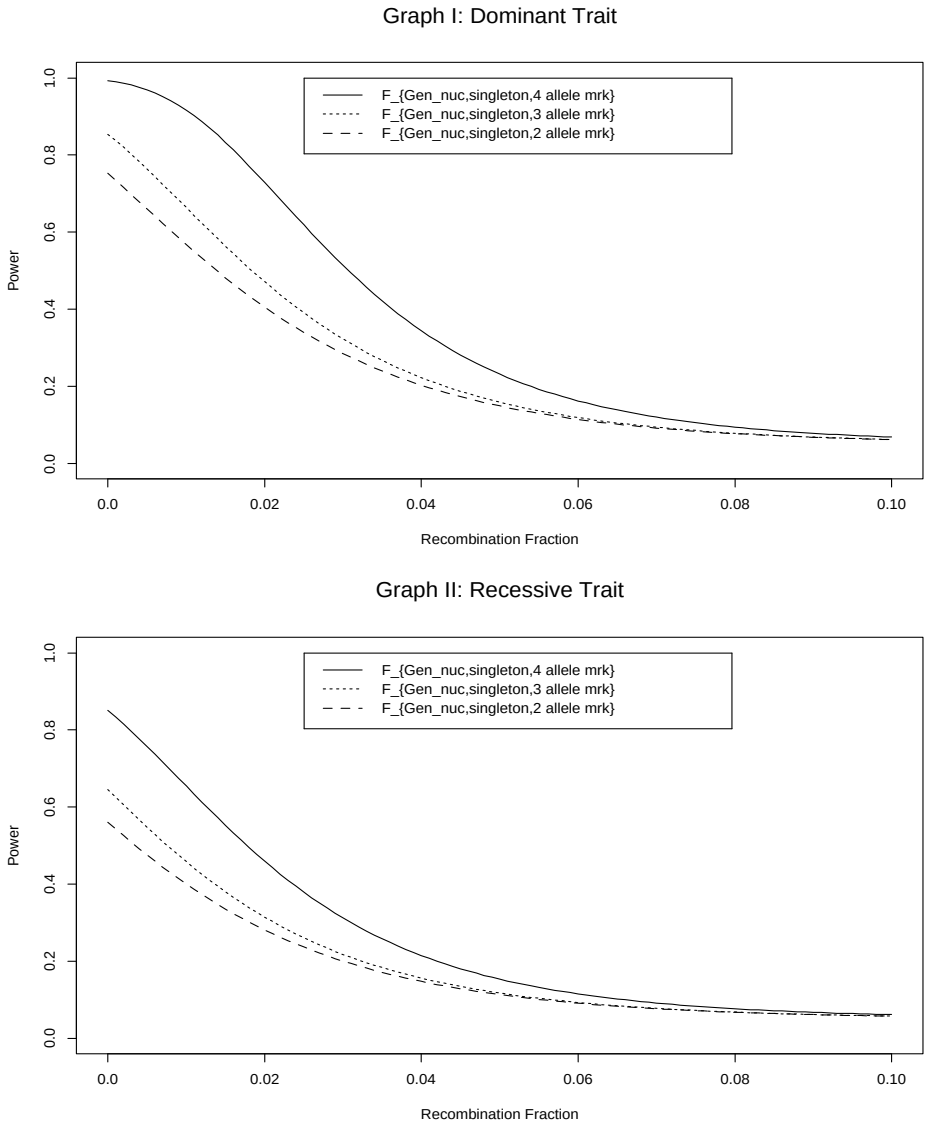

Fig. 5. Power curves of $F_{Gen\_Nuc,singleton}$ for 2, 3 and 4 allele markers against the recombination fraction at 0.05 significant level, when $q_1 = 0.25, \sigma_G^2 = 0.75, A = 20, h^2 = 0.25$ for a dominant trait $a = d = 1.0$, Graph I; and a recessive trait $a = 1.0$ and $d = -0.5$, Graph II. For a 2 allele marker, $p_1 = 0.50, k_1 = k_2 = 100$; For a 3 allele marker, $p_1 = 0.4, p_2 = 0.3, k_1 = 100, k_2 = k_3 = 50$; For a 4 allele marker, $p_i = 0.25, k_i = 50, i = 1, \cdots, 4$.

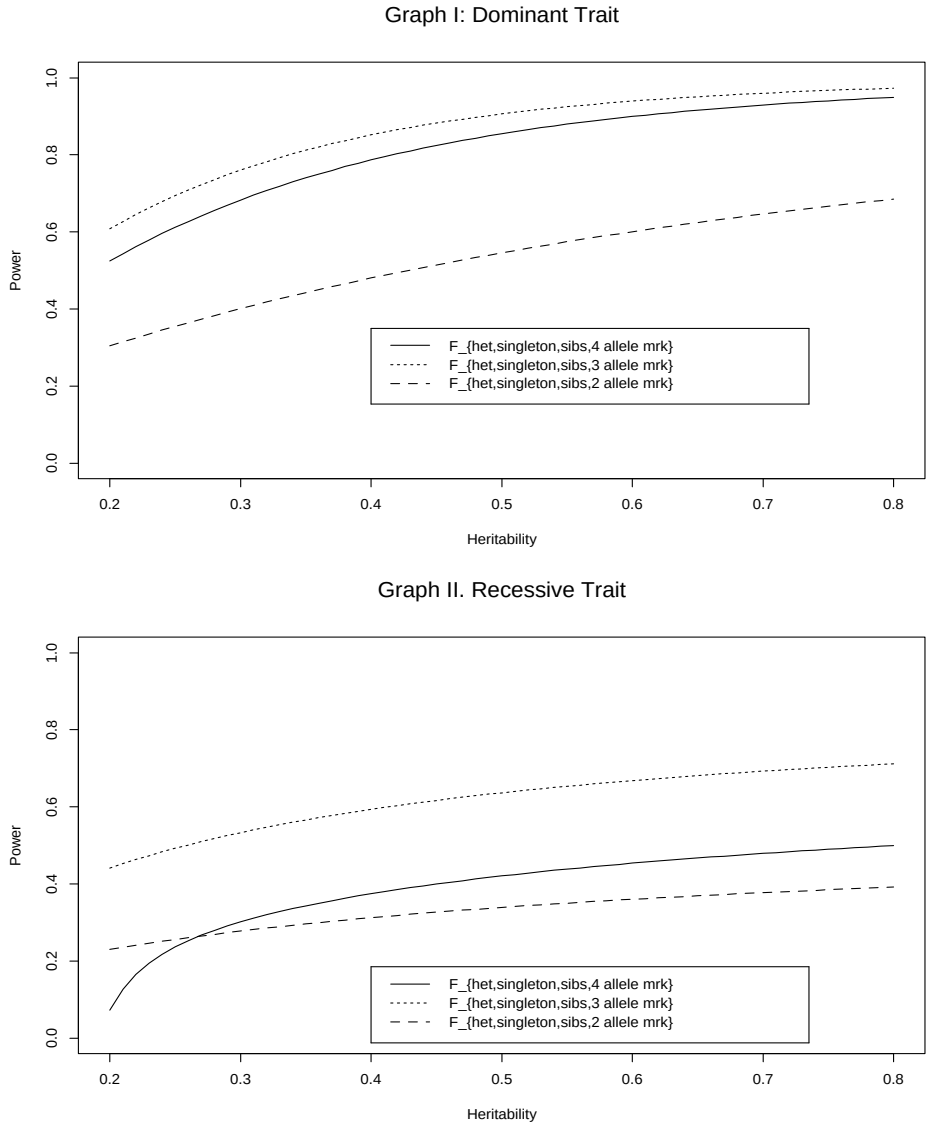

Fig. 6. Power curves of $F_{het,singleton,sibs}$ for 2, 3 and 4 allele markers against the heritability at 0.05 significant level. For a 2 allele marker, $p_1 = 0.90, p_2 = 0.10$; For a 3 allele marker, $p_1 = 0.5, p_2 = 0.45, p_3 = 0.05$; For a 4 allele marker, $p_1 = 0.45, p_2 = p_3 = 0.25, p_4 = 0.05$. All other parameters are the same as those in Figure 2.

Elston 1972). Two regions of potential linkage to autosomal markers were detected with $\log_e$(slope) on chromosomes 4, and 16 (Daniel et al. 1996).

In the four alleles typed, the frequency of one allele is too low (around 3%). When we use the four alleles in data analysis, the convergence is problematic and the results are not stable. This may be due to large number of parameters for the data set. To reduce the number of parameters and to make the results stable, we collapse each of the 4 allele markers to be 3 allele marker. Table I shows the results of test statistics F_{het} and F_{Gen_Nuc}, the results from Fan and Xiong (2002), and Daniel et al. (1996). Three markers, D4S1450, D16S515 and D16S289 show association with the asthma phenotypic trait $\log_e$slope at significant levels 0.05. The results confirms the findings in Fan and Xiong (2002) and Daniel et al. (1996).

Table I. Results of test statistics of asthma data.

Marker Locus	**P-Values of F_{Het}**	**P-Values of F_{Gen_Nuc}**	**P-Values of Fan and Xiong (2002)**	**P-Values of Daniel et al. (1996)**
D4S1540	0.03	0.003	0.02	< 0.05
D16S515	< 0.0001	< 0.0001	< 0.04	< 0.05
D16S289	0.001	< 0.0001	< 0.0001	< 0.05

2.6. Discussion

Mixed models are explored to study association between a multiple allele and a QTL. There are two types of nuclear families in terms of the information of transmission of parental alleles. One is the data of offspring with manifest transmitted alleles from

at least one heterozygous parent. The association study is based on the difference between the conditional mean of trait value given an allele is transmitted and that of trait value given the allele is not transmitted from a heterozygous parent. The other is the data of offspring from nuclear family including homozygous parents. In this case, general association study is based on the difference between the conditional mean of trait value given an allele is transmitted from a parent and the population mean. Using these theoretical bases, mixed models and their test statistics are derived to demonstrate advantage of the method proposed. By power calculation and comparison, the proposed test statistics with a multiple alleles marker have higher power than that with new collapsed bi-alleles marker if the marker allele frequencies are evenly distributed. Therefore, it is more advantageous to use a multiple allele marker for association study in the presence of linkage. It is shown that the power is high when the trait locus is tightly linked to the marker ($\theta < 0$); otherwise, the power decreases very rapidly once the trait locus is getting far away from the marker ($\theta > 0.02$). The proposed models are used to analyze chromosomes 4 and 16 data of the Oxford asthma data, Genetic Analysis Workshop 12.

Fan and Xiong (2003) conducted both linkage analysis in the presence of association and the association study in the presence of linkage. However, it is not clear how to conduct linkage analysis in the presence of association since the way to reduce the number of parameters is not clear for a multiple-allelic marker. In this chapter II, we assume that data are available for all members in a nuclear family. It may not be possible for late onset genetic diseases to obtain the parental data. It would be interesting if the methods and models in this chapter can be extended to apply for sibship data.

CHAPTER III

LINKAGE AND ASSOCIATION STUDY BASED ON SIBSHIP DATA*

3.1. Introduction

Linkage and linkage disequilibrium mappings, two major approaches for genetic studies of human diseases, have been developing in the recent years. There have been lots of interests in joint analyses of both mappings. Separate analysis of either LD mapping or linkage analysis utilizes only part of the available information; LD mapping uses information of LD, on the other hand, linkage analysis uses information of linkage. A combined analysis utilizes both LD and linkage information, and has more power to find putative QTL. For qualitative traits, several studies have shown that combination of LD and linkage mapping is advantageous over separate approach (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 and LD analysis. Fulker et al. (1999) present variance component models to perform integrated linkage and LD mapping based on sibpairs data. Sham et al. (2000) carried out theoretical analyses for power of linkage versus association mapping of quantitative traits based on model in Fulker et al. (1999). Abecasis et al. (2000,2001) generalized the method of Fulker et al. (1999) to analyze data of nuclear families and general pedigrees. For natural populations, Wu et al. (2002) utilized mixture models in joint linkage and LD mapping of QTL. In these studies for the combined analysis, the investigators usually use only one marker in their analyses.

---

*Reprinted with permission from "High Resolution Joint Linkage Disequilibrium and Linkage Mapping of Quantitative Trait Loci Based on Sibship Data" by Ruzong Fan, Jeesun Jung, 2003. *Human Heredity*, Vol. 56, 166–187. by S. Karger AG Basel.

As the dense marker maps such as single nucleotide polymorphisms(SNPs) and high resolution micro-satellite markers are available (The International SNP Map Work Group, 2001; Broman et al. 1998; Kong et al. 2002), it is natural to generalize single marker to multiple markers mapping. Using two flanking markers, Fan and Xiong (2002) proposed a linear regression model to conduct high resolution LD mapping based on population data. The linear regression model incorporated genetic effect decomposed into additive and dominant effects. Fan and Xiong (2003) presented a variance component model which combined linkage and LD mapping. The models employing two flanking markers consider a linear model and variance covariance structure simultaneously to accommodate both population and nuclear family data.

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 recruit parental data. One way to study late-onset disorders is to perform sib-pair 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 resolution joint LD and linkage mapping of QTL based on sibship data. Here, population data are included by treating an independent individual as a single sibship.

In variance component model, a linear regression model and variance covariance structures are introduced to describe a quantitative trait. Association test is based on differences in mean coefficients of linear model. Linkage test is based on differences in covariances according to the identical-by-decent (IBD) status between sib pairs at a candidate locus. Hence, we simultaneously perform joint LD and linkage interval mapping using two flanking markers. Until now, the interval mapping studies published to date are mainly limited to use only the additive genetic variance. There is no explicit formulas to include both additive and dominant genetic variances in the

interval mappings. In this chapter, we derive formulas to calculate covariance of traits between sibships including both additive and dominant variances. To investigate the performance of the formulas, we calculate the numerical values via the formulas and get satisfactory approximations. The non-centrality parameters of test statistics are calculated to compare the power and sample size for cases of sibpairs and general sibships. The non-centrality parameters for linkage analysis are derived based on standard statistical theory, those for LD analysis are calculated by general theory of linear model. Comparison of the power and sample size of LD mapping, and the power of linkage mapping with or without dominant variance is performed. By simulation and theoretical analysis, we compare the results with those of an association between family and association within family ("AbAw") approach from Fulker et al. (1999). The method is applied to Genetic Analysis Workshop (GAW) 12 German asthma data (Meyers, Wjst and Aber, 2001).

3.2. Methods

3.2.1. Linear Model

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 $\mathbf{y} = (y_1, \cdots, y_l)^\tau$,

genotypes at marker A by a vector $(A_1, A_2, \cdots, A_l)^\tau$, and genotypes at marker B by a vector $(B_1, B_2, \cdots, B_l)^\tau$. 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

$$L = -\frac{l}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{y} - X\mu)^\tau \Sigma^{-1}(\mathbf{y} - X\mu). \tag{3.1}$$

The notations of model (3.1) are defined as follows. Σ is a $l \times l$ variance-covariance matrix defined as $\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1l} \\ \rho_{21} & 1 & \cdots & \rho_{2l} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{l1} & \rho_{l2} & \cdots & 1 \end{pmatrix} \sigma^2$, 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 sharing 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 shared by the i-th child and the j-th child at the putative QTL Q are IBD (Pratt et al. 2000; Zhu and Elston 2000). To introduce the mean component $X\mu$ for log-likelihood (3.1), we consider the following regression (Fan and Xiong 2002, 2003)

$$y_i = \beta + w_i\gamma + x_{Ai}\alpha_A + x_{Bi}\alpha_B + z_{Ai}\delta_A + z_{Bi}\delta_B + G_i + H_i + e_i, \tag{3.2}$$

where β is the overall mean, w_i is a row vector of covariates such as gender 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

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

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

$\alpha_A, \alpha_B, \delta_A$ and δ_B are the coefficients of the dummy variables x_{Ai}, x_{Bi}, z_{Ai} and z_{Bi}. X is the design matrix based on regression (3.2), 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 (3.2) as follows. Let μ_{ij} be the effect of genotype $Q_i Q_j, i, j = 1, 2$, $\mu_{12} = \mu_{21}$. Denote the overall population mean by $\mu_0 = \mu_{11} q_1^2 + 2\mu_{12} q_1 q_2 + \mu_{22} q_2^2$, 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 (3.2) 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 $L_1, \cdots, L_I$, 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 $\mathbf{y} = (\mathbf{y}_1^\tau, \cdots, \mathbf{y}_I^\tau)^\tau$, the total variance-covariance matrix by $\Sigma = diag(\Sigma_1, \cdots, \Sigma_I)$, and model matrix $X = (X_1^\tau, \cdots, X_I^\tau)^\tau$. 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 $\mu = (\beta, \gamma, \alpha_A, \alpha_B, \delta_A, \delta_B)^\tau, \sigma^2_{ga}, \sigma^2_{gd}, \sigma^2_{Ga}, \sigma^2_{Gd}, \sigma^2_s$, and σ^2_e. 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), average effect of gene substitution of QTL is $\alpha_Q = a + (q_2 - q_1)d$ and dominant deviation $\delta_Q = 2d$. The additive variance $\sigma^2_{ga} = 2q_1q_2\alpha^2_Q$ and the dominant variance $\sigma^2_{gd} = (q_1q_2)^2\delta^2_Q$. To test the linkage of the trait locus to a particular position in the genome, the null hypothesis is $H_0 : \sigma^2_{ga} = \sigma^2_{gd} = 0$ and the alternative hypothesis is $H_A : \sigma^2_{ga} > 0$ or $\sigma^2_{gd} > 0$. The corresponding LRT is a mixture of χ^2 variables (Self and Liang 1987). If only the additive variance σ^2_{ga} (or dominant variance σ^2_{gd}) is modeled, the null hypothesis is $H_0 : \sigma^2_{ga} = 0$ (or $H_0 : \sigma^2_{gd} = 0$), and the alternative hypothesis is $H_A : \sigma^2_{ga} > 0$ (or $\sigma^2_{gd} > 0$). Then the corresponding LRT is a $\frac{1}{2} : \frac{1}{2}$ mixture of χ^2_1 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

$$V_A = \begin{pmatrix} 2P_aP_A & 2D_{AB} \\ 2D_{AB} & 2P_bP_B \end{pmatrix}, \text{ and } V_D = \begin{pmatrix} P_a^2P_A^2 & D_{AB}^2 \\ D_{AB}^2 & P_b^2P_B^2 \end{pmatrix}. \tag{3.3}$$

Like Fan and Xiong (2002), we can show that the coefficients of regression equation (3.2) are

$$\begin{pmatrix} \alpha_A \\ \alpha_B \end{pmatrix} = V_A^{-1} \begin{pmatrix} 2D_{AQ} \\ 2D_{QB} \end{pmatrix} \alpha_Q, \begin{pmatrix} \delta_A \\ \delta_B \end{pmatrix} = V_D^{-1} \begin{pmatrix} D_{AQ}^2 \\ D_{QB}^2 \end{pmatrix} \delta_Q. \tag{3.4}$$

Equations (3.4) imply that regression (3.2) 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 incorporated 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 (3.4). **First**, suppose that the presence of linkage is verified 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. The existence of LD between markers and trait locus Q cab be tested by $H_0 : \alpha_A = \alpha_B = \delta_A = \delta_B = 0$ vs H_A : at least one of $\alpha_A, \alpha_B, \delta_A$, and δ_B is not 0. The test 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 verified 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 (3.4) can be replaced by $\alpha_A = D_{AQ}\alpha_Q/(P_a P_A), \delta_A = D_{AQ}^2\delta_Q/(P_a^2 P_A^2)$. Suppose that the

presence of linkage is supported by both $\sigma^2_{ga} > 0$ and $\sigma^2_{gd} > 0$. 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 $\sigma^2_{ga} > 0$, but the dominant variance σ^2_{gd} 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 $\sigma^2_{gd} > 0$, but the additive variance σ^2_{ga} 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 (3.2) 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 (3.2) for high resolution joint LD and linkage mapping. In this second step, the significant variables among $\sigma^2_{ga}, \sigma^2_{gd}, \alpha_A, \alpha_B, \delta_A$ 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.

3.2.2. 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 $\mathrm{Cov}(y_1, y_2|I_A, I_B) = \hat{\pi}_{ijQ}\sigma^2_{ga} + \hat{\Delta}_{ijQ}\sigma^2_{gd} + \sigma^2_{Ga}/2 + \sigma^2_{Gd}/4 + \sigma^2_s = \hat{\rho}_{ij}\sigma^2$, where $\hat{\pi}_{ijQ} = \mathrm{E}(\pi_{ijQ}|I_A, I_B)$, π_{ijQ} is the proportion of allele IBD at putative QTL Q, $\hat{\Delta}_{ijQ} = \mathrm{E}(\Delta_{ijQ}|I_A, I_B)$ 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 sib-pair 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 sharing at marker A and marker B by sib-pair 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

$$\begin{aligned} \beta_{\pi A} &= \frac{(1-2\theta_{AQ})^2 - (1-2\theta_{AB})^2(1-2\theta_{QB})^2}{1-(1-2\theta_{AB})^4} \\ \beta_{\pi B} &= \frac{(1-2\theta_{QB})^2 - (1-2\theta_{AB})^2(1-2\theta_{AQ})^2}{1-(1-2\theta_{AB})^4} \\ \alpha_\pi &= \frac{1-\beta_{\pi A}-\beta_{\pi B}}{2}. \end{aligned} \tag{3.5}$$

Let $\Delta_{ijA}, \Delta_{ijB}$ be the probability of sharing 2 alleles IBD at markers A and B for the

sib-pair i and j, respectively. We propose to estimate Δ_{ijQ} by

$$\hat{\Delta}_{ijQ} = \alpha + \beta_A \pi_{ijA} + \beta_B \pi_{ijB} + r_A \Delta_{ijA} + r_B \Delta_{ijB}. \tag{3.6}$$

In Appendices H, I and J, we show that under the assumption of no interference,

$$\begin{aligned} r_A &= \frac{(1-2\theta_{AQ})^4 - (1-2\theta_{QB})^4(1-2\theta_{AB})^4}{1-(1-2\theta_{AB})^8} \\ r_B &= \frac{(1-2\theta_{QB})^4 - (1-2\theta_{AQ})^4(1-2\theta_{AB})^4}{1-(1-2\theta_{AB})^8} \\ \beta_A &= \beta_{\pi A} - r_A, \beta_B = \beta_{\pi B} - r_B \\ \alpha &= \frac{(1-\psi_A)^2(1-\psi_B)^2}{[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]^2}, \end{aligned} \tag{3.7}$$

where $\beta_{\pi A}, \beta_{\pi B}$ are given in equations (3.5) (Fulker and Cardon 1994), $\psi_A = \theta_{AQ}^2 + (1-\theta_{AQ})^2$ and $\psi_B = \theta_{QB}^2 + (1-\theta_{QB})^2$. When we assume that the positions of marker A and marker B are known, θ_{AB} can be calculated through a Haldane's function $\theta = [1-\exp(-2\lambda)]/2$ under assumption of no interference, where λ is map distance.

3.3. Test Statistics and Non-Centrality Parameter

3.3.1. Association Study

We assume that the data are composed of three sub-samples: n independent individuals, m independent sib-pairs, 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 sib-pairs 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 $\mu = (\beta, \alpha_A, \alpha_B, \delta_A, \delta_B)^\tau$. 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 $\hat{\mu} = \left[X^\tau \hat{\Sigma}^{-1} X\right]^{-1} X^\tau \hat{\Sigma}^{-1} \vec{y} = \left[\sum_{i=1}^{I} X_i^\tau \hat{\Sigma}_i^{-1} X_i\right]^{-1} \sum_{i=1}^{I} X_i^\tau \hat{\Sigma}_i^{-1} \vec{y}_i$. 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

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

with the non-centrality parameter $\lambda = (H\mu)^\tau \left[H[X^\tau \Sigma^{-1} X]^{-1} H^\tau\right]^{-1} H\mu$. Under the assumption of large sample sizes n, m and k, we show in Appendix L that

$$\sum_{i=1}^{n+m+k} X_i^\tau \Sigma_i^{-1} X_i \approx diag(a_1, a_2 V_A, a_3 V_D)/\sigma^2, \tag{3.8}$$

where a_1, a_2 and a_3 are constants given by equations (L.4) in Appendix L.

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 (3.4) and (3.8) that

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

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., $\sigma^2_{gd} > 0$ 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 (3.4) and (3.8) that

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

In the presence of both additive and dominant effects, i.e., $\sigma^2_{ga} > 0$ and $\sigma^2_{gd} > 0$, 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 $\lambda_{A,ad} \approx [1/\sigma^2] \Big[a_2 \sigma^2_{ga} R^2_{AQ} + a_3 \sigma^2_{gd} R^4_{AQ} \Big]$, for the null hypothesis $H_{A,ad} : \alpha_A = \delta_A = 0$. Correspondingly, we denote the F-test statistic by $F_{A,ad}$. Similarly, $\lambda_{A,a} \approx [a_2/\sigma^2] \sigma^2_{ga} R^2_{AQ}$ is the non-centrality parameter of the test statistic $F_{A,a}$ for the 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 $\lambda_{A,d} \approx [a_3/\sigma^2] \sigma^2_{gd} R^4_{AQ}$.

3.3.2. 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.

Consider a sib-ship of l children. Under the null hypothesis of no linkage between the trait locus and the markers, the correlation of each sib-pair is $\rho = \frac{\sigma^2_{ga}}{2\sigma^2} + \frac{\sigma^2_{gd}}{4\sigma^2} + \frac{\sigma^2_{Ga}}{2\sigma^2} + \frac{\sigma^2_{Gd}}{4\sigma^2} + \frac{\sigma^2_s}{\sigma^2}$. Hence, we have twice the expected log-likelihood

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

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

$$\begin{aligned}
C_k &= \mathrm{Cov}(y_i, y_j|\pi_{ijA} = k/2)/\sigma^2 = (\sigma^2_{ga} + \sigma^2_{gd})P(\pi_{ijQ} = 1|\pi_{ijA} = k/2)/\sigma^2 \\
&\quad + \frac{\sigma^2_{ga}}{2}P(\pi_{ijQ} = 1/2|\pi_{ijA} = k/2)/\sigma^2 + [\sigma^2_{Ga}/2 + \sigma^2_{Gd}/4 + \sigma^2_s]/\sigma^2, k = 0, 1, 2.
\end{aligned}$$

From Haseman and Elston (1972), Table IV, or Sham et al. (2000), Table 1, we have

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

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

$$\mathrm{E}\left(2L_{random,A}\right) = -l - l\log\left[2\pi\sigma^2\right]$$

$$-\sum_{\pi_{12A}}\cdots\sum_{\pi_{l-1,lA}} P(\pi_{12A})\cdots P(\pi_{l-1,lA})\log\det\begin{pmatrix} 1 & C_{2\pi_{12A}} & \cdots & C_{2\pi_{1lA}} \\ C_{2\pi_{21A}} & 1 & \cdots & C_{2\pi_{2lA}} \\ \vdots & \vdots & \cdots & \vdots \\ C_{2\pi_{l1A}} & C_{2\pi_{l2A}} & \cdots & 1 \end{pmatrix},$$

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

$$\lambda_{linkage,A}=\log\left[1-\rho^2\right]-\sum_{k=0}^{2}P(\pi_{12A}=k/2)\log\left[1-C_k^2\right]. \tag{3.9}$$

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

$$\begin{aligned} C_{k_1k_2} &= \mathrm{Cov}(y_i,y_j|\pi_{ijA}=k_1/2,\pi_{ijB}=k_2/2)/\sigma^2 \\ &= \Big[(\sigma^2_{ga}+\sigma^2_{gd})P(\pi_{ijQ}=1|\pi_{ijA}=k_1/2,\pi_{ijB}=k_2/2) \\ &\quad +\frac{\sigma^2_{ga}}{2}P(\pi_{ijQ}=1/2|\pi_{ijA}=k_1/2,\pi_{ijB}=k_2/2)+\sigma^2_{Ga}/2+\sigma^2_{Gd}/4+\sigma^2_s\Big]/\sigma^2. \end{aligned} \tag{3.10}$$

To calculate the quantities $C_{k_1k_2}$, we need the joint distribution of π_{ijA},π_{ijQ} and π_{ijB} of a sib-pair i and j under the alternative hypothesis of linkage. Based on Table II, we can calculate $C_{ij},i,j=0,1,2$, which are given in Appendix K. We have twice the expected log-likelihood under the alternative hypothesis of linkage

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

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 $\lambda_{linkage,AB} = \mathrm{E}\,(2L_{random,AB}) - \mathrm{E}\,(2L_{Null})$. If the sibship consists of two offspring, then

$$\lambda_{linkage,AB} = \log\left[1-\rho^2\right] - \sum_{i,j=0}^{2} P(\pi_{12A} = i/2)P(\pi_{12B} = j/2)\log\left[1 - C_{ij}^2\right]. \quad (3.11)$$

The correlation quantitative $C_{k_1k_2}$ between the sibpair i and j are derived in Appendix K.

Table II: Joint distribution of π_Q, π_A and π_B of a sib-pair. Here subscripts ij are omitted from π_{ijQ}, π_{ijA} and π_{ijB}. **Prob.** = Probability.

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	$\psi_A^2/4\cdot$ $(1-\psi_B)^2$	$\psi_A(1-\psi_A)/2\cdot$ $\psi_B(1-\psi_B)$	$(1-\psi_A)^2/4\cdot$ ψ_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	$\psi_A(1-\psi_A)/2$ $\cdot 2\psi_B(1-\psi_B)$	$(1-2\psi_A+2\psi_A^2)/2$ $\cdot(1-2\psi_B+2\psi_B^2)$	$\psi_A(1-\psi_A)/2$ $\cdot 2\psi_B(1-\psi_B)$	$\Big[[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]^2 +$ $[(1-\psi_A)\psi_B + \psi_A(1-\psi_B)]^2\Big]/2$
	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	$(1-\psi_A)^2/4\cdot$ $2\psi_B(1-\psi_B)$	$\psi_A(1-\psi_A)/2\cdot$ $(1-2\psi_B+2\psi_B^2)$	$\psi_A^2/4\cdot$ $2\psi_B(1-\psi_B)$	$[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]\cdot$ $[(1-\psi_A)\psi_B + \psi_A(1-\psi_B)]/2$
	0	$(1-\psi_A)^2/4\cdot$ $(1-\psi_B)^2$	$\psi_A(1-\psi_A)/2\cdot$ $\psi_B(1-\psi_B)$	$\psi_A^2/4\cdot$ ψ_B^2	$[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]^2/4$
Total Prob.		1/4	1/2	1/4	1

3.4. Estimates of the Probability of Sharing 2 Alleles IBD for Sibs

Tables III and IV give the interval estimates of $\hat{\Delta}_Q$ by π_A, π_B, Δ_A and Δ_B under Haldane's function. Table III takes a map distance $\lambda_{AB} = 20$ cM, and Table IV 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 sub-intervals. 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 forth location are intermediates between location 3 and location 5. In Table III, 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 IV, the estimates $\hat{\Delta}_Q$ on the third location tends 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 $r_A \approx \frac{(1-4\cdot 2\theta_{AQ})-(1-4\cdot 2\theta_{QB})(1-4\cdot 2\theta_{AB})}{1-(1-8\cdot 2\theta_{AB})} \approx \frac{(1-8\theta_{AQ})-(1-8\theta_{QB}-8\theta_{AB})}{16\theta_{AB}} \approx \frac{-\theta_{AQ}+\theta_{QB}+(\theta_{AQ}+\theta_{QB})}{2\theta_{AB}} = \frac{\theta_{QB}}{\theta_{AB}}$. 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 III are approximately equal to the average

Table III. 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	1.00
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 IV. 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

of Δ_A and Δ_B.

3.5. Power Comparison

3.5.1. Comparisons with the "AbAw" Approach of Fulker

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 1000 sib-pairs in Table V. The results of "AbAw" approach by Fulker and Abecasis et al. are directly taken from Table 5, p1625, Sham et al. (2000). The QTL is assumed to be additive with $\sigma^2_{ga} = 0.2$. The shared residual environment variance, σ^2_s, is set to be either 0 or 0.4, such as those in Tables 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 bi-allelic with equal allele frequencies. The measure D_{AQ} of LD varies from complete disequilibrium, 0.25, to weak disequilibrium, 0.025. In Table V, the statistic $F_{A,a}$ is approximately distributed as non-central $\chi^2(1)$, since the sample size of 1000 sib-pairs 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 sib-pairs are generated by 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 V, 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).

Table V: Empirical values vs. theoretical expectations of statistics, compared with results of Table 5, Sham et al. (2000), when $\sigma^2_{ga} = 0.2, \sigma^2_{gd} = \sigma^2_{Ga} = \sigma^2_{Gd} = 0$. A sample of 50,000 sib-pairs 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. **Abbreviations.** 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 $H_N : \alpha_A = 0$. $F = \frac{(H\hat{\mu})^\tau [H(X^\tau \hat{\Sigma}^{-1} X)^{-1} H^\tau]^{-1} (H\hat{\mu})(N-2)}{Y^\tau [\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} X (X^\tau \hat{\Sigma}^{-1} X)^{-1} X^\tau \hat{\Sigma}^{-1}] Y}$, $\mu = (\beta, \alpha_A)^\tau$, and $H = (0, 1)$. (*), 36.52 in Sham et al. (2000), Table 5, should be 33.52.

D_{AQ}	$\sigma^2_s = 0.0, \sigma^2_e = 0.8$												
	Sham et.al (2000), theory		**Current Method**										
			theory		$\theta_{AQ} = 0$			$\theta_{AQ} = 0.02$			$\theta_{AQ} = 0.04$		
	BP	**WP**	$\lambda_{A,a} + 1$	α_A	LRT	$F_{A,a}$	$\hat{\alpha}_A$	LRT	$F_{A,a}$	$\hat{\alpha}_A$	LRT	$F_{A,a}$	$\hat{\alpha}_A$
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

D_{AQ}	$\sigma^2_s = 0.4, \sigma^2_e = 0.4$												
	Sham et.al (2000), theory		**Current Method**										
			theory		$\theta_{AQ} = 0$			$\theta_{AQ} = 0.02$			$\theta_{AQ} = 0.04$		
	BP	**WP**	$\lambda_{A,a} + 1$	α_A	LRT	$F_{A,a}$	$\hat{\alpha}_A$	LRT	$F_{A,a}$	$\hat{\alpha}_A$	LRT	$F_{A,a}$	$\hat{\alpha}_A$
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

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.

3.5.2. Comparisons of Sample Sizes and Power for LD Mapping

In the sample size and power calculations, we take an additive polygenic variance $\sigma^2_{Ga} = 0.10$, polygenic dominant variance $\sigma^2_{Gd} = 0.05$, and shared environment residual variance $\sigma^2_s = 0$. For sib-pairs, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$. For tri-sibships, $\pi_A = \pi_B = \Delta_A = \Delta_B = 0.5$ for sib-pair 1 and 2; $\pi_A = \pi_B = \Delta_B = 0.5, \Delta_A = 0.25$ for sib-pair 1 and 3; and $\pi_A = \pi_B = 0.5, \Delta_A = \Delta_B = 0.25$ for sib-pair 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 $h^2 = \sigma^2_{ga}/\sigma^2$. 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 7 gives the required number of sib-pairs (Graphs I and II) and tri-sibships (Graphs III and IV) 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, for a mode of dominant inheritance $a = d = 1.0$ (Graphs I and II), and a mode of recessive inheritance $a = 1.0, d = -0.5$ (Graphs III and IV), respectively. In the figure, we take equal allele frequencies $q_1 = P_A = P_B = 0.50$, LD coefficients $D_{AB} = 0.10, D_{AQ} = D_{QB} = 0.15$, and map distances $\lambda_{AB} = 5cM, \lambda_{AQ} = \lambda_{QB} = 2.5cM$. We can see the following: (1) For both dominant and recessive traits, the required number of sib-pairs 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 (Graphs I and III); (2) For dominant traits, the required number of sib-pairs 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 large than 0.1 (Graph I); the required number of sib-pairs of test statistic $F_{AB,ad}$ is similar to that of $F_{AB,a}$, and the required number of sib-pairs 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 (Graph III); (4) The required number of sib-pairs or tri-sibships of test statistics $F_{AB,d}$ and $F_{A,d}$ is much bigger, especially for recessive trait (Graphs II and IV).

Figure 8 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 $P_A = 0.5$ (Graphs I and II), $q_1 = 0.5$ (Graphs III and IV), $P_B = 0.50, n = 60, m = 30, k = 20$, $\lambda_{AB} = 5cM, \lambda_{AQ} = \lambda_{QB} = 2.5cM$, 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 7. Interestingly, the power of statistics $F_{AB,ad}$ and $F_{AB,a}$ depends heavily on the trait allele frequency q_1 (Graphs I and II), but not so much on the marker allele frequency P_A (Graphs III and IV). 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 9 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 $q_1 = P_A = P_B =$

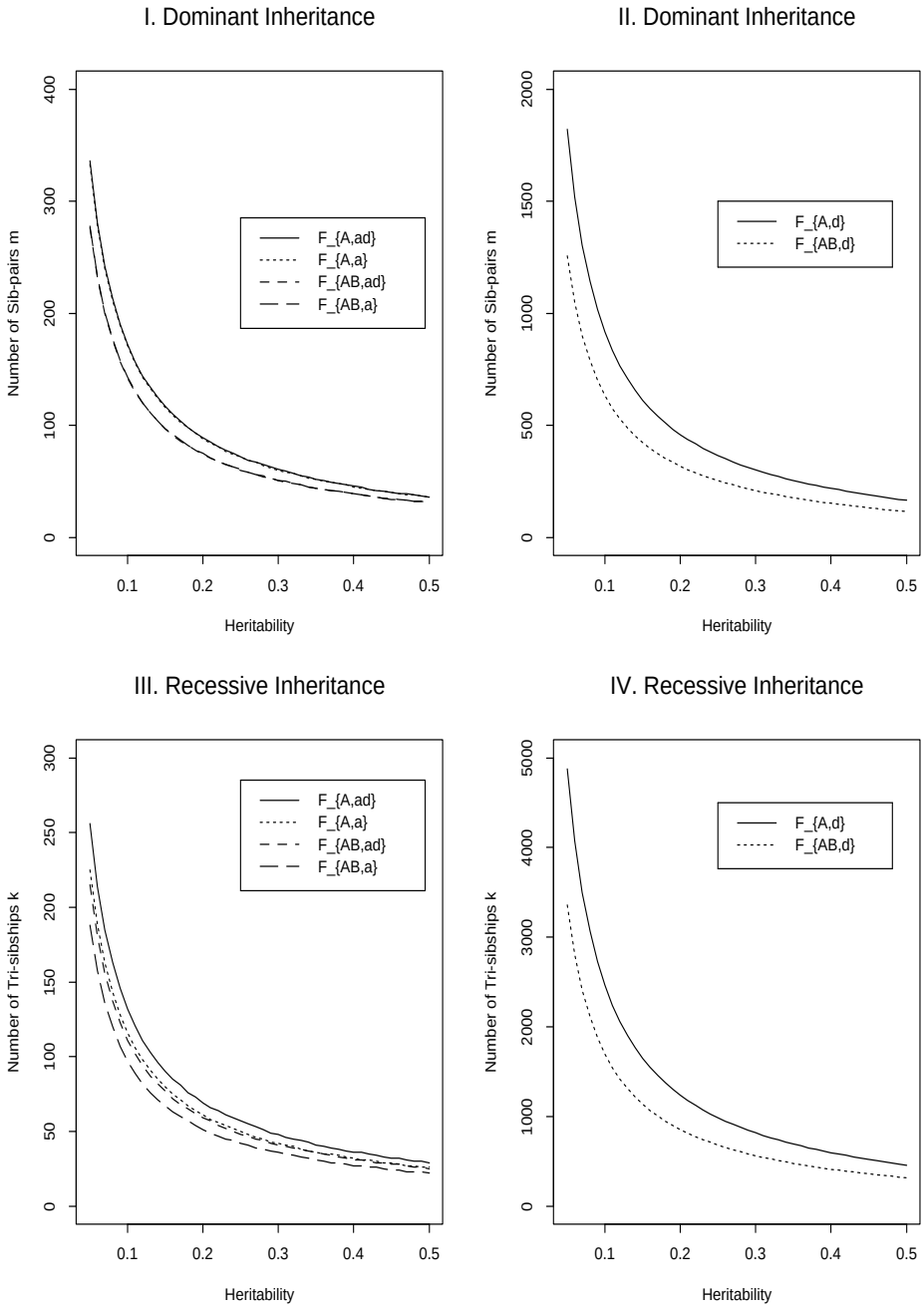

Fig. 7. Number of sib-pairs (Graphs I and II) or tri-sibships (Graphs III and IV) of test statistics $F_{AB,ad}$, $F_{AB,a}$,$F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against the heritability h^2 at 0.01 significant level and 0.80 power.

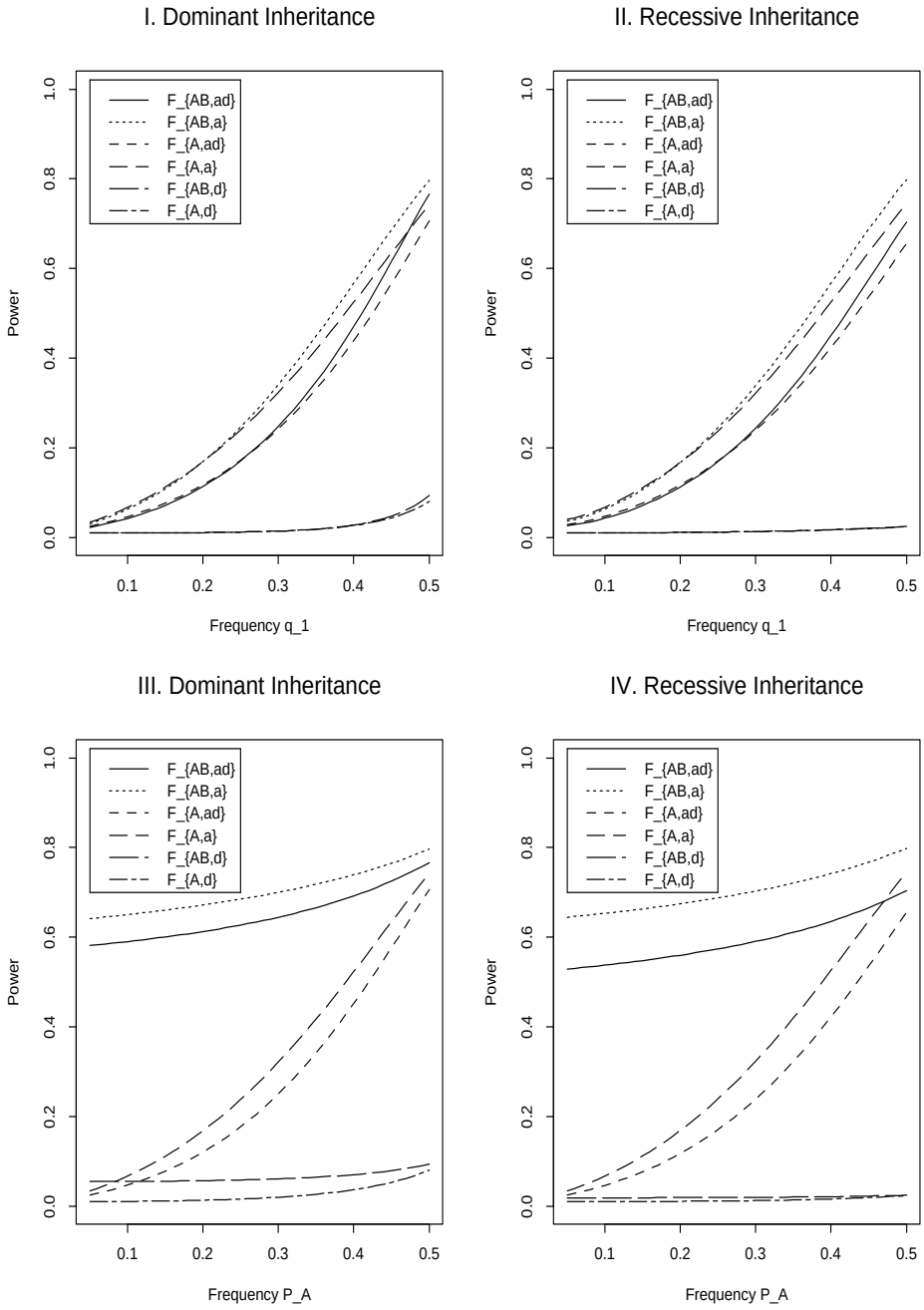

Fig. 8. Power of test statistics $F_{AB,ad}$, $F_{AB,a}$,$F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against trait frequency q_1 or marker allele frequency P_A at 0.01 significant level, when $P_A = 0.5$ (Graphs I and II), $q_1 = 0.5$ (Graphs III and IV).

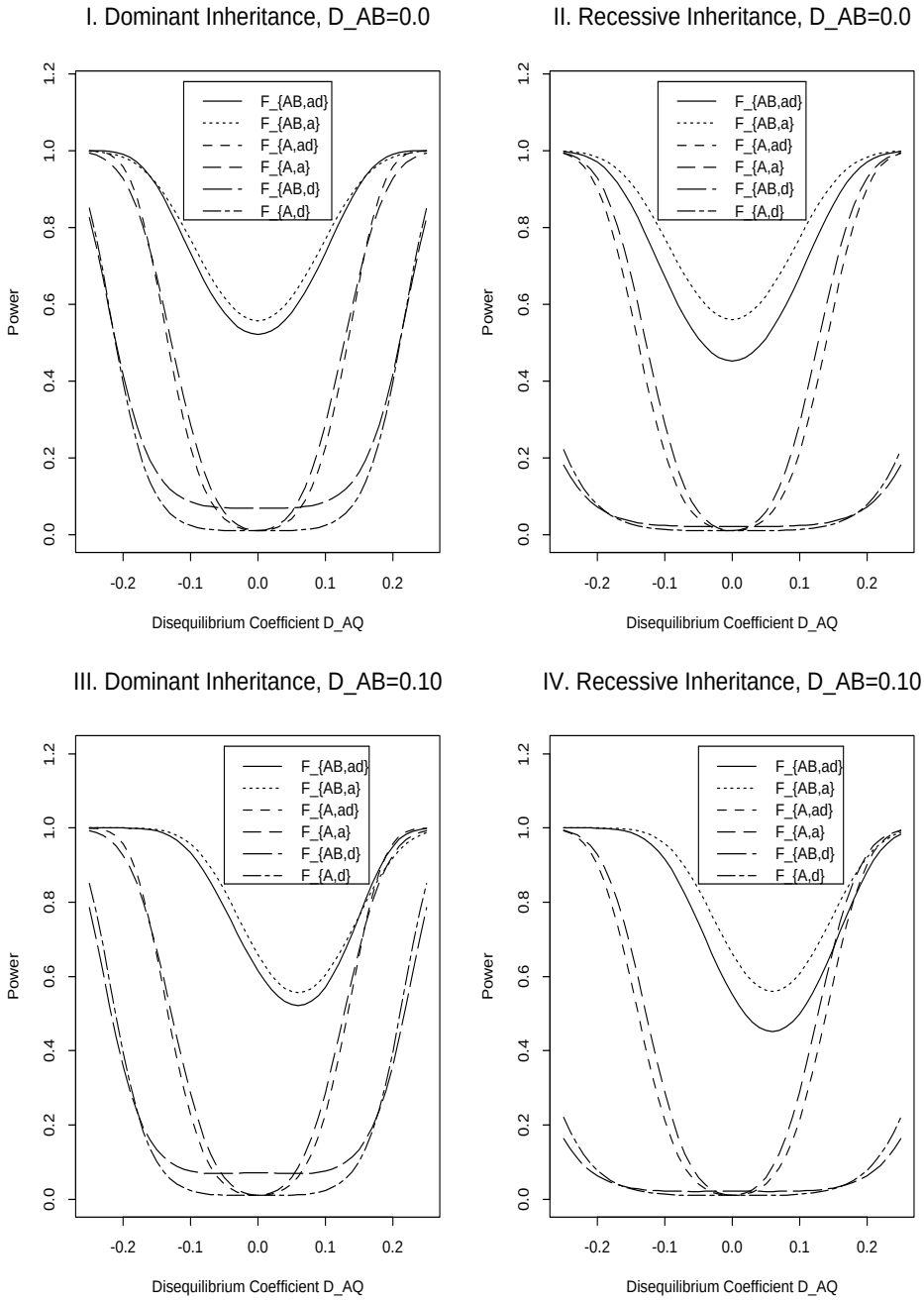

Fig. 9. Power of test statistics $F_{AB,ad}$, $F_{AB,a}$,$F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against LD coefficient D_{AQ} at 0.01 significant level.

$0.50, D_{QB} = 0.15, n = 60, m = 30, k = 20$, $\lambda_{AB} = 5cM, \lambda_{AQ} = \lambda_{QB} = 2.5cM$, 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 $D_{AQ} = 0$ (Graphs I and II). If LD measure D_{AB} is highly positive (Graphs III and IV, $D_{AB} = 0.10$), 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 ($|D_{AQ}| > 0.20$, Graphs I and III). For recessive traits, the power of $F_{AB,d}$ and $F_{A,d}$ is very low (Graphs II and IV).

Figure 10 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 $q_1 = P_A = P_B = 0.50, n = 60, m = 30, k = 20$, $\lambda_{AB} = 5cM, \lambda_{AQ} = \lambda_{QB} = 2.5cM$, 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 (Graphs I and II, $D_{AB} = 0.10, D_{AQ} = D_{QB} = 0.15$), 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 (Graphs III and IV, $D_{AB} = 0.05, D_{AQ} = D_{QB} = 0.08$), 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 $D_{AQ}(T) = D_{AQ}(0)(1 - \theta_{AQ})^T$. Sim-

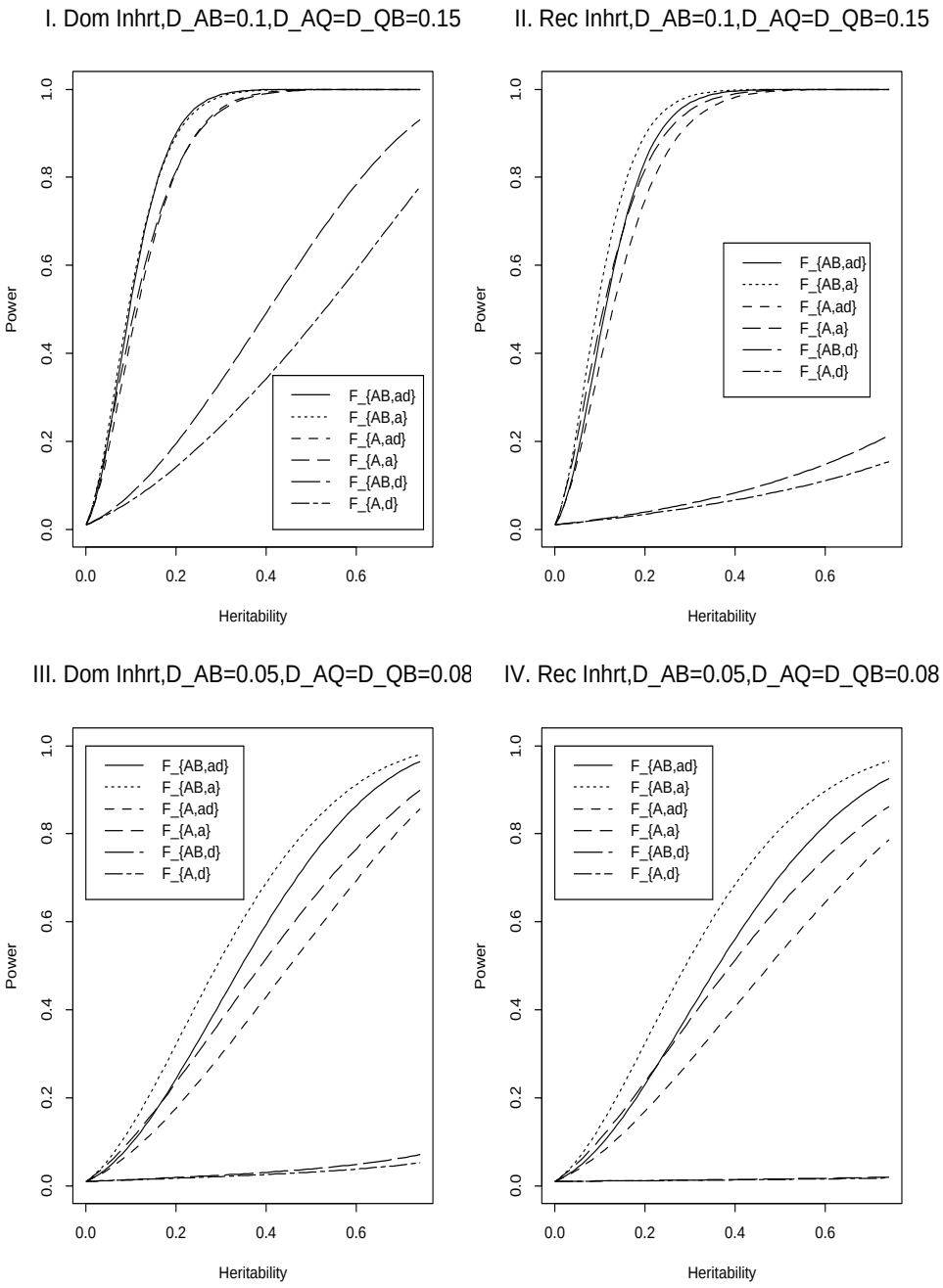

Fig. 10. Power of test statistics $F_{AB,ad}$, $F_{AB,a}$,$F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against heritability h^2 at 0.01 significant level.

ilarly, the other LD coefficients are $D_{AB}(T) = D_{AB}(0)(1-\theta_{AB})^T$ and $D_{QB}(T) = D_{QB}(0)(1-\theta_{QB})^T$. In Figure 11, Graphs I and II 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 $q_1 = P_A = P_B = 0.50, n = 60, m = 30, k = 20$, $\lambda_{AB} = 4.5cM$, 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 0cM, and marker B is located at 4.5cM. 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 $1.5cM$ from marker A). When trait locus Q locates in the region which is larger than $1.5cM$ 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. Graphs III and IV of Figure 11 show the power of test statistics $F_{AB,ad}$ for different mutation ages against the position of markers A and B at 0.01 significant level. In the two graphs, the trait locus Q locates at 10cM; 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 2.5cM away from the QTL.

3.5.3. Comparisons of Sample Sizes and Power 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 $m = 250$ sib pairs. Multiplying $\lambda_{linkage,AB}$ of (3.11) given in Appendix D by m, we calculate the non-centrality

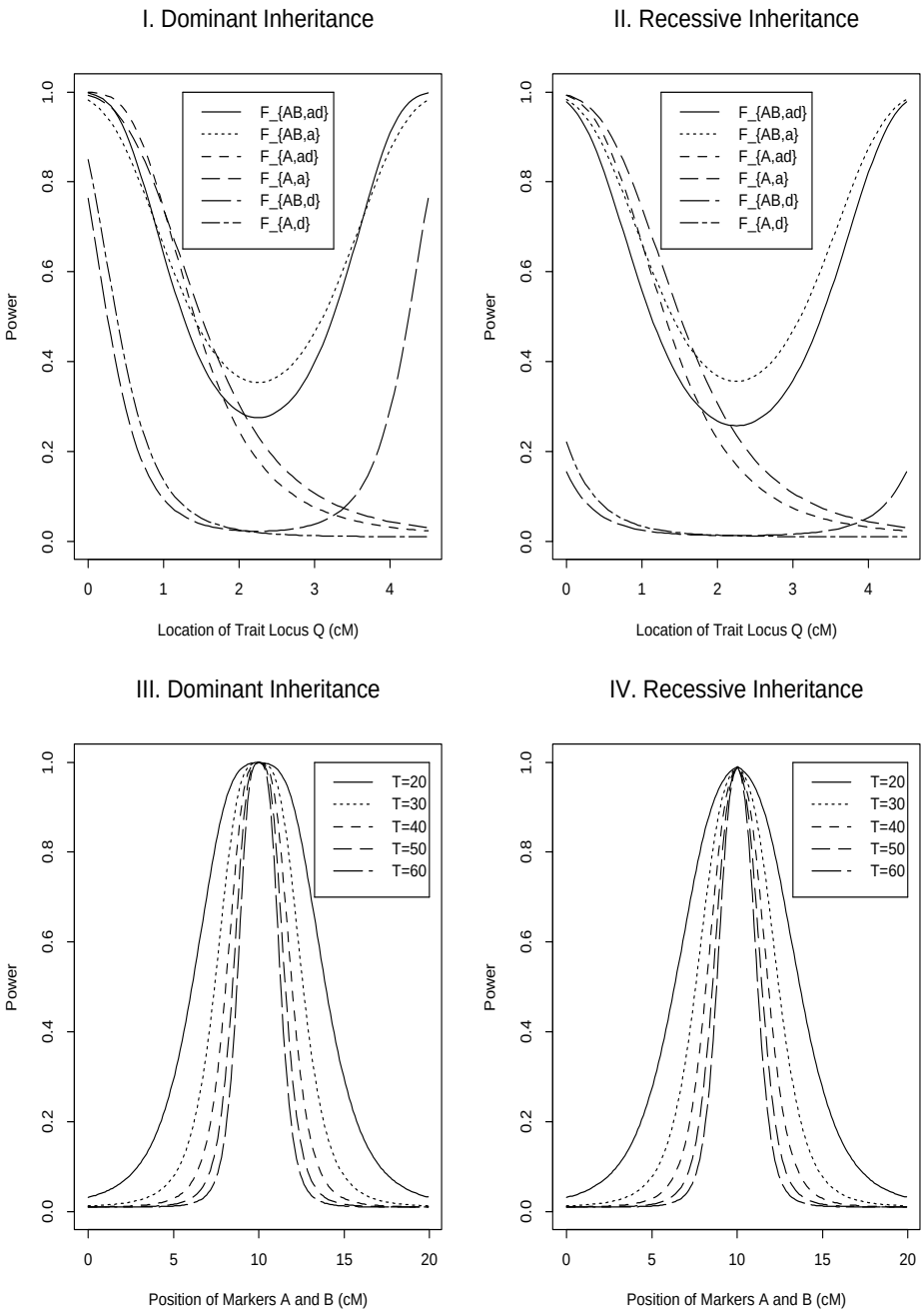

Fig. 11. **Graphs I and II**. Power of test statistics $F_{AB,ad}$, $F_{AB,a}$,$F_{AB,d}$, $F_{A,ad}$, $F_{A,a}$, and $F_{A,d}$ against position of trait locus Q at 0.01 significant level. **Graphs III and IV**. Power of test statistics $F_{AB,ad}$ of different mutation ages against position of markers A and B at 0.01 significant level. The trait locus Q locates at 10cM. The two markers A and B flank the trait locus Q. The other parameters are the same as Graphs I and II.

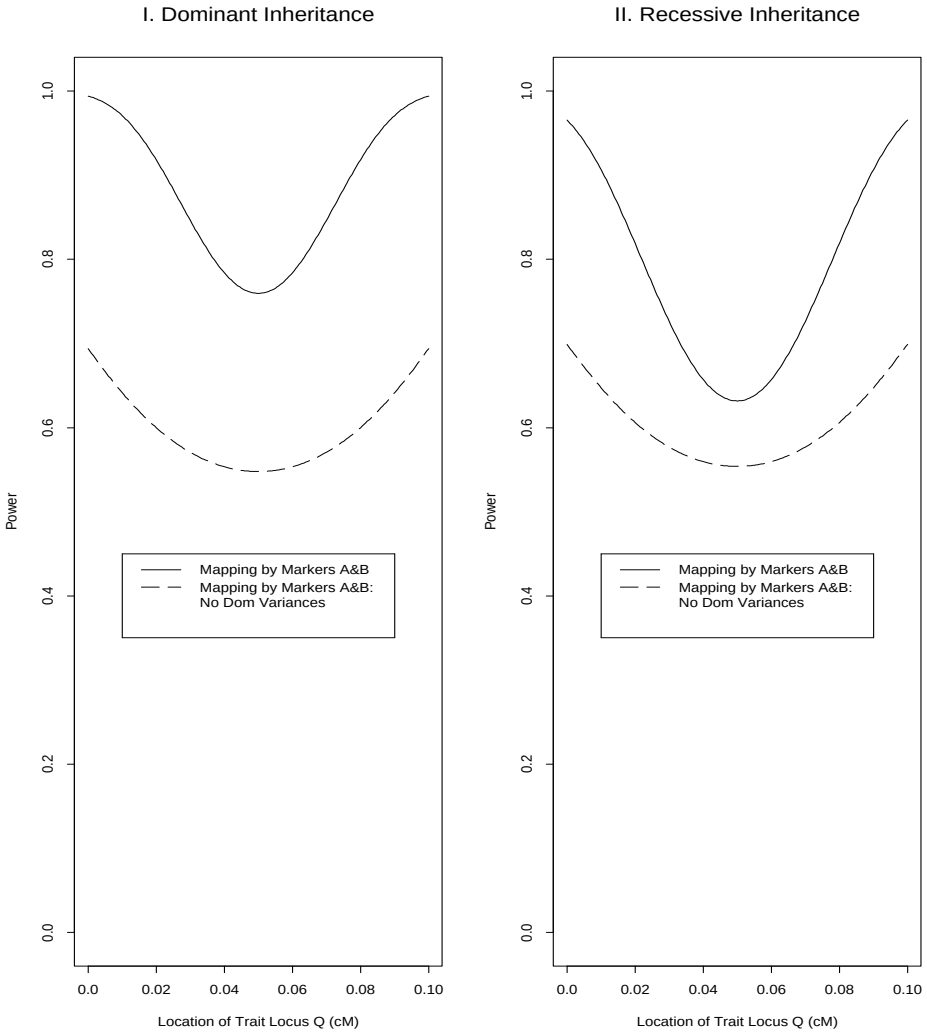

Fig. 12. Power curves of the interval mapping by markers A and B with or without dominant variances against the recombination fraction θ_{AQ} at 0.05 significant level, when $h^2 = 0.35, \lambda_{AB} = 10cM, m = 250, \sigma^2_{Ga} = 0.10, \sigma^2_{Gd} = 0.05, \sigma^2_s = 0$, for a dominant trait $a = d = 1.0, q_1 = 0.60$; and a recessive trait $a = 1.0, d = -0.9, q_1 = 0.40$. Marker A locates at 0cM, and marker B locates at 10cM.

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 $\lambda_{AB} = 10cM$. Marker A locates at 0cM, and marker B locates at 10cM. Figure 12 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 Graph I, we assume $a = d = 1.0$. For a mode of recessive inheritance in Graph II, we assume $a = 1.0, d = -0.9$. By assuming there is no dominance variance at the putative trait locus Q, we include σ^2_{ga} but not σ^2_{gd} in calculating the correlation of sib-pairs. 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.

3.6. Application

We apply the method in this chapter to the Genetic Analysis Workshop 12 German asthma data (Meyers, Wjst and Ober 2001). The data consist of 97 nuclear families, including 415 persons. Seventy-four families have 2 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 146.7cM and marker D1S502 at position 151.2cM 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 149.85cM (p-value, 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 149.85cM ($\hat{\delta}_A = 1.16$, with a p-value 0.0475 by LRT and a p-value 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).

3.7. Discussion

Variance component models are explored to perform the combined linkage and LD mapping based on sibship data with no parental data. The models simultaneously incorporate both linkage information in variance covariance structure of sibship and LD information in the mean coefficients. The mean coefficients account for both LD and the genetic effects such as additive and dominant effects. The linear model of high resolution LD mapping method of Fan and Xiong (2002) is generalized from population to pedigree data, as we consider the variance covariance of pedigree in the model (Fan and Xiong, 2003). In this chapter, we develop the method to accommodate sibship data and population. In the presence of linkage to a particular chromosome region, test of association between QTL and markers is based on coefficient of linear equations. By power and sample size comparisons, generally the power of test statistics for two markers is higher than that for one markers. Furthermore, the power of testing additive genetic effect is higher than that of testing both additive and dominant genetic effect because of an increase of degrees of freedom. In theoretical and simulation study, powers of the proposed model are higher than any of between-pairs and within-pairs ("AbAw") approaches of Fulker et al. (1999) if only one marker is used in analysis. Moreover, the methods are applied to GAW 12 German asthma data and find some effective results.

Fulker and Cardon (1994) suggested the interval mapping approach which has an advantage in detecting the exact location QTL. We propose a way to calculate the

probability of sharing both trait alleles IBD for sibships conditional on the information of flanking markers. Using the formulas of Fulker and Cardon (1994) and the proposed formulas of the probability of sharing both allele IBD in this chapter, we can calculate the trait covariance which is decomposed into additive and dominant genetic variances weighted by IBD status. By numerical calculation and power comparisons, including both additive and dominant variances in the models has a merit in linkage interval mapping when dominant variances exist.

It would be interesting to generalize the proposed method in terms of several views. We generalize the method to use multiple bi-allele markers in the next chapter. It is worthwhile that multi-allelic markers such as micro-satellites or haplotype block could be applied to these models. Since LD mapping is affected very heavily by population subdivisions and admixtures, there is a need to develop methodologies which can deal with the problem in joint LD and linkage mapping. The proposed methods can be applied to general pedigree data.

CHAPTER IV

LINKAGE AND ASSOCIATION MAPPING BY MULTIPLE MARKERS

4.1. Introduction

In linkage disequilibrium (LD) mapping or association study, it is interesting in developing models which use multiple markers simultaneously for high resolution mapping of genetic traits. Usually, mapping single marker on chromosome has low resolution and methods utilizing different markers may lead to different results which make the interpretation complicated. The models using multiple markers may give a consistent result, and lead to greater resolution. Moreover, as large numbers of single nucleotide polymorphisms (SNPs) are available and high throughput genotyping approaches are emerging, there is a need to work out high resolution mapping.

In chapter III, variance component models using two markers are proposed for high resolution mapping of quantitative trait loci (QTL) based on population and pedigree data (Fan and Jung 2003; Fan and Xiong 2002, 2003; Zhao et al. 2001). The genetic effects are orthogonally decomposed into summation of additive and dominant effects. In Abecasis et al. (2000, 2001), Cardon 2000, Fulker et al. (1999) and Sham et al. (2000), an association between-family and association within-family ("AbAw") approach is proposed to decompose the genetic association into effects of between-pairs and within-pairs. The models in chapter III differ from "AbAw" approach in the following views: (1) The "AbAw" approach uses only one marker in analysis, but we use two bi-allelic markers; (2) The way of modeling mean coefficients is different. Fan and Jung (2003) compare our method with the "AbAw" approach, and find that our method is more advantageous for sib-pair data. One may want to notice that it is not clear how to extend the "AbAw" approach to use more than one

markers in analysis (Dr. Fan's communications with Dr. Abecasis and Dr. Sham).

Models in this chapter extend those of the previous chapter, and investigate variance component models in fine association QTL mapping using multiple bi-allelic markers. The models jointly take linkage and linkage disequilibrium information into account. The linkage information is modeled in the variance covariance matrix, and the linkage disequilibrium information is modeled in mean coefficients of trait values like the " AbAw" approach does. By modeling the linkage information in the variance covariance matrix, we may take the advantage of much research of variance component models (Almasy and Blangero 1998; Amos 1994; Amos et al. 1989; Fulker et al. 1995; George et al. 1999; Goldgar and Oniki 1992; Haseman and Elston 1972; Pratt et al. 2000). In the mean time, the linkage disequilibrium information is incorporated into the mean coefficients through indicator variables of marker genotypes, whose validity can be justified intuitively (Fan and Xiong 2000, pages 608-609).

Using the models developed in this chapter, test statistics can be derived for high resolution association mapping. The procedure is to perform appropriate linkage analysis based on a sparse genetic map for prior linkage evidence. Then association study can be worked out using a dense genetic map in the presence of prior linkage information. Likelihood ratio tests (LRT) can be carried out in high resolution association study. For large sample data, likelihood ratio criteria are accurate. Based on the general theory of linear models, F-test statistics can be built to test the association between trait locus and markers in the presence of prior linkage evidence (Graybill 1976). The analytical formulae for the non-centrality parameter approximations are derived for the F-test statistics. The merits of the proposed method are investigated in terms of power and sample size comparison. Using simulation program LDSIMUL kindly provided by Dr. Abecasis, simulation study is performed to explore the power and type I error rates of the proposed test statistics. The proposed methods are com-

pared with the " AbAw" approach (Abecasis, Cardon, and Cookson 2000). Moreover, the method is applied to the Genetic Analysis Workshop (Gaw) 12 German asthma data (Meyers, Wjst and Ober 2001; Wjst et al. 1999).

4.2. Model

Assume that k bi-allelic markers $M_j, j = 1, \cdots, k$ are typed in a region of one chromosome. Suppose a quantitative trait locus Q is located in the region, which has two alleles Q_1 and Q_2 with frequencies q_1 and q_2, respectively. For marker M_j, there are two alleles M_j with frequency P_{M_j} and m_j with frequency P_{m_j}, respectively. For a nuclear family of l children and two parents, let $\mathbf{y} = (y_f, y_m, y_1, \cdots, y_l)^\tau$ be their quantitative traits vector, let $G_j = (G_{fj}, G_{mj}, G_{1j}, \cdots, G_{lj})$ be genotypes at j-th marker locus M_j. Here y_f is a trait value of the father, G_{fj} is the genotype of the father at j-th marker. Likewise, the mother and the i-th child with subscript m and i, respectively. The log-likelihood function for these data is

$$L = -\frac{l+2}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{y} - X\eta)^\tau \Sigma^{-1}(\mathbf{y} - X\eta). \tag{4.1}$$

The components of model (4.1) are defined as follows.

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ 0 & 1 & \rho_0 & \rho_0 & \cdots & \rho_0 \\ \rho_0 & \rho_0 & 1 & \rho_{12} & \cdots & \rho_{1l} \\ \rho_0 & \rho_0 & \rho_{21} & 1 & \cdots & \rho_{2l} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_0 & \rho_0 & \rho_{l1} & \rho_{l2} & \cdots & 1 \end{pmatrix} \sigma^2$$

is a $(l+2) \times (l+2)$ variance-covariance matrix, where $\sigma^2 = \sigma_g^2 + \sigma_H^2 + \sigma_e^2$. Here σ_g^2 is variance explained by the putative QTL Q, σ_H^2 is the variance of familial effects which include shared environment variance and half of the additive polygenic variance, and σ_e^2 is error variance. The genetic variance $\sigma_g^2 = \sigma_{ga}^2 + \sigma_{gd}^2$ is decomposed into additive

and dominant components. $\rho_0 = (\sigma^2_{ga}/2 + \sigma^2_H)/\sigma^2$ is correlation between parents and children, $\rho_{ij} = \rho_{ji} = (\pi_{ijQ}\sigma^2_{ga} + \Delta_{ijQ}\sigma^2_{gd} + \sigma^2_H)/\sigma^2$ is the correlation between the i-th child and the j-th child, π_{ijQ} is the proportion of alleles sharing 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 shared by the i-th child and the j-th child at the putative QTL Q are IBD (Cotterman 1940; Lange 2002; Pratt et al. 2000; Zhu and Elston 2000). For the mean component $X\eta$ of log-likelihood (4.1), we consider

$$y_i = \beta + w_i\gamma + \sum_{j=1}^{k} x_{ij}\alpha_j + \sum_{j=1}^{k} z_{ij}\delta_j + H_i + e_i. \tag{4.2}$$

where β is overall mean, w_i is a row vector of covariates such as gender and age, γ is a column vector of regression coefficients of w_i, and e_i is error term. Assume that e_i is normal $N(0, \sigma^2_e)$. H_i is the familial effect. Assume that H_i is normal $N(0, \sigma^2_H)$. Moreover, H_i and e_i are independent. For $j = 1, \cdots, k$, α_j and δ_j are regression coefficients of the dummy variables x_{ij} and z_{ij}, respectively. Hence, $\eta = (\beta, \gamma^\tau, \alpha_1, \cdots, \alpha_k, \delta_1, \cdots, \delta_k)^\tau$ is a vector of regression coefficients and X is model matrix. Here x_{ij} and z_{ij} are indicator variables, and are defined as follows

$$x_{ij} = \begin{cases} 2P_{m_j} & \text{if } G_{ij} = M_jM_j \\ P_{m_j} - P_{M_j} & \text{if } G_{ij} = M_jm_j \\ -2P_{M_j} & \text{if } G_{ij} = m_jm_j \end{cases} \quad \text{and} \quad z_{ij} = \begin{cases} -P^2_{m_j} & \text{if } G_{ij} = M_jM_j \\ P_{m_j}P_{M_j} & \text{if } G_{ij} = M_jm_j \\ -P^2_{M_j} & \text{if } G_{ij} = m_jm_j \end{cases}.$$

Regression (4.2) uses multiple markers and is a natural generalization of model of our previous work. The objective is to fully use marker information for fine high resolution mapping of QTL.

4.3. Parameter Estimation

4.3.1. Regression Coefficients and Association Study

Denote the measure of LD between trait locus Q and marker M_i by $D_{M_iQ} = P(M_iQ_1) - P_{M_i}q_1$, $i = 1, \cdots, k$, and the measure of LD between marker M_i and marker M_j by $D_{M_iM_j} = P(M_iM_j) - P_{M_i}P_{M_j}, i < j, i, j = 1, \cdots, k$. Let the additive and dominant variance-covariance matrices be

$$V_A = 2\begin{pmatrix} P_{M_1}P_{m_1} & D_{M_1M_2} & \cdots & D_{M_1M_k} \\ D_{M_1M_2} & P_{M_2}P_{m_2} & \cdots & D_{M_2M_k} \\ \vdots & \vdots & \cdots & \vdots \\ D_{M_1M_k} & D_{M_2M_k} & \cdots & P_{M_k}P_{m_k} \end{pmatrix}, V_D = \begin{pmatrix} P^2_{M_1}P^2_{m_1} & D^2_{M_1M_2} & \cdots & D^2_{M_1M_k} \\ D^2_{M_1M_2} & P^2_{M_2}P^2_{m_2} & \cdots & D^2_{M_2M_k} \\ \vdots & \vdots & \cdots & \vdots \\ D^2_{M_1M_k} & D^2_{M_2M_k} & \cdots & P^2_{M_k}P^2_{m_k} \end{pmatrix}.$$

In Appendix M, the coefficients of regression (4.2) are derived as

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = V_A^{-1}\begin{pmatrix} 2D_{M_1Q} \\ \vdots \\ 2D_{M_kQ} \end{pmatrix}\alpha_Q \text{ and } \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_k \end{pmatrix} = V_D^{-1}\begin{pmatrix} D^2_{M_1Q} \\ \vdots \\ D^2_{M_kQ} \end{pmatrix}\delta_Q. \tag{4.3}$$

Equations (4.3) show that the parameters of LD (i.e., D_{M_iQ} and $D_{M_iM_j}$) and gene effect (i.e., α_Q and δ_Q) are contained in the mean coefficients. Model (4.2) simultaneously takes care of the LD and the effects of the putative trait locus Q. The gene substitution effect α_Q is contained in α_i; and the dominant effect δ_Q is contained in $\delta_i, i = 1, \cdots, k$. Therefore, regression (4.2) orthogonally decomposes genetic effect into summation of additive and dominant effects.

Assume that all markers M_i and M_j are in linkage equilibrium (i.e., $D_{M_iM_j} = 0, i, j = 1, \cdots, k, i \neq j$). The coefficients of additive and dominant effects are given by $\alpha_1 = \frac{D_{M_1Q}}{P_{M_1}P_{m_1}}\alpha_Q$, $\cdots$, $\alpha_k = \frac{D_{M_kQ}}{P_{M_k}P_{m_k}}\alpha_Q$ and $\delta_1 = \frac{D^2_{M_1Q}}{P^2_{M_1}P^2_{m_1}}\delta_Q, \cdots, \delta_k = \frac{D^2_{M_kQ}}{P^2_{M_k}P^2_{m_k}}\delta_Q$. That means markers $M_1, \cdots, M_k$ independently contribute to the analysis of the trait values. Usually, the markers M_i can be in LD, especially when they are locate in a narrow chromosome region. Equations (4.3) rightly use the LD information of

markers M_i in the analysis.

Linkage analysis can be performed by considering a reduced variance component model $y_i = \beta + w_i\gamma + H_i + e_i$. This initial study can identify prior linkage evidence of the trait values to a specific chromosome region based on a sparse genetic map. Suppose that prior linkage evidence is provided by an initial linkage study. Based on a dense genetic map, high resolution association mapping of the QTL can be carried out by fitting the full model (4.2). First, assume that linkage is confirmed in a chromosome region by the significant presence of both the gene substitution and dominant effects, i.e., $\alpha_Q \neq 0$ and $\delta_Q \neq 0$. Based on equations (4.3), the existence of LD between markers M_i $(i = 1, \cdots, k)$ and trait locus Q can be tested by $H_{ad} : \alpha_1 = \cdots = \alpha_k = \delta_1 = \cdots = \delta_k = 0$. Second, assume that linkage is supported by the significant presence of the gene substitution effect, but not the dominant effect, i.e., $\alpha_Q \neq 0$ and $\delta_Q = 0$. The existence of LD can be tested by $H_a : \alpha_1 = \cdots = \alpha_k = 0$. Third, assume that linkage is supported by the significant presence of the dominant effect, but not the gene substitution effect, i.e., $\alpha_Q = 0$ and $\delta_Q \neq 0$. The existence of LD can be tested by $H_d : \delta_1 = \cdots = \delta_k = 0$.

Evidence of association can be evaluated by likelihood ratio test (LRT) procedure. For instance, let L_{ad} be the log-likelihood under the alternative hypothesis of H_{ad}, and L_0 be the log-likelihood under the null hypothesis H_{ad}. Then, the quantity $2[L_{ad} - L_0]$ is asymptotically distributed as χ^2. Notice that there are only k measures of LD, $D_{M_1Q}, \cdots, D_{M_kQ}$, under the alternative hypothesis H_{ad}. In data analysis, the number of coefficients $\alpha_i, \delta_i, i =, 1 \cdots, k$, which are significantly different from 0, should be less than or equal to k. This number is the degrees of freedom of the likelihood ratio test $2[L_{ad} - L_0]$. For large sample data, the likelihood ratio test is accurate based on the statistical theory. In this paper, we will develop a F-test procedure based on linear model theory (Graybill 1976). Before that, we will discuss

the variance-covariance first.

4.3.2. Variance-Covariances

Denote the recombination fraction between trait locus Q and marker M_i by $\theta_{M_iQ}, i = 1, \cdots, k$. Likewise, the recombination fraction between markers M_i and M_j are defined by $\theta_{M_iM_j}$. Following Fulker et al. (1995)) and Alamsy and Blangero (1998), we propose a multi-point interval mapping method to estimate the proportion π_{ijQ} of allele sharing IBD at a putative QTL Q for a sib-pair i and j by

$$\begin{aligned} \hat{\pi}_{ijQ} &= \mathrm{E}\left(\pi_{ijQ}|I_{M_1}, I_{M_2}, \cdots, I_{M_k}\right) \\ &= \alpha_\pi + \beta_{\pi M_1}\pi_{ijM_1} + \beta_{\pi M_2}\pi_{ijM_2} + \cdots + \beta_{\pi M_k}\pi_{ijM_k}, \end{aligned} \tag{4.4}$$

where π_{ijM_l} is the proportions of alleles sharing IBD at the marker M_l for $l = 1, \cdots, k$. The coefficients $\alpha_\pi, \beta_{\pi M_1}, \cdots, \beta_{\pi M_k}$ are derived in Appendix N as follows

$$\begin{pmatrix} \beta_{\pi M_1} \\ \beta_{\pi M_2} \\ \vdots \\ \beta_{\pi M_k} \end{pmatrix} = \begin{pmatrix} 1 & (1-2\theta_{M_1M_2})^2 & \cdots & (1-2\theta_{M_1M_k})^2 \\ (1-2\theta_{M_1M_2})^2 & 1 & \cdots & (1-2\theta_{M_2M_k})^2 \\ \vdots & \vdots & \vdots & \vdots \\ (1-2\theta_{M_1M_k})^2 & (1-2\theta_{M_2M_k})^2 & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ (1-2\theta_{M_2Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \end{pmatrix}.$$

And α_π is estimated as $\alpha_\pi = 1 - \beta_{\pi M_1} - \beta_{\pi M_2} - \cdots - \beta_{\pi M_k}$. If marker M_l coincides with QTL Q, it can be shown that $\beta_{\pi M_l} = 1$ and $\alpha_\pi = 0$, $\beta_{\pi M_i} = 0, i \neq l$. Hence $\hat{\pi}_{ijQ} = \pi_{ijM_l}$. To estimate Δ_{ijQ} of the probability of sharing 2 alleles IBD for a sib-pair, consider

$$\begin{aligned} \hat{\Delta}_{ijQ} &= \mathrm{E}\left(\Delta_{ijQ}|I_{M_1}, I_{M_2}, \cdots, I_{M_k}\right) \\ &= \alpha + \beta_{M_1}\pi_{ijM_1} + \cdots + \beta_{M_k}\pi_{ijM_k} + r_{M_1}\Delta_{ijM_1} + \cdots + r_{M_k}\Delta_{ijM_k}, \end{aligned} \tag{4.5}$$

where Δ_{ijM_l} is the probability of sharing 2 allele IBD at marker M_l for $l = 1, \cdots, k$. The coefficients $(r_{M_1}, \cdots, r_{M_k})^\tau$ are derived in Appendix O as follows

$$\begin{pmatrix} r_{M_1} \\ r_{M_2} \\ \vdots \\ r_{M_k} \end{pmatrix} = \begin{pmatrix} 1 & (1-2\theta_{M_1M_2})^4 & \cdots & (1-2\theta_{M_1M_k})^4 \\ (1-2\theta_{M_1M_2})^4 & 1 & \cdots & (1-2\theta_{M_2M_k})^4 \\ \vdots & \vdots & \vdots & \vdots \\ (1-2\theta_{M_1M_k})^4 & (1-2\theta_{M_2M_k})^4 & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} (1-2\theta_{M_1Q})^4 \\ (1-2\theta_{M_2Q})^4 \\ \vdots \\ (1-2\theta_{M_kQ})^4 \end{pmatrix}.$$

The remaining coefficients are given in Appendix O by $\begin{pmatrix} \beta_{M_1} \\ \beta_{M_2} \\ \vdots \\ \beta_{M_k} \end{pmatrix} = \begin{pmatrix} \beta_{\pi M_1} \\ \beta_{\pi M_2} \\ \vdots \\ \beta_{\pi M_k} \end{pmatrix} - \begin{pmatrix} r_{M_1} \\ r_{M_2} \\ \vdots \\ r_{M_k} \end{pmatrix}$. The α in equation (4.5) is $\alpha = 1 - \beta_{M_1} - \cdots - \beta_{M_k} - r_{M_1} - \cdots - r_{M_k}$. Again, if marker M_l coincides with QTL Q, it can be shown that $\hat{\Delta}_{ijQ} = \Delta_{ijM_l}$.

4.4. Test Statistics and Non-centrality Parameter

4.4.1. Combined analysis of population and family data

We assume that the data are composed of three sub-sample: n individuals of a population, m trio families with both parents and a single child, and s nuclear families each has both parents and two offspring. Furthermore, we assume that n, m and s are sufficiently large, so that large sample theory applies. We may include data of nuclear families with both parents and more than two offspring. The principle of the following paragraphs can be extended to such families if the number of the families is large enough to apply the large sample theory.

The coefficients of regression (4.2) can be written as $\eta = (\beta, \alpha_1, \cdots, \alpha_k, \delta_1, \cdots, \delta_k)^\tau$ if there are no covariates. Consider the overall log-likelihood $L = \sum_{i=1}^{I} L_i, I = n + m + s$, where L_i is the log-likelihood of trait value $\mathbf{y}_i$ of the i-th family or individual. Let Σ_i be the variance-covariance matrix of trait value $\mathbf{y}_i$, and X_i be its design matrix. Denote the all trait values by $\mathbf{y} = (\mathbf{y}_1^\tau, \cdots, \mathbf{y}_I^\tau)^\tau$, the total variance-covariance matrix by $\Sigma = diag(\Sigma_1, \cdots, \Sigma_I)$, and model matrix by $X = (X_1^\tau, \cdots, X_I^\tau)^\tau$.

Let $N = n + 3m + 4s$ be the total number of individuals. The estimate of η is $\hat{\eta} = \left[X^\tau \hat{\Sigma}^{-1} X\right]^{-1} X^\tau \hat{\Sigma}^{-1} \mathbf{y} = \left[\sum_{i=1}^{I} X_i^\tau \hat{\Sigma}_i^{-1} X_i\right]^{-1} \sum_{i=1}^{I} X_i^\tau \hat{\Sigma}_i^{-1} \mathbf{y}_i$.

The non-centrality parameters of appropriate test statistics of genetic effects and LD coefficients can be calculated as like subsection 3.3.1. First, one may construct test statistic for each of three hypotheses: $H_{ad} : \alpha_1 = \cdots = \alpha_k = \delta_1 = \cdots = \delta_k = 0$; $H_a : \alpha_1 = \cdots = \alpha_k = 0$; $H_d : \delta_1 = \cdots = \delta_k = 0$. The non-centrality parameter of each hypothesis can be calculated using the theory in Chapter 6, Graybill (1976). Let H be $q \times (2k+1)$ matrix of rank q. The test statistic for hypothesis $H\eta = 0$ is

$$F = \frac{(H\hat{\eta})^\tau [H(X^\tau \hat{\Sigma}^{-1} X)^{-1} H^\tau]^{-1} (H\hat{\eta})}{\mathbf{y}^\tau (\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} X (X^\tau \hat{\Sigma}^{-1} X)^{-1} X^\tau \hat{\Sigma}^{-1}) \mathbf{y}} \frac{(N - 2k - 1)}{q}$$

with non-central $F(q, N - (2k+1))$ distribution. The non-centrality parameter is $\lambda = (H\eta)^\tau [H(X^\tau \Sigma^{-1} X)^{-1} H^\tau]^{-1} (H\eta)$. Under the assumption of large sample sizes n, m and s, we show in Appendix P that

$$X^\tau \Sigma^{-1} X = \sum_{i=1}^{n+m+s} X_i^\tau \Sigma_i^{-1} X_i \approx diag(a_1, a_2 V_A, a_3 V_D)/\sigma^2, \tag{4.6}$$

where a_1, a_2 and a_3 are constants given by equations (P.7) in Appendix P.

The additive variance $\sigma^2_{ga} = 2q_1 q_2 \alpha_Q^2$ and the dominant variance $\sigma^2_{gd} = (q_1 q_2)^2 \delta_Q^2$ are expressed in terms of the average effect of gene substitution α_Q and the dominance deviation δ_Q. Let I_k and I_{2k} be k and $2k$ dimension identity matrices. Moreover, let $O_{k \times l}$ be $k \times l$ zero matrix. To test hypothesis $H_a : \alpha_1 = \cdots = \alpha_k = 0$, the test matrix $H = (O_{k \times 1}, I_k, O_{k \times k})$. Let us denote the test statistic as $F_{k,a}$. The non-centrality parameter is approximated by

$$\lambda_{k,a} \approx \frac{a_2}{\sigma^2} (\alpha_1, \cdots, \alpha_k) V_A \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \frac{4a_2}{\sigma^2} \alpha_Q^2 (D_{M_1 Q}, \cdots, D_{M_k Q}) V_A^{-1} \begin{pmatrix} D_{M_1 Q} \\ \vdots \\ D_{M_k Q} \end{pmatrix}$$

$$= \frac{a_2\sigma^2_{ga}}{\sigma^2 q_1 q_2}(D_{M_1Q},\cdots,D_{M_kQ})(V_A/2)^{-1}\begin{pmatrix} D_{M_1Q} \\ \vdots \\ D_{M_kQ}\end{pmatrix}.$$

To test hypothesis $H_d : \delta_1 = \cdots = \delta_k = 0$, the test matrix $H = (O_{k\times 1}, O_{k\times k}, I_k)$. Let us denote the test statistic as $F_{k,d}$. The non-centrality parameter is approximated by

$$\begin{aligned}\lambda_{k,d} &\approx \frac{a_3}{\sigma^2}(\delta_1,\cdots,\delta_k)V_D\begin{pmatrix}\delta_1\\ \vdots \\ \delta_k\end{pmatrix} = \frac{a_3}{\sigma^2}\delta_Q^2(D^2_{M_1Q},\cdots,D^2_{M_kQ})V_D^{-1}\begin{pmatrix} D^2_{M_1Q} \\ \vdots \\ D^2_{M_kQ}\end{pmatrix} \\ &= \frac{a_3\sigma^2_{gd}}{\sigma^2 q_1^2 q_2^2}(D^2_{M_1Q},\cdots,D^2_{M_kQ})V_D^{-1}\begin{pmatrix} D^2_{M_1Q} \\ \vdots \\ D^2_{M_kQ}\end{pmatrix}.\end{aligned}$$

To test hypothesis $H_{ad} : \alpha_1 = \cdots = \alpha_k = \delta_1 = \cdots = \delta_k = 0$, the test matrix $H = (O_{2k\times 1}, I_{2k})$. Let us denote the test statistic as $F_{k,ad}$. The non-centrality parameter is $\lambda_{k,ad} \approx \lambda_a + \lambda_d$, i.e., $\lambda_{k,ad}$ is decomposed into the summation of additive and dominant non-centrality parameters.

4.4.2. Nuclear family

To make comparison with the results of Table 4 of Abecasis, Cardon, and Cookson (2000), we consider I families each has both parents and l offspring. Let $N = I(l+2)$ be the total number of individuals. The other notations are defined in a similar way as above. Suppose that variance-covariance matrices of the I families are the same, i.e., $\Sigma_1 = \cdots = \Sigma_I$. Denote $\Sigma_i^{-1} = \frac{1}{\sigma^2}(\gamma_{hj})_{(l+2)\times(l+2)}$. If the sample sizes N is large enough, we show in Appendix Q that

$$X^\tau\Sigma^{-1}X/I = \sum_{i=1}^{I} X_i^\tau\Sigma_i^{-1}X_i/I \approx diag(\sum_{h,j}\gamma_{hj}, b_1V_A, b_2V_D)/\sigma^2, \tag{4.7}$$

where b_1 and b_2 are constants given by equations (Q.1) in Appendix Q. The approximation of non-centrality parameter of statistic $F_{k,a}$ is

$$\lambda_{k,a} \approx \frac{b_1 I \sigma_{ga}^2}{\sigma^2 q_1 q_2} (D_{M_1Q}, \cdots, D_{M_kQ})(V_A/2)^{-1} \begin{pmatrix} D_{M_1Q} \\ \vdots \\ D_{M_kQ} \end{pmatrix}.$$

4.5. Type I Error Rates

To evaluate the type I error rates of the proposed method, nuclear families are generated by simulation program LDSIMUL provided by Dr. Abecasis. Five test cases are considered in type I error rate calculation, which are taken from Table 2 of Abecasis, Cardon, and Cookson (2000). Trait values are constructed by normal distribution with mean 0 and total variance $\sigma^2 = 100$ except test case of **Admixture**. Here $\sigma^2 = \sigma_{ga}^2 + \sigma_H^2 + \sigma_e^2$ is the summation of the additive major gene effect σ_{ga}^2, the variance of familial effects σ_H^2, and the error variance σ_e^2. In each model except the **Admixture**, a bi-allelic marker M_1 is simulated with allele frequency $P_{M_1} = 0.5$. In the test cases of **Null, Familiality,** and **Admixture**, no major gene effect is assumed, i.e., $\sigma_{ga}^2 = 0$. In the test cases of **Linkage** and **Composite**, major gene effect is assumed, and marker M_1 coincides with the QTL Q, i.e., recombination fraction $\theta_{M_1Q} = 0$; in the meantime, linkage equilibrium is assumed between QTL Q and the marker M_1, i.e., $D_{M_1Q} = 0$. In the test case of **Admixture**, population admixture is generated by mixing families equally drawn from one of the two sub-populations A and B. In both sub-populations A and B, no major gene effect or familial effect is assumed, i.e., $\sigma_{ga}^2 = \sigma_H^2 = 0$. However, the trait mean of sub-population A is fixed as 10 and the variance is fixed as 100, and the marker allele frequency P_{M_1} is taken as 0.7 in sub-population A.

The trait mean of sub-population B is fixed as 0 and the variance is fixed as 100,

Table VI. Type I Error Rates (%) at a 0.05 significant level. The parameters are the same as those of Table 2 of Abecasis, Cardon, and Cookson (2000). The total variance is fixed as $\sigma^2 = 100$ (see text for explanation of **Admixture** case). **Null**: no major gene effect or familial effect $\sigma_g^2 = \sigma_H^2 = 0$; **Familiality**: large familial effect $\sigma_H^2 = 50$, but no major gene effect $\sigma_g^2 = 0$; **Admixture**: no major gene effect or familial effect $\sigma_g^2 = \sigma_H^2 = 0$, but with population admixture; **Linkage**: large linkage effect $\sigma_g^2 = \sigma_{ga}^2 = 30, \theta_{M_1Q} = 0$, but no familial effect $\sigma_H^2 = 0$; **Composite**: large linkage effect $\sigma_g^2 = \sigma_{ga}^2 = 20, \theta_{M_1Q} = 0$, and large familial effect $\sigma_H^2 = 30$. There is no linkage disequilibrium between QTL and marker M_1 ($D_{M_1Q} = 0$).

Offspring in Each family	**Test Case**	**Error Rates When Total No. of Offspring is**					
		120		**240**		**480**	
		LRT	$\hat{F}_{1,a}$	LRT	$\hat{F}_{1,a}$	LRT	$\hat{F}_{1,a}$
1	Null	6.5	7.0	5.1	6.5	5.8	6.9
	Familiality	5.4	8.3	5.2	8.1	5.3	9.5
	Admixture	6.4	9.7	5.2	9.3	5.3	8.9
2	Null	4.6	2.9	4.8	2.8	4.5	2.9
	Familiality	4.2	4.4	3.6	3.8	4.7	4.2
	Admixture	5.0	5.2	6.1	5.4	4.9	4.3
	Linkage	5.5	4.9	5.0	3.9	5.0	4.6
	Composite	5.6	7.0	5.8	6.2	5.6	5.5
4	Null	4.9	1.7	4.3	1.5	3.6	1.2
	Familiality	5.2	4.8	4.2	3.4	4.8	3.3
	Admixture	5.5	3.2	5.4	3.5	4.2	2.6
	Linkage	5.3	3.6	5.4	3.7	4.9	3.8
	Composite	5.3	4.9	5.3	3.4	4.1	2.6
8	Null	4.2	1.4	5.0	1.0	4.7	1.0
	Familiality	4.7	4.5	5.1	4.8	4.4	3.6
	Admixture	3.5	2.6	5.5	3.2	4.4	3.1
	Linkage	6.1	3.7	4.3	2.8	4.6	2.8
	Composite	5.8	4.5	5.5	3.8	3.7	2.8

and the marker allele frequency P_{M_1} is taken as 0.3 in sub-population B. Therefore, the total variance in the mixing population is $\sigma^2 = 125$. The admixture contributed to $(10-0)^2/[4] = 0.20$ of the total variance. The other related parameters are given in the legend of Table VI.

Table VI presents type I error rates of likelihood ratio tests and F-test statistics. The type I error rates are calculated as the proportions of 1000 simulation data sets which give significant result at a 0.05 significant level based on $F_{1,a}$ and likelihood ratio test statistic, respectively. The results show that the type I error rates of likelihood ratio tests are around the 0.05 nominal significant level in most cases. Hence, the proposed model works well. The type I error rates of trio families (i.e., family with only one offspring) are usually higher than those of nuclear family data which contain multiple offspring. In particular, the type I error rates of F-test are high for trio families. For nuclear family data which contain multiple offspring, the type I error rates of F-test are similar or smaller than those of the likelihood ratio tests. In an association study, false positives due to population stratifications are usually a big issue. From the results of Table VI, the type I error rates in the **Admixture** case are reasonable for nuclear family data which contain multiple offspring. For trio families, the type I error rates of F-test in the **Admixture** case are high.

4.6. Powers and Their Comparison

4.6.1. Comparison with the "AbAw" approach

Denote the heritability by h^2, which is defined as $h^2 = \sigma^2_{ga}/\sigma^2$ (Falconer and Mackay 1996). To compare the method proposed in this paper with the "AbAw" approach of Abecasis, Cardon, and Cookson (2000), we present power comparison in Table VII. The parameters are the same as those of Table 4 of Abecasis, Cardon, and Cookson

(2000): $q_1 = P_{M_1} = 0.5, h^2 = 0.1, \sigma^2 = 100, \sigma^2_{ga} = 10, \sigma^2_s = 30, \sigma^2_e = 60$. Besides, $D' = D_{M_1Q}/D_{max}$ and $D_{max} = min(P_{M_1}, q_1) - P_{M_1}q_1$. In the columns of ACC, the results are taken from Table 4 of Abecasis, Cardon, and Cookson (2000). In the columns $(F_{1,a}, \hat{F}_{1,a}, LRT)^\tau$, the power of $F_{1,a}$ is calculated based on approximation of non-centrality parameter $\lambda_{1,a}$ of test statistic $F_{1,a}$ at a 0.001 significant level; the power of $\hat{F}_{1,a}$ and LRT are calculated as the proportions of 1000 simulation data sets which give significant result at the 0.001 significant level based on $F_{1,a}$ and likelihood ratio test statistic, respectively. For each simulated dataset, certain number nuclear families are simulated via LDSIMUL. For instance, for one sib per family, 480 trio families are simulated in each simulated dataset.

The results of Table VII clearly show that the proposed F-tests $F_{1,a}$ and likelihood ratio tests are much more powerful than the "AbAw" approach. When $D' = D_{M_1Q}/D_{max} > 25\%$, it is possible to achieve considerable power. When $D' = D_{M_1Q}/D_{max} > 50\%$, the statistic $F_{1,a}$ is powerful since the power is higher than $(F_{1,a}, \hat{F}_{1,a}, LRT) = (0.560, 0.333, 0.322)$ for a sample with a total number of 480 sibs. Moreover, the power to detect association decreases as the size of sibship increases. Hence, families of large sibship sizes contain less LD information than families of small sibship sizes. The readers may want to notice that this result is consistent with findings in Fan and Xiong (2003). In Figure 3 of Fan and Xiong (2003), p131, population based method is shown to be more powerful than the family based method for the same number of individuals.

In addition, the results of Table VII show that the empirical power of $\hat{F}_{1,a}$ is similar to that of likelihood ratio test. This implies that in large sample, the two tests provide similar power. For nuclear families of small sibship size (i.e., number of sibs is ≤ 4), the empirical power of $\hat{F}_{1,a}$ and likelihood ratio test (LRT) is similar to the power based on the theoretical approximations $\lambda_{1,a}$ of $F_{1,a}$.

Table VII: Power comparison with results of Table 4 of Abecasis, Cardon, and Cookson (2000). In the columns of ACC, the power is taken from Table 4 of Abecasis, Cardon, and Cookson (2000). In the columns $(F_{1,a}, \hat{F}_{1,a}, LRT)^\tau$, the power of $F_{1,a}$ is calculated based on approximation of non-centrality parameter $\lambda_{1,a}$ of test statistic $F_{1,a}$ at 0.001 significant level; the power of $\hat{F}_{1,a}$ and LRT are calculated as the proportions of 1000 simulation data sets which give significant result at 0.001 significant level based on $F_{1,a}$ and likelihood ratio test statistic, respectively. For each simulated dataset, certain number nuclear families are simulated via LDSIMUL.

D' %	**One sib per family**		**Two sibs per family**		**Three sibs per family**		**Four sibs per family**		**Five sibs per family**		**Six sibs per family**		**Eight sibs per family**	
	No. of families													
	480		240		160		120		96		80		60	
	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT	ACC	$F_{1,a}$ $\hat{F}_{1,a}$ LRT
0	0.2	0.1 0.0 0.0	0.0	0.1 0.0 0.0	0.1	0.1 0.1 0.1	0.1	0.1 0.4 0.4	0.2	0.1 0.1 0.1	0.1	0.1 0.0 0.0	0.0	0.1 0.2 0.2
25	2.1	33.1 35.2 34.9	1.8	15.4 16.7 16.5	2.0	10.6 11.1 10.9	2.6	8.5 7.6 7.4	3.0	7.4 4.4 4.3	2.1	6.7 4.3 4.1	2.1	5.8 3.8 3.7
50	19.5	99.2 99.5 99.5	22.9	89.4 89.3 89.6	24.8	78.6 80.3 80.0	26.7	70.7 66.9 66.9	26.7	65.2 54.8 53.9	27.2	61.2 45.3 45.1	23.9	56.0 33.3 32.2
75	69.3	100 100 100	72.6	100 100 100	74.2	99.8 99.8 99.8	76.9	99.3 99.6 99.6	76.0	98.7 96.9 96.9	76.5	98.1 95.4 95.3	75.4	96.9 86.2 85.3
100	97.4	100 100 100	97.7	100 100 100	98.3	100 100 100	98.4	100 100 100	98.2	100 100 100	98.4	100 100 100	98.5	100 99.9 99.9

For nuclear families of large sibship size (i.e., number of sibs is ≥ 5), the empirical power of $\hat{F}_{1,a}$ and likelihood ratio test (LRT) is smaller than the power based on the theoretical approximations $\lambda_{1,a}$ of $F_{1,a}$. Hence, the approximations of non-centrality parameter $\lambda_{1,a}$ is accurate in the case of small sibship size, but less accurate in the case of large sibship size.

4.6.2. Comparisons of Sample Size and Power of LD mapping

Power and sample size calculations are performed to investigate the merits of the proposed method. Figure 13 shows the power curves of the test statistics $F_{4,a}, F_{3,a}$, $F_{2,a}$,$F_{4,d}$,$F_{3,d}$, and $F_{2,d}$ against the linkage disequilibrium coefficient D_{M_1Q} at a 0.01 significant level for a dominant mode of inheritance ($a = d = 1.0$) and a recessive mode of inheritance ($a = 1.0, d = -0.5$). The related parameters are given in the legend of the figure. Generally, the power of $F_{4,a}$ using 4 markers in the model is higher than that of $F_{3,a}$ using 3 markers, which in turn is higher than that of $F_{2,a}$ using 2 markers. Hence, multiple marker analysis is advantageous. The power of $F_{k,d}$ is usually minimal unless the LD between locus Q and marker M_1 is very strong for the dominant mode of inheritance. Figure 14 provides the power of the test statistics $F_{4,a}, F_{3,a}, F_{2,a}$, $F_{4,d}, F_{3,d}$, and $F_{2,d}$ against heritability h^2 at a 0.01 significant level for a dominant mode of inheritance ($a = d = 1.0$) and a recessive mode of inheritance ($a = 1.0, d = -0.5$), respectively. In addition to the merits shown in Figure 13, the power of the test statistics $F_{4,a}, F_{3,a}, F_{2,a}$ is high when heritability h^2 is larger than 0.10 for both modes of inheritance.

Figure 15 shows the power of test statistics $F_{4,a}$, $F_{3,a}$, $F_{2,a}$, and $F_{1,a}$ against the trait allele frequency q_1 (Graph I) or marker allele frequency P_{M_1} (Graph II) at a 0.01 significant level for an additive mode of inheritance $a = 1.0, d = 0.0$, respectively. The other parameters are given in the legend of the figure. From Graph I of the

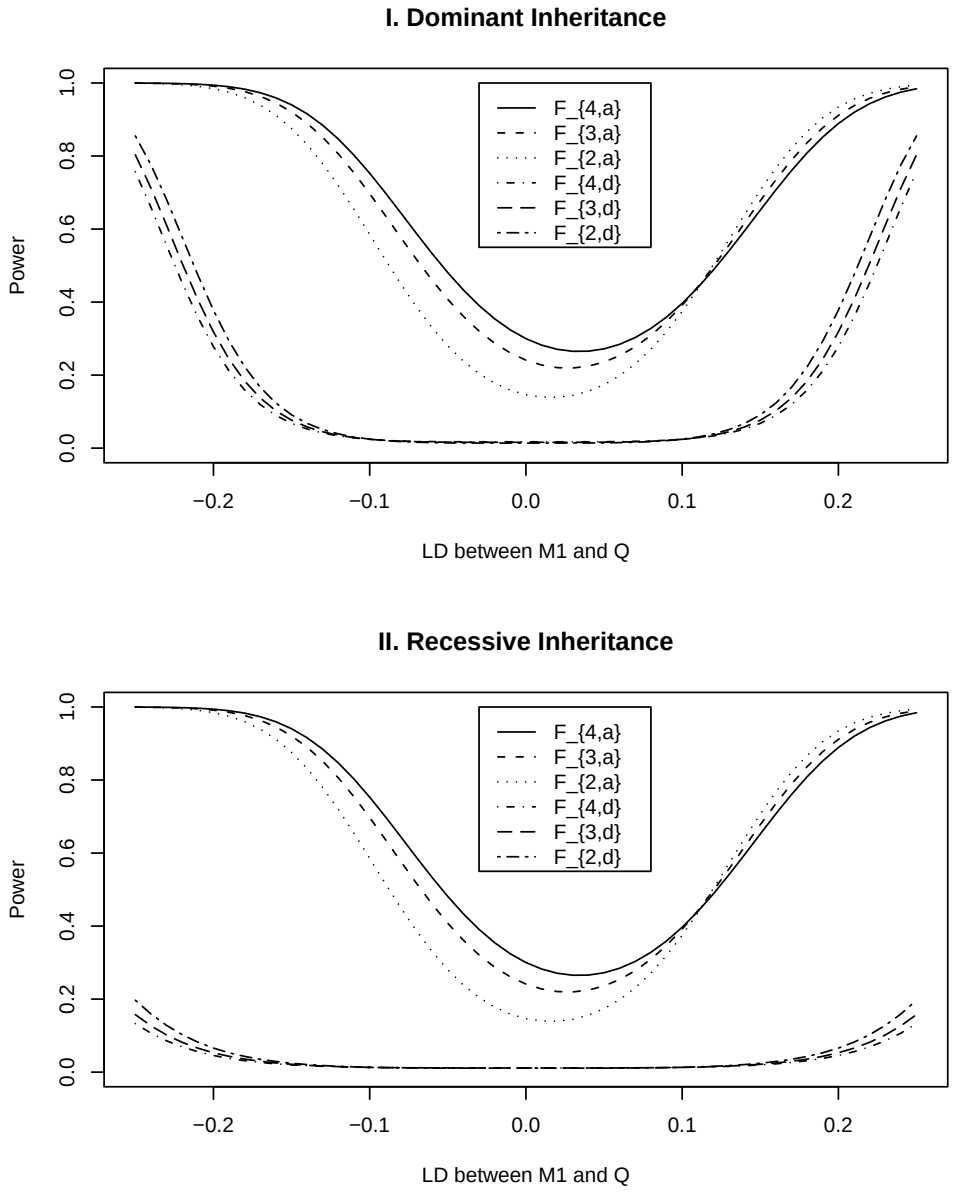

Fig. 13. Power curves of test statistics $F_{4,a}$, $F_{3,a}$, $F_{2,a}$, $F_{4,d}$, $F_{3,d}$, and $F_{2,d}$ against the measure of LD between M_1 and Q at a 0.01 significant level, when $q_1 = 0.50, P_{M_i} = 0.50, i = 1, 2, 3, 4, D_{M_iQ} = 0.08, i = 2, 3, 4, D_{M_iM_j} = 0.05, i \neq j$, $\pi_{12Q} = 0.5, \delta_{12Q} = 0.25$, heritability $h^2 = 0.15$, familial effect variance $\sigma_H^2 = 0.10$, and sample size $n = 40, m = 30, s = 20$ for a dominant mode of inheritance $a = d = 1.0$ (Graph I), and a recessive mode of inheritance $a = 1.0, d = -0.5$ (Graph II), respectively.

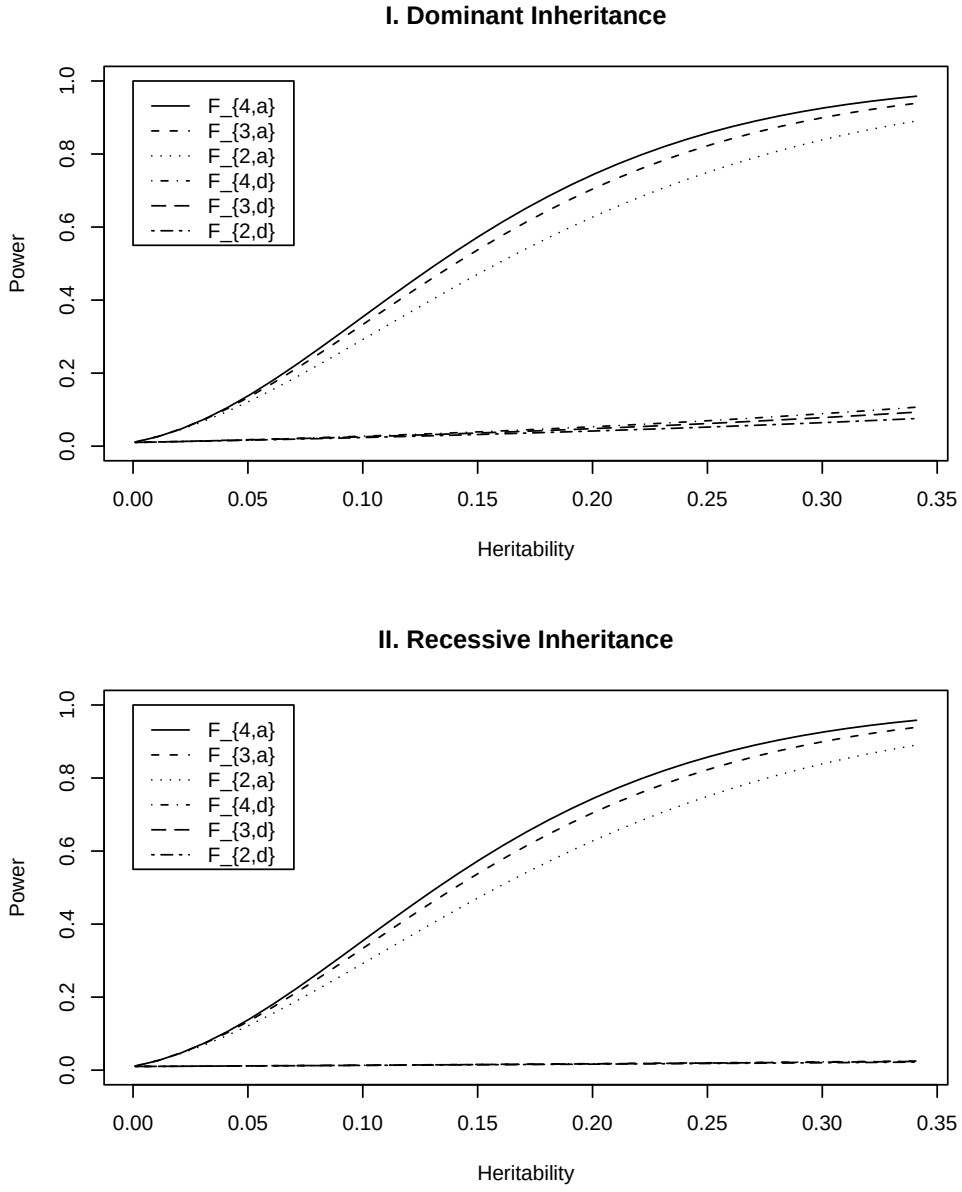

Fig. 14. Power of test statistics $F_{4,a}$, $F_{3,a}$, $F_{2,a}$, $F_{4,d}$, $F_{3,d}$, and $F_{2,d}$ against the heritability h^2 at a 0.01 significant level, when $q_1 = 0.5, P_{M_i} = 0.5, D_{M_iQ} = 0.1, D_{M_iM_j} = 0.05, i, j = 1, 2, 3, 4, i \neq j$, $\pi_{12Q} = 0.5, \delta_{12Q} = 0.25$, $\sigma_H^2 = 0.1$, and sample size $n = 40, m = 30, s = 20$ for a dominant mode of inheritance $a = d = 1.0$ (Graph I), and a recessive mode of inheritance $a = 1.0, d = -0.5$ (Graph II), respectively.

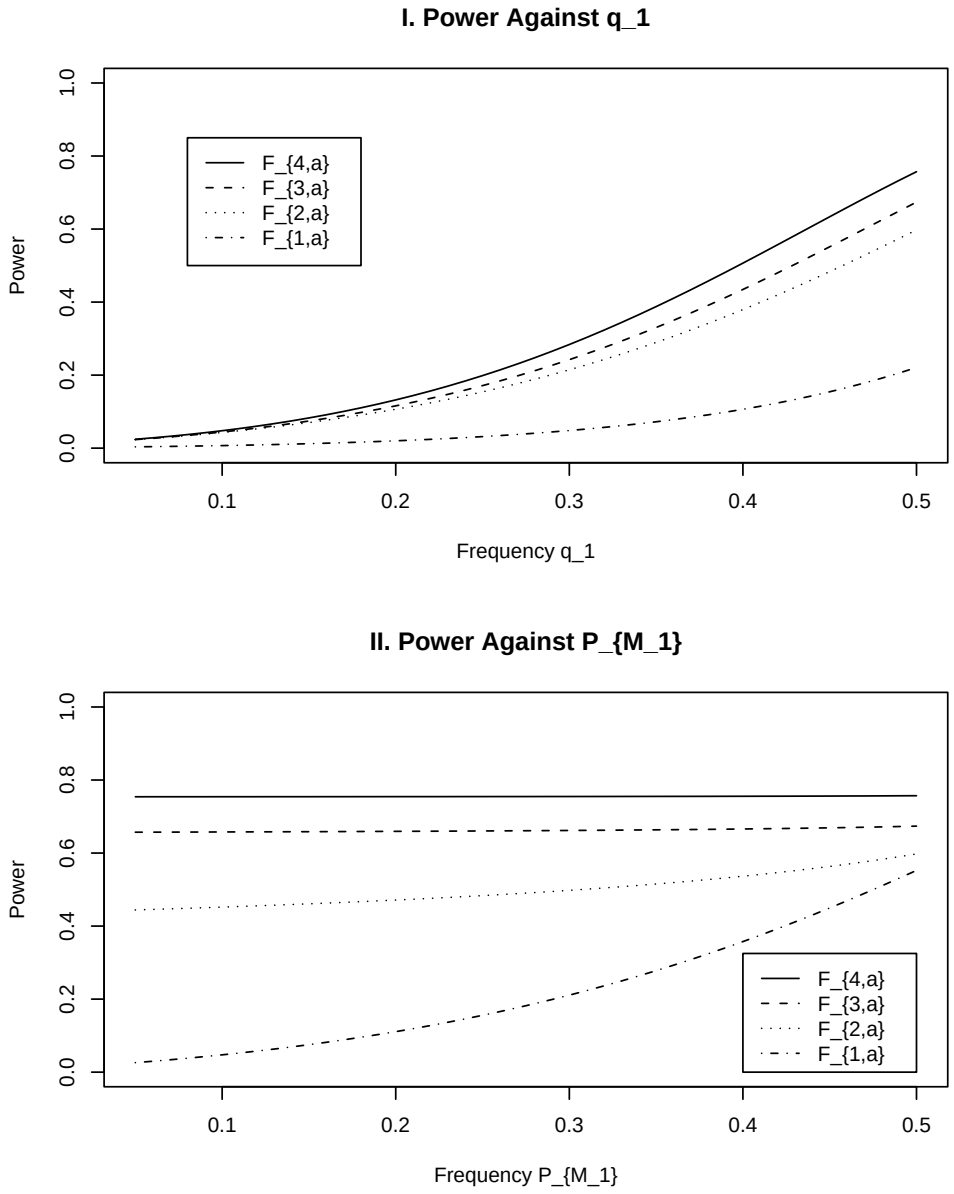

Fig. 15. Power of test statistics $F_{4,a}$, $F_{3,a}$, $F_{2,a}$, and $F_{1,a}$ against the trait allele frequency q_1 (Graph I) or marker allele frequency P_{M_1} (Graph II) at a 0.01 significant level for an additive mode of inheritance $a = 1.0, d = 0.0$, when $P_{M_1} = 0.5$ or $q_1 = 0.5$, respectively. The other parameters are given by $h^2 = 0.15$, $P_{M_i} = 0.5$, $\pi_{12Q} = 0.5$, $\delta_{12Q} = 0.25$, $\sigma_H^2 = 0.1$, $D_{M_iQ} = [\min(P_{M_i}, q_1) - P_{M_i}q_1]/2$, $D_{M_1M_i} = [\min(P_{M_1}, P_{M_i}) - P_{M_1}P_{M_i}]/2, i = 2, 3, 4$ and $D_{M_iM_j} = 0.05, i, j = 2, 3, 4, i \neq j$ and sample size $n = 40, m = 30, s = 20$.

figure 15, it can be seen that the power of $F_{k,a}$ increases as the trait allele frequency q_1 increases. Graph II of the figure 15 shows that the power of $F_{4,a}$ and $F_{3,a}$ is almost constant; besides, the power of $F_{2,a}$ increases slowly, and the power of $F_{1,a}$ increases as the marker allele frequency P_{M_1} increases. In general, the power of $F_{4,a}$ and $F_{3,a}$ heavily depends on the trait allele frequency q_1, but not on the marker allele frequency P_{M_1}.

Assume that the LD is due to historical mutations of T generations ago at QTL Q. At the initial generation when the mutation occurred, the LD coefficient is $D_{M_iQ}(0) = P(M_iQ)(0) - q_1P_{M_i}$, where $P(M_iQ)(0)$ is frequency of haplotype M_iQ. The LD coefficient is reduced by a factor $1 - \theta_{M_iQ}$ in each subsequent generation. The LD between marker M_i and Q is $D_{M_iQ}(T) = D_{M_iQ}(0)(1 - \theta_{M_iQ})^T$ at the current generation. Assume that the marker M_1 locates at position 0cM, marker M_2 locates at position 1cM, marker M_3 locates at position 2cM, and marker M_4 locates at position 3cM. Under the assumption of no interference, we may calculate the recombination fraction $\theta_{M_iM_j} = [1 - \exp(-2\Omega_{M_iM_j})]/2$ by Haldane's map function, where $\Omega_{M_iM_j}$ is map distance between marker M_i and marker M_j. Similarly, the recombination fraction θ_{M_iQ} can be calculated by the distance Ω_{M_iQ} between QTL Q and marker M_i, $i = 1, \cdots, 4$. Suppose that the QTL Q is located along the horizontal axis, i.e., it moves from 0cM to 3cM. Figure 16 shows the power curves of the test statistics $F_{4,a}, F_{4,ad}, F_{3,a}, F_{3,ad}, F_{2,a}$, and $F_{2,ad}$ against the location of QTL Q for a dominant mode of inheritance ($a = d = 1$) and a recessive mode of inheritance ($a = 1.0, d = -0.5$), respectively. The powers of $F_{4,a}$ and $F_{4,ad}$ with 4 markers in the model are generally high across the location of QTL Q, since at least one marker is close to the QTL Q. The power of $F_{3,a}$ and $F_{3,ad}$ using 3 markers in the model is similar to that of 4 markers, except that QTL Q locates far above from marker M_3, i.e., $\lambda_{M_1Q} \geq 2.3cM$. The power of $F_{2,a}$ and $F_{2,ad}$ using two markers in the model is

high when the QTL is close to markers M_1 and M_2. However, once the QTL is far above from marker M_2 (i.e., $\lambda_{M_1Q} \geq 1.3cM$), the power of $F_{2,a}$ and $F_{2,ad}$ using two markers in the model decreases very quickly. Figure 16 implies that multiple marker LD analysis has high power in fine mapping of QTL. Moreover, the power of test statistics $F_{k,a}$ which only tests additive effect is higher than that of $F_{k,ad}$ which tests both additive and dominant effect through the proposed model. The reason is the number of degrees of freedom of test statistics increases if dominant effect is added to the test statistics. Figure 17 shows the power curves of test statistic $F_{4,ad}$ against position of markers $M_1, \cdots, M_4$ for different mutation age at a 0.01 significant level. The trait locus Q locates at position 10cM. The four markers flank the trait locus Q; two markers are on each side of the QTL with equal distance to the each other as follows: $M_2 = 5 + M_1/2, M_3 = 15 - M_1/2, M_4 = 20 - M_1$. Here M_i also denotes the location in cM of marker M_i. As age of mutation is getting old, the power decreases and the power can be high only when the markers are close to the trait locus.

Figure 18 shows that the required number of trio families or families with both parents and 2 offspring for the test statistics $F_{4,a}, F_{3,a}, F_{2,a}$ and $F_{1,a}$ against heritability h^2 at a significant level 0.01 and power 0.8. For a favorable case (Graphs I and III), the parameters are given by $q_1 = P_{M_i} = 0.5, D_{M_iM_j} = 0.05$ and $D_{M_iQ} = 0.1$ for $i, j = 1, \cdots, 4, i \neq j$. For a less favorable case (Graphs II and IV), the parameters are given by $q_1 = 0.2, P_{M_i} = 0.8$, $D_{M_iM_j} = 0.0$ and $D_{M_iQ} = 0.03$ for $i, j = 1, \cdots, 4, i \neq j$. For the favorable case, the required number of families of test statistics $F_{4,a}$ and $F_{3,a}$ is less than 200 and that of $F_{2,a}$ is less than 600 if heritability h^2 is larger than 0.1. For the less favorable case, the required number of families of test statistics $F_{4,a}$ and $F_{3,a}$ is less than 500 and that of $F_{2,a}$ is less than 700 if heritability h^2 is larger than 0.1. The required number of families of test statistics $F_{1,a}$ is very large for both favorable and less favorable cases.

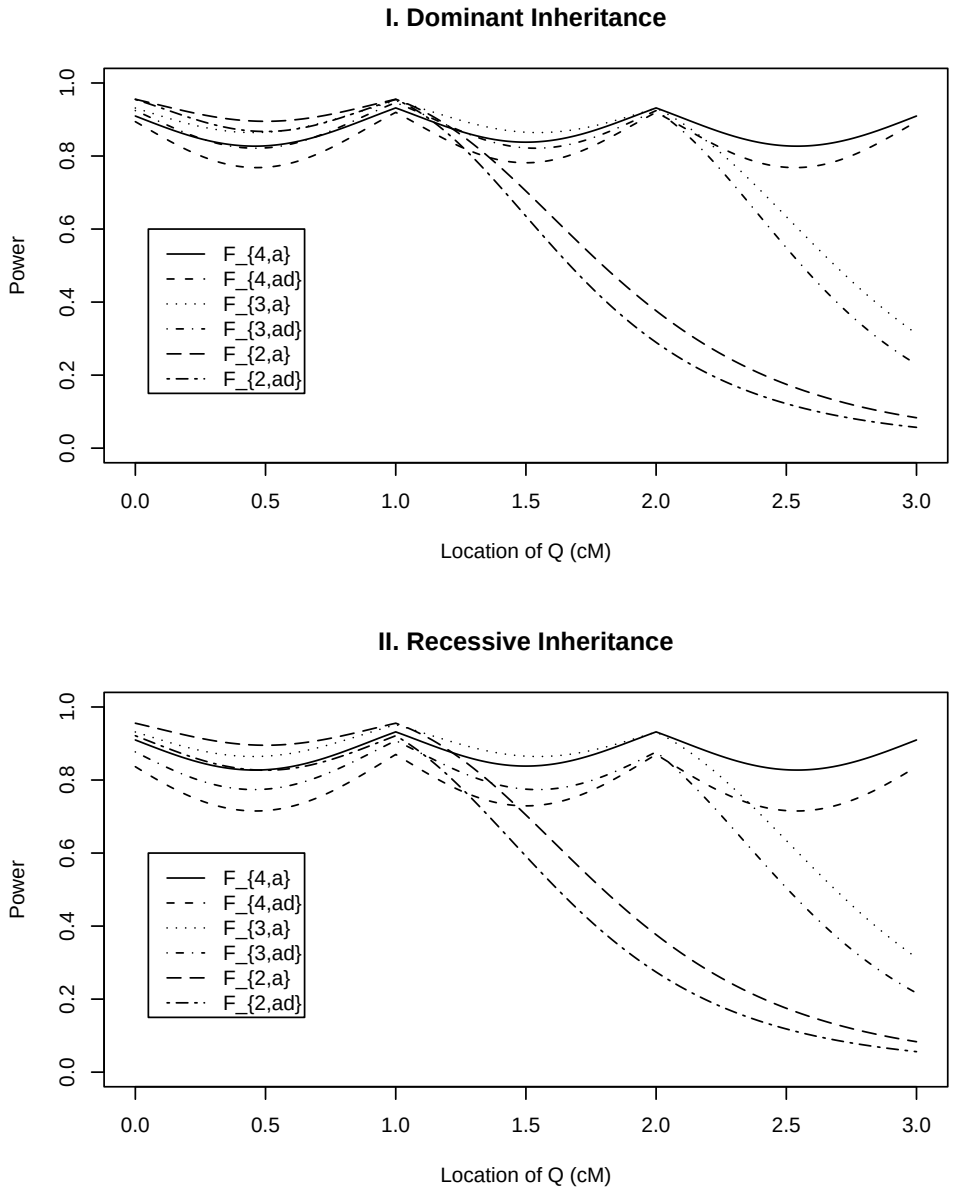

Fig. 16. Power of test statistics $F_{4,a}$, $F_{4,ad}$, $F_{3,a}$, $F_{3,ad}$, $F_{2,a}$, and $F_{2,ad}$ against location of QTL Q at a 0.01 significant level. The parameters are given by $q_1 = 0.5, P_{M_i} = 0.5, D_{M_iQ}(0) = 0.15$, $D_{M_iM_j} = 0.05, i, j = 1, \cdots, 4, i \neq j$, $\pi_{12Q} = 0.5, \delta_{12Q} = 0.25$, familial effect variance $\sigma_H^2 = 0.10$, heritability $h^2 = 0.15$, and sample size $n = 100, m = 50, s = 30$, mutation age $T = 60$ for a dominant mode of inheritance $a = d = 1.0$ (Graph I), and a recessive mode of inheritance $a = 1.0, d = -0.5$ (Graph II), respectively. Marker M_1 locates at position 0cM, marker M_2 locates at position 1cM, marker M_3 locates at position 2cM, and marker M_4 locates at position 3cM. The location of QTL Q is along the horizontal axis, i.e., it moves from 0cM to 3cM.

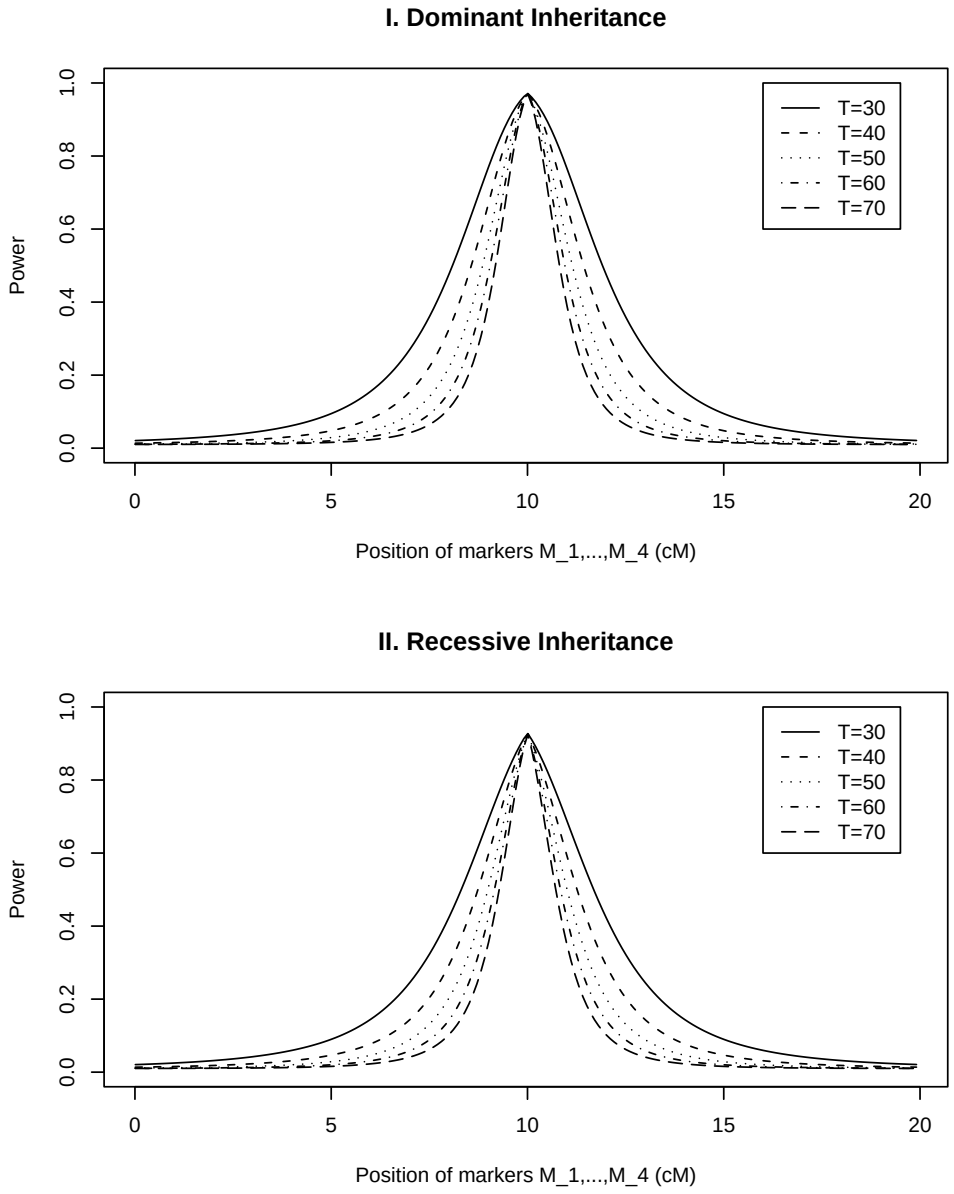

Fig. 17. Power of test statistic $F_{4,ad}$ for mutation age $T = 30$, $T = 40$, $T = 50$, $T = 60$, $T = 70$ against position of markers $M_i, i = 1, \cdots, 4$ at a 0.01 significant level. The QTL Q locates at position 10cM. The four markers flank the trait locus Q; two markers are on each side of the QTL with equal distance to the each other as follows: $M_2 = 5 + M_1/2, M_3 = 15 - M_1/2, M_4 = 20 - M_1$. $q_1 = 0.5, P_{M_i} = 0.5, D_{M_iQ}(0) = 0.15$, $D_{M_iM_j} = 0.05, i, j = 1, \cdots, 4, i \neq j$, heritability $h^2 = 0.15$, familial effect variance $\sigma_H^2 = 0.1$, and sample size $n = 40, m = 30, s = 20$ for a dominant mode of inheritance $a = d = 1.0$ (Graph I), and a recessive mode of inheritance $a = 1.0, d = -0.5$ (Graph II), respectively.

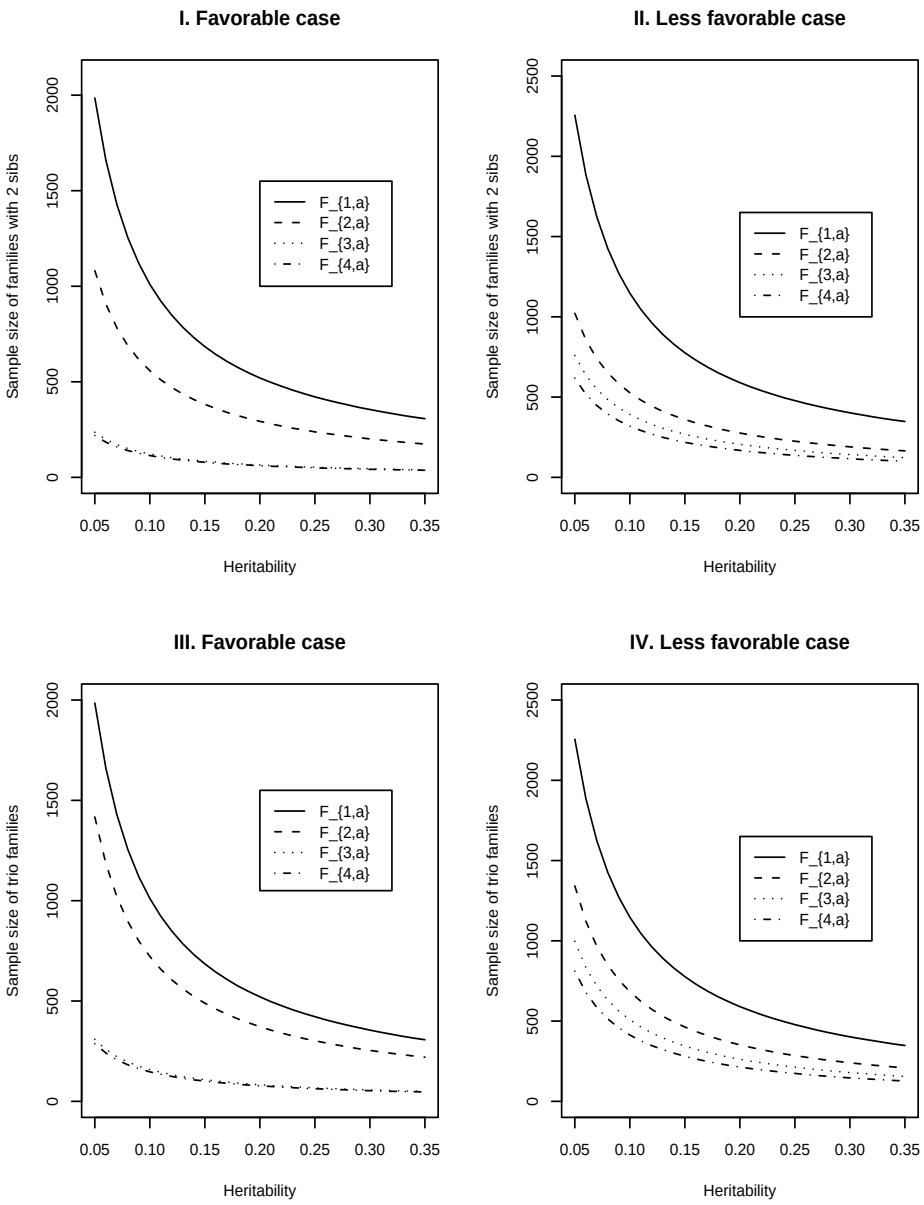

Fig. 18. Sample size of test statistics $F_{1,a}$, $F_{2,a}$, $F_{3,a}$, and $F_{4,a}$ against heritability h^2 at a 0.01 significant level and 0.80 power for a dominant mode of inheritance $a = d = 1.0$. For favorable case (Graph I and Graph III), $q_1 = 0.5, P_{M_i} = 0.5, D_{M_iM_j} = 0.05, D_{M_iQ} = 0.1, i, j = 1, 2, 3, 4, i \neq j$; for less favorable case (Graph II and Graph IV), $q_1 = 0.2, P_{M_i} = 0.8, D_{M_iM_j} = 0.0, D_{M_iQ} = 0.03, i, j = 1, 2, 3, 4, i \neq j$. In addition, the familial effect variance $\sigma_H^2 = 0.1$.

4.7. Application

The proposed method is applied to the Genetic Analysis Workshop 12 German asthma data (Meyers, Wjst and Ober 2001). The data consist of 97 nuclear families, including 415 persons. Seventy-four families have 2 children, 19 have three children, and 4 have four children. Wjst et al. (1999) perform linkage analysis for total serum IgE by nonparametric statistic of MAPMAKER/SIBS 2.1. Three markers on chromosome 1 are shown to be linked with IGE level, i.e., marker D1S207 at position 118.1cM, marker D1S221 at position 146.7cM and marker D1S502 at position 151.2cM. In Fan and Jung (2003), we analyze the data using sib-ships, and confirm the result of Wjst et al. (1999). By the method proposed in this paper, we analyze the data again. The dominant variance of log(IGE) is significantly higher than 0 at position 149.85cM (p-value, 0.00075; compared with the p-value 0.01 in Fan and Jung 2003). On this basis, we collapse alleles 6, 8 and 10 as allele M_1 at marker D1S207, and others as allele m_1. At marker D1S221, alleles 5, 6 and 7 are collapsed as allele M_2, and other alleles as allele m_2. At marker D1S502, we collapse alleles 7, 8, and 12 as allele M_3, and others as allele m_3. Then, we find that coefficient δ_2 is significantly different from 0 at position 149.85cM, with a p-value 0.034 by likelihood ratio test (compared with the p-value 0.0475 in Fan and Jung 2003) and a p-value 0.034 by F test (compared with the p-value 0.0484 in Fan and Jung 2003). The estimation is $\hat{\delta}_2 = 0.76$. Hence, we are able to confirm the result of Wjst et al. (1999), and find that marker D1S221 is associated with log(IGE).

Compared with the results in the previous chapter, the evidence in the above paragraph is stronger since the p-values are smaller. There are two reasons for this. In the method of this chapter, all family members are used with three markers in analysis, while sibships are analyzed with only two markers in the previous chapter.

Hence, the proposed model improves the performance of the methods in the chapter III.

4.8. Discussion

Based on multiple bi-allelic markers, variance component models are proposed for high resolution linkage disequilibrium mapping of QTL in the presence of prior linkage evidence. The models are extended by method using two bi-allele markers in analysis, and incorporate genetic-marker information into the models (Fan and Jung 2003; Fan and Xiong 2002, 2003). With analytical derivation, it is shown that linkage disequilibrium measures and genetic effects are incorporated in the mean coefficients. Using the information of sharing IBD of multiple markers, a multi-point interval mapping method is provided to estimate the proportion of allele sharing IBD and probability of sharing 2 allele IBD at a putative QTL for a sib-pair. It is shown that recombination fractions, i.e., linkage information, are contained in variance covariance matrices. Therefore, the proposed methods model both association and linkage in a unified model.

After comparing with the "AbAw" approach, it is found that the method proposed in this chapter is more powerful and advantageous in terms of simulation study and power calculation. By power and sample size comparison, it is shown that models which use more markers may have higher power than models which use less markers. The multiple marker analysis can be more advantageous, and has high power in fine mapping QTL.

Type I error calculations are performed in this chapter. We allow for the very extreme form of population admixture, in which each family is drawn from a different stratum (Abecasis, Cardon, and Cookson 2000). Type I error rates of the proposed

test statistics are calculated to investigate the behaviors of the test statistics under the null distribution. Five test cases including population admixture are considered to investigate the type I error rates, which leads to reasonable result. The likelihood ratio tests are less likely to be influenced by population admixture.

In a QTL mapping study, a strategy may be taken as follows. First, linkage analysis can be carried out using a sparse genetic map. Then, association study can be performed using a dense genetic map for high resolution mapping. The basic idea is to take the advantage of linkage analysis for a prior linkage information. In the meantime, the advantage in high resolution of association study can be taken for fine mapping a genetic trait. It is well known that linkage analysis is robust, i.e., the false positive rates are not high. However, the resolution of linkage analysis can be low. On the other hand, the resolution of association study is high. But, association study is prone to false positives caused by population stratifications. Using the method proposed in this chapter, it is more likely to avoid high false positive rates by performing association study in the presence of prior linkage. The low resolution of a prior linkage analysis can be remedied by the follow-up high resolution association study.

So far, only one trait locus Q is assumed to be located in the chromosome region. Suppose that there are multiple QTL in the region. The regression equation (4.2) can still be used in QTL mapping. Besides, suppose that the trait value is influenced by unlinked trait loci in different regions. Then model (4.1) needs to be generalized to use markers from different regions in analysis (Hoh and Ott 2003). If multiple trait loci are present, other issues such as epistasis need more in depth investigation. For IBD estimation, we follow the method proposed by Fulker et al. (1995) and Alamsy and Blangero (1998). If there is LD between the trait and markers, LD among markers would also be expected, and needs to be incorporated in estimating proportion of

sharing IBD. However, it is not clear how to achieve this. This is a very interesting and important research area for future study. Better estimates of the proportion of allele sharing IBD would lead to a fitted variance covariance structure which is a better approximation of the true variance covariance structure. This would improve the performance of the proposed models.

CHAPTER V

CONCLUSION

5.1. Summary and Discussion

In a QTL mapping study, one may carry out both linkage analysis and association study. Linkage analysis is based on family data, and is useful in localizing a genetic trait locus in a broad chromosome region. Therefore, linkage analysis can provide suggestive linkage between a putative trait locus and a marker locus based on a sparse marker map. In addition, linkage analysis is robust to the population stratification which heavily affects the results of population-based association study. Association study, on the other hand, is useful in fine gene mapping of genetic trait locus since the allelic association due to LD usually operates over very short genetic distance. Hence, association study can provide high resolution in genetic trait mapping. However, association study is prone to false positive caused by population stratifications. As we develop methods proposed in chapters III and IV, it is more likely to avoid high false positive rates by performing association study in the presence of prior linkage. The low resolution of a prior linkage analysis can be remedied by the follow-up high resolution association study.

In the recent years, there has been great interest in association study of quantitative trait loci (QTL). Allison (1997) proposed various Transmission Disequilibrium (TD)-type tests which accommodate either selected sampling or sampling based on selection of extreme phenotypes among the offspring. George et al. (1999) proposed a TDT in pedigree data by multiple regression. Zhang and Zhao (2001) propose a quantitative similarity-based test to identify association between a bi-allelic marker and a quantitative. Using a bi-allelic marker, Fan and Xiong (2003) proposed mixed

models to perform both linkage analysis in the presence of association and association study in the presence of linkage. For multiple allele marker, only association study in the presence of linkage is conducted by mixed model in the chapter II because the way to reduce the number of parameter is not clear. The association study shows that the method employing a multiple allele has higher power than that using a bi-alleles marker if the marker allele frequencies are evenly distributed.

"AbAw" approach, a combined linkage and association mapping, is developed to decompose association effect into within and between family components (Abecasis et al. 2000, 2001; Cardon 2000; Fulker et al. 1999; Sham et al. 2000). Xiong and Jin (2000) proposed a maximum likelihood based linkage and linkage disequilibrium analysis for genome-wide screens that can be applied to general pedigrees. Wu et al. (2002) made use of mixture models in joint linkage and LD mapping. However, most research limits on using one bi-allelic marker at a time to model the combined study. The methods presented in chapters III and IV propose to use multiple markers in order to model the association and linkage together. Both chapters show that models which use more markers may have higher power than models which use less markers. The multiple marker analysis can be more advantageous, and has high power and better effect in fine mapping QTL.

In association study, population stratification can lead to high false positives (Ewens and Spielman, 1995). Zhao and Xiong (2002) presented unbiased quantitative population association tests to investigate the issue. In the chapter IV, we calculate type I error rate of the proposed test statistics to investigate the behavior of test statistics under the null hypothesis. Then we compare the results with those of "AbAw" in Abecasis et al. (2000) and find that the method proposed in chapter IV is more likely to avoid high false positive rates.

5.2. Open Problems

5.2.1. Association Study by Mixed Model

In chapter II, we assume that all members of nuclear family are available. With the information of transmitted and non-transmitted alleles from parents, the mixed model is built in order to study association. But there are some situations which parental information is not available with several reasons such as late onset diseases and financial problems. It would bring an interest if the methods proposed in chapter II can be extended to study the data without parental data.

The mixed models in the chapter II do not take interactions into account. There may exist an interaction between genetic effects and environments in the certain situation. Van den Oord and Sneider (2002) proposed a general model to study an interaction of the multiple etiological factors and other genetic effects such as age dependency. It would be interesting if the proposed model can be extended to consider the interaction between genetic effects and environment effects.

5.2.2. Association Study by Variance Component Model

Genotyping information is usually given in a genetics study. The methods developed in chapters III and IV can be directly used in analyzing quantitative trait and genotyping data of nuclear families by combining linkage and association information together. One may insist on using haplotype data to map QTL which can be constructed based on genotyping data. We may be interested in comparing our approach with an approach of haplotype data.

The potential problem of the method using multiple markers in chapters III and IV is that degrees of freedom of test statistics can be large as we add the number of markers, and the large numbers of degree of freedom may cause power to decrease.

Moreover, the number of LD measures can be large. The selection of appropriate markers for analysis is one of important problems to be carefully considered. The optimal number of markers needed depends on not only specific trait in a study, but also the LD measures among the QTL and the markers. It would not be a good idea to use many bi-allelic markers in the model. More markers will lead to higher degrees of freedom which cause lower power. Usually, using three or four relevant markers in analysis would be worthwhile, since it may not only have higher power than one or two marker analysis, but also have lower degrees of freedom and number of LD measures than more than four markers.

The other problem is the existence of dominant trait effect. If the dominant effect is present, one may lose power by excluding it from the models, (Fan and Xiong, 2002). However, one may get low power during simultaneous test of additive and dominant effect, if the dominant effect is not significantly present to influence the trait values, due to the increase of degrees of freedom of test statistics.

Only one trait locus Q is assumed to be considered in order to localize it on a chromosome region until now. Suppose that there are multiple quantitative trait loci (QTL) in the region. The regression equation in chapter IV can still be used in QTL mapping. Besides, suppose that the trait value of interest is influenced by unlinked trait loci in different regions. Then model proposed in chapter IV needs to be generalized to use markers from different regions in analysis (Hoh and Ott 2003). If multiple trait loci are present, other issues such as epistasis are needed to be considered. For estimation of proportion of sharing IBD, we follow the method proposed by Fulker et al. (1995) and Alamsy and Blangero (1998). If there is LD between the trait and markers, LD among markers would also be expected, and needs to be incorporated in estimating IBD. However, it is not clear how to achieve them. This is a very interesting and important research area for future study. Better estimated proportion

of sharing IBD would lead to a fitted variance covariance structure which is a better approximation of the true variance covariance structure. This would improve the performance of the proposed models.

REFERENCES

Abecasis, G., Cardon, L. and Cookson, W. (2000). A general test of association for quantitative traits in nuclear families. *American Journal of Human Genetics* **66**, 279–292.

Abecasis, G., Cookson, W., and Cardon, L. (2000). Pedigree tests of linkage disequilibrium. *European Journal of Human Genetics* **8**, 545–551.

Abecasis, G., Cookson, W., and Cardon, L. (2001). The power to detect linkage disequilibrium with quantitative traits in selected samples. *American Journal of Human Genetics* **68**, 1463–1474.

Allison, D. (1997). Transmission-disequilibrium tests for quantitative traits. *American Journal of Human Genetics* **60**, 676–690.

Almasy, L. and Blangero, J. (1998). Multipoint quantitative trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198–1211.

Almasy, L., Williams, J., Dyer, and T., Blangero, J. (1999). Quantitative trait locus detection using combined linkage/disequilibrium analysis. *Genetic Epidemiology* **17**,(Suppl 1):S31–S36.

Amos, C. (1994). Robust variance-components approach for assessing linkage in pedigrees. *American Journal of Human Genetics* **54**, 534–543.

Amos, C., Elston, R., Wilson, A., and Bailey-Wilson, J.(1989) A more powerful robust sib-pair test of linkage for quantitative traits. *Genetics Epidemiology* **6**, 435–449.

Broman, K., Murray, J., Sheffied, V., White, R. and Weber, J.(1998) Comprehensive human genetic map: Individual and sex-specific variation in recombination. *American Journal of Human Genetics* **63**, 861–869.

Cardon, L.(2000). A sib-pair regression model of linkage disequilibrium for quantitative traits. *Human Heredity* **50**, 350–358.

Cookson, W. and Abecasis, G. (2001). Oxford genome screen for asthma-associated traits. *Genetic Epidemiology* **21**, (Suppl 1):S1–S3.

Cotterman, C. (1974) A calculus for statistico-genetics. In *Genetics and Social Structure:mathematical structuralism in population genetics and social theory.*, P. Ballonoff(ed), 157–272 Stroudsburg, PA:Dowden, Hutchinson & Ross.

Daniel, S., Bhattacharrya, S., James, A., Leaves, N., Young, A. et al. (1996). A genome-wide search for quantitative trait loci underlying asthma. *Nature* **383**, 247–250.

Elston, R. and Keats, B. (1985) Genetic analysis workshop III: Sib pair analyses to determine linkage groups and to order loci. *Genetic Epidemiology* **2**, 211–213.

Ewens, W. and Spielman, R. (1995) The transmission/disequilibrium test: History, subdivision, and admixture. *American Journal of Human Genetics* **57**, 455–464.

Falconer, D., and Mackay, T. (1996). *Introduction to Quantitative Genetics*, 4th edition, Prentice Hall, London and New York: Longman.

Fan, R., Floros, J., and Xiong, M. (2002). Models and tests of linkage and association studies of QTL for multi-allele marker loci. *Human Heredity* **53**, 130–145.

Fan, R. and Jung, J. (2003) High resolution joint linkage disequilibrium and linkage mapping of quantitative trait loci based on sibship data. *Human Heredity*, **56** 166–187.

Fan, R. and Xiong, M. (2002) High resolution mapping of quantitative trait loci by linkage disequilibrium analysis. *European Journal of Human Genetics* **10**, 607–

615.

Fan R and Xiong, M (2003) Combined high resolution linkage and association mapping of quantitative trait loci. *European Journal of Human Genetics* **11**, 125–137.

Fulker, D. and Cardon, L. (1994) A sib-pair approach to interval mapping of quantitative trait loci. *American Journal of Human Genetics* **54**, 1092–1103.

Fulker, D., Cherny, S. and Cardon, L. (1995) Multiple interval mapping of quantitative trait loci, using sib-pairs. *American Journal of Human Genetics* **56**, 1224–1233.

Fulker, D., Cherny, S., Sham, P., and Hewitt, J. (1999). Combined linkage and association sib-pair analysis for quantitative traits. *American Journal of Human Genetics* **64**, 259–267.

George, V., Tiwari, H., Zhu, X., and Elston, R. (1999). A test of transmission/ disequilibrium for quantitative traits in pedigree data, by multiple regression. *American Journal of Human Genetics* **65**, 236–245.

Goldgar, D. (1990) Multipoint analysis of human quantitative genetic variation. *American Journal of Human Genetics* **47**, 957–967.

Goldgar, D. and Oniki, R. (1992) Comparison of a multipoint identity-by-descent method with parametric multipoint linkage analysis for mapping quantitative traits. *American Journal of Human Genetics* **50**, 598–606.

Göring HHH and 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. *American Journal of Human Genetics* **66**, 1310–1327.

Graybill, F. (1976). *Theory and Application of the Linear Model*. Pacific Grove, CA:Duxbury press.

Haley, C. and Knott, S. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hartl, D. and Clark, A. (1989) *Principles of Population Genetics*, Second edition. Sunderland, MA: Sinauer Associates.

Harville, D. (1997) *Matrix Algebra from a Statistician's Perspective*. NY:Springer.

Haseman, J. (1970) *The Genetic Analysis of Quantitative Traits Using Twin and Sib Data.* Ph.D. dissertation, Chapel Hill:University of North Carolina.

Haseman, J. and Elston, R. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* **2**, 3–19.

Hedrick, P. (1987) Gametic disequilibrium measures: Proceed with caution. *Genetics* **117**, 331–341.

Hoh,J. and Ott, J. (2003) Mathematical multi-locus approaches to localizing complex human trait genes. *Nature Reviews: Genetics* **4**, 701–709.

Horvath, S. and Laird, N. (1998) A discordant-sibship test for disequilibrium and linkage: No need for parental data. *American Journal of Human Genetics* **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, R. (1993) Interval mapping of multiple quantitative trait loci. *Genetics* **135**, 205–211.

Kong, A., Gudbjartsson, D., Sainz, J., Jonsdottir, G., Gudjonsson, S. et al. (2002) A high resolution recombination map of the human genome. *Nature Genetics* **31**, 241–247.

Lander, E. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lange, K. (2002) *Mathematical and Statistical Methods for Genetic Analysis.* NY:Springer.

Lewontin, R. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 67.

Martin, E., Monks, S., Warren, L., and Kaplan, N. (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *American Journal of Human Genetics* **67** 146–154.

Meyers, D., Wjst, M. and 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. *Genetic Epidemiology* **21** (Suppl 1):S4-S8.

Miller, J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistics* **5**, 746–762.

Monks, S. and Kaplan, N. (2000) Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. *American Journal of Human Genetics* **66**, 576–592.

Pinheiro, J. (1994). *Topics in Mixed-effects Models.* Ph.D. dissertation, Madison:University of Wisconsin-Madison.

Pinheiro, J. and Bates, D. (2000). Mixed-effects models in S and S-plus. NY:Spinger

Verlag.

Pratt, S., Daly, M., and Kruglyak, (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *American Journal of Human Genetics* **66**, 1153–1157.

Rabinowitz, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Human Heredity* **47**, 342–350.

Schaid, D. and Li, H. (1997) Genotype relative-risks and association tests for nuclear families with missing parents. *Genetic Epidemiology* **14**, 1113–1118.

Schaid, D. and Rowland, C. (1998) Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. *American Journal of Human Genetics* **63**, 1492–1506.

Schork, N. (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: Efficiency, power, and modeling considerations. *American Journal of Human Genetics* **53**, 1306–1319.

Self, S. and Liang, K. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of American Statistics Association* **82**, 605–610.

Sham, P., Cherny, S., Purcell, S. and Hewitt, J. (2000). Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *American Journal of Human Genetics* **66**, 1616–1630.

Sham, P. and Curtis, D. (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Annals Human Genetics* **59**, 323–336.

Spielman, R. and Ewens, W. (1996). The TDT and other family-based tests for

linkage disequilibrium and association. *American Journal of Human Genetics* **59**, 983–989.

Spielman, R. and Ewens, W. (1998) A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test. *American Journal of Human Genetics* **62**, 450–458.

Spielman, R., McGinnis, R. and Ewens, W. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* **52**, 506–516.

Stuart, A. and Ord, J. (1991). *Kendall's Advanced Theory of Statistics*, Vol. 2: Classical Inference and Relationships, 5th edition. Oxford:Oxford University.

Van den Oord, E. and Sneider, H. (2002). Including measured genotypes in statistical models to study the interplay of multiple factors affecting complex traits. *Behavior Genetics* **32**, 1–22.

Weiss, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases I. *Journal of American Statistical Association* **66**, 345–350.

Weiss, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases II. *Journal of American Statistical Association* **68**, 428–430.

Wjst, M., Fischer, G., Immervoll, T., Jung, M., Saar, K. et al. (1999) A genome-wide search for linkage to asthma. *Genomics* **58**, 1–8.

Wu, R., Ma, C., Casella, G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* **160**, 779–792.

Xiong, M., Jin, L. (2000) Combined linkage and linkage disequilibrium mapping for genome screens. *Genetic Epidemiology* **19**, 211–234.

Xiong, M., Krushkal, J. and Boerwinkle, E. (1998). TDT statistics for mapping quantitative loci. *Annals of Human Genetics* **62**, 431–452.

Xu, S. and Atchley, W. (1995) A random model approach to interval mapping of quantitative trait loci. *Genetics* **141**, 1189–1197.

Zhang, S. and Zhao, H. (2001) Quantitative similarity-based association tests using population samples. *American Journal of Human Genetics* **69**, 601–614.

Zhao, J., Li, W., and Xiong, M. (2001) Population based linkage disequilibrium mapping of QTL: An application to simulated data in an isolated population. *Genetic Epidemiology* **21**, (S1):S655-659.

Zhao, J., and Xiong, M. (2002) Unbiased quantitative population association test. *American Journal of Human Genetics* **71**, (Supplement):568, Poster 2336.

Zhu, X. and Elston, R. (2000) Power comparison of regression methods to test quantitative traits for association and linkage. *Genetic Epidemiology* **18**, 322–330.

Zhu, X. and Elston, R. (2001) Transmission/disequilibrium tests for quantitative traits. *Genetic Epidemiology* **20**, 57–74.

APPENDIX A

Without loss of generality, assume that $k = 2$ and $n = 3$ in Figure 1. Let TM_1 be the abbreviation of the "transmitted marker allele for child 1", and NM_1 be the abbreviation of the "non-transmitted marker allele for child 1", from the heterozygous mother M_iM_j in Figure 1. Similarly, we define the notations TM_i, NM_i, $i = 2, 3$. Denote $A = (TM_1 = M_i, NM_1 = M_j, TM_2 = M_i, NM_2 = M_j)$. Let S_{7kl} be the state where two offspring share two identical trait alleles Q_k and Q_l by descent, and Q_l is from the heterozygous father and Q_k is from the mother; S_{8klr} be the state where two offspring share one identical trait allele Q_k by descent, and the other two alleles Q_l and Q_r are not identical by descent; and S_{9krls} be the state where two offspring share no identical trait alleles by descent, and two alleles Q_l, Q_s are from the heterozygous father, and the other two alleles Q_k, Q_r are from the mother. Then

$$\begin{aligned}\Sigma_{ij,ij} &= \Big[\sum_k \sum_l \mu_{kl}^2 P(A \cap S_{7kl}) + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} P(A \cap S_{8klr}) \\ &\quad + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} P(A \cap S_{9krls})\Big]/(p_i p_j/2) - (\nu - \alpha_{i,j})^2 + \sigma_G^2/2,\end{aligned}$$

where

$$\begin{aligned}P(A \cap S_{7kl}) &= \frac{q_k}{2}\Big(2h_{li}p_j\frac{1-\theta}{2}\frac{1-\theta}{2} + 2h_{lj}p_i\frac{\theta}{2}\frac{\theta}{2}\Big) = q_k\Big(h_{li}p_j(1-\theta)^2 + h_{lj}p_i\theta^2\Big)/4 \\ P(A \cap S_{8klr}) &= \frac{q_l q_r}{2}\Big(2h_{ki}p_j\frac{1-\theta}{2}\frac{1-\theta}{2} + 2h_{kj}p_i\frac{\theta}{2}\frac{\theta}{2}\Big) \\ &\quad + \frac{q_k}{2}(h_{li}h_{rj} + h_{ri}h_{lj})2\theta(1-\theta)/4 \\ &= q_l q_r\Big(h_{ki}p_j(1-\theta)^2 + h_{kj}p_i\theta^2\Big)/4 + q_k(h_{li}h_{rj} + h_{ri}h_{lj})\theta(1-\theta)/4 \\ P(A \cap S_{9krls}) &= \frac{q_k q_r}{2}(h_{li}h_{sj} + h_{si}h_{lj})2\theta(1-\theta)/4 = q_k q_r(h_{li}h_{sj} + h_{si}h_{lj})\theta(1-\theta)/4.\end{aligned}$$

Similarly, denote $B = (TM_1 = M_i, NM_1 = M_j, TM_3 = M_j, NM_3 = M_i)$. We can calculate the conditional covariance of offspring 1 and 3 in Figure 1

$$\begin{aligned}
\Sigma_{ij,ji} &= \Sigma_{ji,ij} = \mathrm{Cov}(y_1, y_3) \\
&= \Big[\sum_k \sum_l \mu_{kl}^2 P(B \cap S_{7kl}) + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} P(B \cap S_{8klr}) \\
&\quad + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} P(B \cap S_{9krls})\Big]/(p_i p_j/2) \\
&\quad -(\nu - \alpha_{i,j})(\nu - \alpha_{j,i}) + \sigma_G^2/2,
\end{aligned}$$

where

$$\begin{aligned}
P(B \cap S_{7kl}) &= \frac{q_k}{2}\Big(2h_{li}p_j + 2h_{lj}p_i\Big)\theta(1-\theta)/4 = q_k\Big(h_{li}p_j + h_{lj}p_i\Big)\theta(1-\theta)/4 \\
P(B \cap S_{8klr}) &= \frac{q_l q_r}{2}\Big(2h_{ki}p_j + 2h_{kj}p_i\Big)\theta(1-\theta)/4 \\
&\quad + \frac{q_k}{2}(h_{li}h_{rj} + h_{ri}h_{lj})\frac{\theta^2 + (1-\theta)^2}{4} \\
&= q_l q_r\Big(h_{ki}p_j + h_{kj}p_i\Big)\theta(1-\theta)/4 \\
&\quad + q_k(h_{li}h_{rj} + h_{ri}h_{lj})\frac{\theta^2 + (1-\theta)^2}{8} \\
P(B \cap S_{9krls}) &= \frac{q_k q_r}{2}(h_{li}h_{sj} + h_{si}h_{lj})\frac{\theta^2 + (1-\theta)^2}{4} \\
&= q_k q_r(h_{li}h_{sj} + h_{si}h_{lj})\frac{\theta^2 + (1-\theta)^2}{8}.
\end{aligned}$$

APPENDIX B

Assume that the marker locus and the trait locus are in linkage equilibrium, i.e., $h_{ri} = q_r p_i$ for all r, i. Then we have

$$\begin{aligned}
\alpha_{i,j} &= \sum_{r=1}^{2}(\nu + \mu_r)q_r = \nu + \mu = \alpha \\
\sigma^2_{i,j} &= \sigma^2_e + \sigma^2_G + \sum_{r=1}^{2}\sum_{s=1}^{2}(\nu + \mu_{rs} - \alpha)^2 q_r q_s = \sigma^2 \\
\Sigma_{ij,ij} &= \sum_k \sum_l \mu^2_{kl} q_k q_l [(1-\theta)^2 + \theta^2]/2 + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} q_k q_l q_r /2 \\
&\quad + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} q_k q_r q_l q_s \theta(1-\theta) - (\nu-\alpha)^2 + \sigma^2_G/2 = \Sigma_{ts} \\
\Sigma_{ij,ji} &= \sum_k \sum_l \mu^2_{kl} q_k q_l (1-\theta)\theta + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} q_k q_l q_r /2 \\
&\quad + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} q_k q_r q_l q_s [\theta^2 + (1-\theta)^2]/2 - (\nu-\alpha)^2 + \sigma^2_G/2 = \Sigma_{td}.
\end{aligned}$$

Notice that $\alpha, \sigma^2, \Sigma_{ts}$ and Σ_{td} do not depend on subscripts i and j.

APPENDIX C

Assume that the recombination fraction $\theta \approx 0$, i.e. there is tight linkage between the trait locus and the marker. Then $P(Q_r M_i, M_j) \approx h_{ri} p_j$. Therefore, we have

$$\begin{aligned}
\alpha_{i,j} &\approx \sum_{r=1}^{2} (\nu + \mu_r) h_{ri}/p_i = \alpha_i \\
\sigma^2_{i,j} &\approx \sigma^2_e + \sigma^2_G + \sum_{r=1}^{2}\sum_{s=1}^{2} (\nu + \mu_{rs} - \alpha_i)^2 q_s h_{ri}/p_i = \sigma^2_e + \sigma^2_G + \Sigma^2_i = \sigma^2_i.
\end{aligned}$$

Note that α_i and Σ^2_i only depend on subscript i. Besides, the covariances $\Sigma_{ij,ij}$ and $\Sigma_{ij,ji}$ can be approximated by

$$\begin{aligned}
\Sigma_{ij,ij} &\approx \Big[\sum_k \sum_l \mu^2_{kl} q_k h_{li} + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} q_l q_r h_{ki}\Big]/(2p_i) - (\nu - \alpha_i)^2 + \sigma^2_G/2 = \Sigma_{i,i} \\
\Sigma_{ij,ji} &\approx \Big[\sum_k \sum_l \sum_r \mu_{kl}\mu_{kr} q_k (h_{li} h_{rj} + h_{ri} h_{lj}) \\
&\quad + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} q_k q_r (h_{li} h_{sj} + h_{si} h_{lj})\Big]/(4p_i p_j) \\
&\quad -(\nu - \alpha_i)(\nu - \alpha_j) + \sigma^2_G/2 = \Sigma_{i,j} = \Sigma_{j,i}.
\end{aligned}$$

Notice that $\Sigma_{i,i}$ only depends on subscript i, but $\Sigma_{i,j} = \Sigma_{j,i}$ depends on both i and j.

APPENDIX D

Let TH denote abbreviation of "transmitted haplotype". Then $P(TH = Q_r M_i)$ $= (1-\theta)h_{ri} + \theta q_r p_i$. Notice that $h_{2i} - q_2 p_i = -h_{1i} + q_1 p_i = -\delta_i$. Like Appendix A of Fan, Floros and Xiong (2002), one may show that

$$\begin{aligned}
\beta_i &= \mathrm{E}\,[Y|TM = M_i] \\
&= \Big[\mathrm{E}\,[Y|TH = Q_1 M_i]P(TH = Q_1 M_i) + \mathrm{E}\,[Y|TH = Q_2 M_i]P(TH = Q_2 M_i)\Big]/p_i \\
&= (1-\theta)\Big[(\nu+\mu_1)h_{1i} + (\nu+\mu_2)h_{2i}\Big]/p_i + \theta\alpha
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\beta_i - \alpha}{1-\theta} &= \Big[(\nu+\mu_1)h_{1i} + (\nu+\mu_2)h_{2i}\Big]/p_i - [(\nu+\mu_1)q_1 + (\nu+\mu_2)q_2] \\
&= (\mu_1 - \mu_2)\delta_i/p_i.
\end{aligned}$$

To calculate the conditional variance, we first notice the conditional variances

$$\sigma^2_{Qk} = \mathrm{Var}(Y|TQ = Q_k) = \sigma^2_e + \sigma^2_G + (\mu_{k1} - \mu_k)^2 q_1 + (\mu_{k2} - \mu_k)^2 q_2, k = 1, 2.$$

The conditional variance

$$\sigma^2_{ir} = \mathrm{Var}(Y|TM = M_i) = \sum_{k=1}^{2}[\sigma^2_{Qk} + (\nu + \mu_k - \beta_i)^2]P(TH = Q_k M_i)/p_i.$$

For two different alleles M_i and M_j, $i \neq j$, the conditional covariance

$$\Sigma_{i,jr} = \mathrm{Cov}(Y_1, Y_2|TM_1 = M_i, TM_2 = M_j) = \Sigma_{ij,ji}.$$

Let $C_i = (TM_1 = M_i, TM_2 = M_i)$.

The probability of C_i is $P(C_i) = \sum_{j\neq i} 2p_ip_j\frac{1}{2}\frac{1}{2} + p_i^2 \cdot 1 \cdot 1 = p_i(1+p_i)/2$. Let S_{7kl}, S_{8klr} and S_{9krls} be similar notations as those in Appendix A. Then

$$\begin{aligned}
\Sigma_{i,ir} &= \mathrm{Cov}(Y_1, Y_2 | TM_1 = M_i, TM_2 = M_i) \\
&= \Big[\sum_k\sum_l \mu_{kl}^2 P(C_i \cap S_{7kl}) + \sum_k\sum_l\sum_r \mu_{kl}\mu_{kr} P(C_i \cap S_{8klr}) \\
&\quad + \sum_k\sum_l\sum_r\sum_s \mu_{kl}\mu_{rs} P(C_i \cap S_{9krls})\Big]/P(C_i) - (\nu - \beta_i)^2 + \sigma_G^2/2,
\end{aligned}$$

where

$$\begin{aligned}
P(C_i \cap S_{7kl}) &= \frac{q_k}{2}\Big(2h_{li}\frac{1-\theta}{2}\frac{1-\theta}{2} + 2q_lp_i\frac{\theta}{2}\frac{\theta}{2} + 2h_{li}p_i\frac{1-\theta}{2}\frac{\theta}{2}2\Big) \\
&= q_k\Big(h_{li}(1-\theta)^2 + q_lp_i\theta^2 + 2h_{li}p_i\theta(1-\theta)\Big)/4 \\
P(C_i \cap S_{8klr}) &= \frac{q_lq_r}{2}\Big[2h_{ki}\frac{1-\theta}{2}\frac{1-\theta}{2} + 2q_kp_i\frac{\theta}{2}\frac{\theta}{2} + 2h_{ki}p_i\frac{\theta}{2}\frac{1-\theta}{2}2\Big] \\
&\quad + \frac{q_k}{2}\Big[2h_{li}h_{ri}\frac{\theta^2+(1-\theta)^2}{4} + 2h_{ri}q_l\theta(1-\theta)/4 + 2h_{li}q_r\theta(1-\theta)/4\Big] \\
&= q_lq_r\Big[h_{ki}(1-\theta)^2 + q_kp_i\theta^2 + 2h_{ki}p_i\theta(1-\theta)\Big]/4 \\
&\quad + q_k\Big[h_{li}h_{ri}[\theta^2+(1-\theta)^2] + (h_{ri}q_l + h_{li}q_r)\theta(1-\theta)\Big]/4 \\
P(C_i \cap S_{9krls}) &= \frac{q_kq_r}{2}\Big[2h_{li}h_{si}\frac{\theta^2+(1-\theta)^2}{4} + 2h_{li}q_s\theta(1-\theta)/4 + 2h_{si}q_l\theta(1-\theta)/4\Big] \\
&= q_kq_r\Big[h_{li}h_{si}[\theta^2+(1-\theta)^2] + (h_{li}q_s + h_{si}q_l)\theta(1-\theta)\Big]/4.
\end{aligned}$$

Assume that the marker M and the trait locus Q are in linkage equilibrium, i.e, $h_{ri} = q_rp_i$ for $r = 1, 2, i = 1, \cdots, m$. Then $\beta_i = \alpha$, $\sigma_{ir}^2 = \sigma^2$, $\Sigma_{i,jr} = \Sigma_{td}$ and

$$\begin{aligned}
\Sigma_{i,ir} &= \sum_k\sum_l \mu_{kl}^2 q_kq_l\frac{\theta^2+(1-\theta)^2+2p_i\theta(1-\theta)}{2(1+p_i)} + \sum_k\sum_l\sum_r \mu_{kl}\mu_{kr}q_kq_lq_r/2 \\
&\quad + \sum_k\sum_l\sum_r\sum_s \mu_{kl}\mu_{rs}q_kq_lq_rq_s\frac{[\theta^2+(1-\theta)^2]p_i + 2\theta(1-\theta)}{2(1+p_i)} - (\nu-\beta_i)^2 + \sigma_G^2/2.
\end{aligned}$$

Assume that there is tight linkage between the trait locus and the marker, i.e., $\theta \approx 0$.

Then $\beta_i \approx \alpha_i$, $\sigma_{ir}^2 \approx \sigma_i^2$, $\Sigma_{i,jr} \approx \Sigma_{i,j}$ and

$$\begin{aligned}\Sigma_{i,ir} \;\approx\; & \Big[\sum_k \sum_l \mu_{kl}^2 q_k h_{li} + \sum_k \sum_l \sum_r \mu_{kl}\mu_{kr}[q_l q_r h_{ki} + q_k h_{li} h_{ri}] \\ & + \sum_k \sum_l \sum_r \sum_s \mu_{kl}\mu_{rs} q_k q_r h_{li} h_{si}\Big]/[4P(C_i)] - (\nu - \alpha_i)^2 + \sigma_G^2/2.\end{aligned}$$

APPENDIX E

The loglikelihood function of model (2.5) is $l = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{I}\log|\Gamma_i| - \frac{1}{2}\sum_{i=1}^{I}\big(\vec{y}_i - X_i\gamma\big)^{\tau}\Gamma_i^{-1}\big(\vec{y}_i - X_i\gamma\big)$. Assume that the data consist of both singleton families and sib-pair families. Suppose there are k_i singleton offspring who receive allele M_i from their heterozygous parents, k_{ii} $(i = 1, 2, \cdots, m)$ sib pairs in each of them both sibs receive allele M_i from their heterozygous parents, and $k_{ij} = k_{ji}, i \neq j$ sib pairs in each of them one sib receives allele M_i from his/her heterozygous parent and the other receives allele M_j from the same heterozygous parent.

Let us denote $\rho^{\tau} = (\rho_1 = \sigma_1^2, \rho_2 = \sigma_2^2, \cdots, \rho_m = \Sigma_m, \rho_{m+1} = \Sigma_{1,1}, \cdots, \rho_{2m} = \Sigma_{m,m}, \rho_{2m+1} = \Sigma_{1,2}, \cdots, \rho_{3m-1} = \Sigma_{1,m}, \cdots, \rho_{2m+m(m-1)/2} = \Sigma_{m-1,m})$. We may get the following expected second partial derivatives for $i, j, k = 1, \cdots, m, i \neq j, i \neq j, j \neq k$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial\gamma\partial\gamma^{\tau}} &= -X^{\tau}\Gamma^{-1}X, \mathrm{E}\Big(\frac{\partial^2 l}{\partial\gamma\partial\rho^{\tau}}\Big) = 0,\\
\mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_i^2}\Big) &= \mathrm{E}\Big(\frac{\partial^2 l}{\partial(\sigma_i^2)^2}\Big) = -\frac{k_i}{2(\sigma_i^2)^2} - \frac{k_{ii}[(\sigma_i^2)^2 + \Sigma_{i,i}^2]}{[(\sigma_i^2)^2 - \Sigma_{i,i}^2]^2} - \sum_{j\neq i}\frac{k_{ij}(\sigma_j^2)^2}{2[\sigma_i^2\sigma_j^2 - \Sigma_{i,j}^2]^2},\\
\mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_{m+i}^2}\Big) &= \mathrm{E}\Big(\frac{\partial^2 l}{\partial\Sigma_{i,i}^2}\Big) = -\frac{k_{ii}[(\sigma_i^2)^2 + \Sigma_{i,i}^2]}{[(\sigma_1^2)^2 - \Sigma_{i,i}^2]^2}, \mathrm{E}\Big(\frac{\partial^2 l}{\partial\Sigma_{i,j}^2}\Big) = -\frac{k_{ij}(\sigma_i^2\sigma_j^2 + \Sigma_{i,j}^2)}{(\sigma_i^2\sigma_j^2 - \Sigma_{i,j}^2)^2},\\
\mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_i\partial\rho_j}\Big) &= \mathrm{E}\Big(\frac{\partial^2 l}{\partial\sigma_i^2\partial\sigma_j^2}\Big) = -\frac{k_{ij}\Sigma_{i,j}^2}{2(\sigma_i^2\sigma_j^2 - \Sigma_{i,j}^2)^2},\\
\mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_i\partial\Sigma_{i,i}}\Big) &= \mathrm{E}\Big(\frac{\partial^2 l}{\partial\sigma_i^2\partial\Sigma_{i,i}}\Big) = \frac{2k_{ii}\sigma_i^2\Sigma_{i,i}}{[(\sigma_i^2)^2 - \Sigma_{i,i}^2]^2}, \mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_i\partial\Sigma_{j,j}}\Big) = \mathrm{E}\Big(\frac{\partial^2 l}{\partial\sigma_i^2\partial\Sigma_{j,j}}\Big) = 0,\\
\mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_i\partial\Sigma_{i,j}}\Big) &= \mathrm{E}\Big(\frac{\partial^2 l}{\partial\sigma_i^2\partial\Sigma_{i,j}}\Big) = \frac{k_{ij}\sigma_j^2\Sigma_{i,j}}{(\sigma_i^2\sigma_j^2 - \Sigma_{i,j}^2)^2}, \mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho_i\partial\Sigma_{j,k}}\Big) = \mathrm{E}\Big(\frac{\partial^2 l}{\partial\sigma_i^2\partial\Sigma_{j,k}}\Big) = 0,\\
\mathrm{E}\Big(\frac{\partial^2 l}{\partial\Sigma_{i,i}\partial\Sigma_{j,j}}\Big) &= \mathrm{E}\Big(\frac{\partial^2 l}{\partial\Sigma_{i,i}\partial\Sigma_{j,k}}\Big) = \mathrm{E}\Big(\frac{\partial^2 l}{\partial\Sigma_{i,j}\partial\Sigma_{k,l}}\Big) = 0, (i,j) \neq (k,l).
\end{aligned}
$$

Assume that $k_i, k_{ii}, k_{ij} \longrightarrow \infty, i, j = 1, \cdots, m$. To make it simple, assume $k_{mm} = \min\{k_i, k_{ii}, k_{ij}\}$. Then we can show that $-\frac{1}{k_{mm}}\frac{\partial^2 l}{\partial\gamma\partial\gamma^{\tau}}$ and $-\frac{1}{k_{mm}}\mathrm{E}\Big(\frac{\partial^2 l}{\partial\rho\partial\rho^{\tau}}\Big)$ are positive

definite. Now we are in a position to use the method in Miller (1977) and Pinheiro (1994) according to the theory of Weiss (1971, 1973). Actually, taking k_{mm} to replace v_j we can see that the key condition, i.e., Assumption 3.1.7 of Pinheiro (1994), p28, holds. Then by the same arguments in Pinheiro (1994), Chapter 3, we can show that $\sqrt{k_{mm}}\hat{\gamma}$ converges to normal in distribution. This implies that the test statistic F_{het} is asymptotically $F_{m-1,n-m}$ by considering the denominator of F_{het} as the estimate of mean squared error, which is independent of the numerator of F_{het} (Pinheiro 1994, pp28-29; Graybill 1976).

In above discussion, we assume that there are sufficiently large data which include both trio families and sib-pair families. In addition, suppose we have nuclear families with any number children. We can show that $-\frac{1}{k_{mm}}\frac{\partial^2 l}{\partial\gamma\partial\gamma^\tau}$ and $-\frac{1}{k_{mm}}\,\mathrm{E}\left(\frac{\partial^2 l}{\partial\rho\partial\rho^\tau}\right)$ are positive definite. Then, we can keep on using the method of Pinherio (1994), chapters 2-3, to show that $\sqrt{k_{mm}}\hat{\gamma}$ is asymptotically normal. Hence, the statistic F_{het} is asymptotically $F(m-1, n-m)$-distributed.

APPENDIX F

If $n_i = 1$ for each family, then there is only one child in each family. Let $k_i, i = 1, 2, \cdots, m$ be the number of offspring who receive allele M_i from their heterozygous parents. Let I_k be identity $k \times k$ matrix. The design matrix and the variance-covariance matrix can be written as $X = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \Gamma = \text{diag}(\sigma_1^2 I_{k_1}, \cdots, \sigma_m^2 I_{k_m})$.

Then we have $X^\tau \Gamma^{-1} X = \text{diag}(k_1/\sigma_1^2, k_2/\sigma_2^2, \cdots, k_m/\sigma_m^2)$. Using a fact of inverse matrix $(A + ab^\tau)^{-1} = A^{-1} - (A^{-1}a)(b^\tau A^{-1})/(1 + b^\tau A^{-1} a)$, we can calculate

$$\begin{aligned} \left(H[X^\tau \Gamma^{-1} X]^{-1} H^\tau\right)^{-1} &= \left[\begin{pmatrix} \sigma_2^2/k_2 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \sigma_m^2/k_m \end{pmatrix} + \frac{\sigma_1^2}{k_1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \right]^{-1} \\ &= \begin{pmatrix} k_2/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & k_m/\sigma_m^2 \end{pmatrix} - \begin{pmatrix} k_2/\sigma_2^2 \\ \vdots \\ k_m/\sigma_m^2 \end{pmatrix} \frac{(\, k_2/\sigma_2^2, \cdots, k_m/\sigma_m^2 \,)}{\frac{k_1}{\sigma_1^2} + \cdots + \frac{k_m}{\sigma_m^2}}. \end{aligned}$$

Therefore, the non-centrality parameter $\lambda_{het,singleton} \approx (H\gamma)^\tau [H(X^\tau \Gamma^{-1} X)^{-1} H^\tau]^{-1} H$ $= \sum_{i=2}^m (\alpha_1 - \alpha_i)^2 k_i/\sigma_i^2 - \left[\sum_{i=2}^m (\alpha_1 - \alpha_i) k_i/\sigma_i^2\right]^2 / [\sum_{i=1}^m k_i/\sigma_i^2]$.

APPENDIX G

Such as in Appendix F, let us denote the variance-covariance matrix of the $\sum_{i=1}^{m} k_i$ singleton offspring by Γ_1, and the related design matrix by X_1. Now let Γ_2 denote the variance-covariance matrix of the $\sum_{i=1}^{m} k_{ii}$ sib-pairs, in each of them both sibs receive the same allele from their heterozygous parents, and X_2 the related design matrix. Then the form of X_2 is similar to X_1 given in Appendix F with different numbers of rows and $\Gamma_2 =$

$$\text{diag}\left(\begin{pmatrix} \sigma_1^2 & \Sigma_{1,1} \\ \Sigma_{1,1} & \sigma_1^2 \end{pmatrix}, \cdots, \begin{pmatrix} \sigma_1^2 & \Sigma_{1,1} \\ \Sigma_{1,1} & \sigma_1^2 \end{pmatrix}, \cdots, \begin{pmatrix} \sigma_m^2 & \Sigma_{m,m} \\ \Sigma_{m,m} & \sigma_m^2 \end{pmatrix}, \cdots, \begin{pmatrix} \sigma_m^2 & \Sigma_{m,m} \\ \Sigma_{m,m} & \sigma_m^2 \end{pmatrix}\right).$$

Let Γ_3 denote the variance-covariance matrix of the $\sum_{i=1}^{m} \sum_{j>i} k_{ij}$ sib pairs, in each of them one sib receives one allele (i.e., $M_i, i = 1, 2, \cdots, m$, respectively) from his/her heterozygous parent and the other receives the other allele (i.e., $M_j, j \neq i, j = 1, 2, \cdots, m$, respectively) from the same heterozygous parent, and X_3 be the related design matrix. The variance-covariance matrix Γ_3 is

$$\text{diag}\left(\begin{pmatrix} \sigma_1^2 & \Sigma_{1,2} \\ \Sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \cdots, \begin{pmatrix} \sigma_1^2 & \Sigma_{1,2} \\ \Sigma_{1,2} & \sigma_2^2 \end{pmatrix}, \cdots, \begin{pmatrix} \sigma_{m-1}^2 & \Sigma_{m-1,m} \\ \Sigma_{m-1,m} & \sigma_m^2 \end{pmatrix}, \cdots, \begin{pmatrix} \sigma_{m-1}^2 & \Sigma_{m-1,m} \\ \Sigma_{m-1,m} & \sigma_m^2 \end{pmatrix}\right).$$

The related design matrix is $X_3 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$. In the same manner of Appendix F, we may obtain that

$$\begin{aligned} X_1^\tau \Gamma_1^{-1} X_1 &= \text{diag}\left(\frac{k_1}{\sigma_1^2}, \frac{k_2}{\sigma_2^2}, \cdots, \frac{k_m}{\sigma_m^2}\right) \\ X_2^\tau \Gamma_2^{-1} X_2 &= \text{diag}\left(\frac{2k_{11}}{\sigma_1^2 + \Sigma_{1,1}}, \frac{2k_{22}}{\sigma_2^2 + \Sigma_{2,2}}, \cdots, \frac{2k_{mm}}{\sigma_m^2 + \Sigma_{m,m}}\right). \end{aligned}$$

After some calculation, one may obtain that

$$X_3^\tau \Gamma_3^{-1} X_3 = \begin{pmatrix} \sum_{i \neq 1} \frac{k_{1i}\sigma_i^2}{\sigma_1^2\sigma_i^2 - \Sigma_{1,i}^2} & -\frac{k_{12}\Sigma_{1,2}}{\sigma_1^2\sigma_2^2 - \Sigma_{1,2}^2} & \cdots & -\frac{k_{1m}\Sigma_{1,m}}{\sigma_1^2\sigma_m^2 - \Sigma_{1,m}^2} \\ -\frac{k_{12}\Sigma_{1,2}}{\sigma_1^2\sigma_2^2 - \Sigma_{1,2}^2} & \sum_{i \neq 2} \frac{k_{2i}\sigma_i^2}{\sigma_2^2\sigma_i^2 - \Sigma_{2,i}^2} & \cdots & -\frac{k_{2m}\Sigma_{2,m}}{\sigma_2^2\sigma_m^2 - \Sigma_{2,m}^2} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{k_{1m}\Sigma_{1,m}}{\sigma_1^2\sigma_m^2 - \Sigma_{1,m}^2} & -\frac{k_{2m}\Sigma_{2,m}}{\sigma_2^2\sigma_m^2 - \Sigma_{2,m}^2} & \cdots & \sum_{i \neq m} \frac{k_{mi}\sigma_i^2}{\sigma_m^2\sigma_i^2 - \Sigma_{m,i}^2} \end{pmatrix}.$$

APPENDIX H

To simplify notations, we omit subscripts ij from $\Delta_{ijQ}, \pi_{ijA}, \pi_{ijB}, \Delta_{ijA}, \Delta_{ijB}$ in the following appendices H, I, and J. Taking the variance-covariance for equation (3.6), we have the following matrix equation to calculate the coefficients

$$\text{Cov}\begin{pmatrix} (\pi_A,\pi_A) & (\pi_B,\pi_A) & (\Delta_A,\pi_A) & (\Delta_B,\pi_A) \\ (\pi_A,\pi_B) & (\pi_B,\pi_B) & (\Delta_A,\pi_B) & (\Delta_B,\pi_B) \\ (\pi_A,\Delta_A) & (\pi_B,\Delta_A) & (\Delta_A,\Delta_A) & (\Delta_B,\Delta_A) \\ (\pi_A,\Delta_B) & (\pi_B,\Delta_B) & (\Delta_A,\Delta_B) & (\Delta_B,\Delta_B) \end{pmatrix}\begin{pmatrix} \beta_A \\ \beta_B \\ r_A \\ r_B \end{pmatrix} = \text{Cov}\begin{pmatrix} (\Delta_Q,\pi_A) \\ (\Delta_Q,\pi_B) \\ (\Delta_Q,\Delta_A) \\ (\Delta_Q,\Delta_B) \end{pmatrix} \quad \text{(H.1)}$$

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 I, we will show that

$$\begin{aligned} \text{Cov}(\Delta_A,\pi_A) &= \text{Cov}(\Delta_B,\pi_B) = 1/8, \text{Cov}(\Delta_B,\pi_A) = \text{Cov}(\Delta_A,\pi_B) = (1-2\theta_{AB})^2/8, \\ \text{Cov}(\Delta_Q,\pi_A) &= (1-2\theta_{AQ})^2/8, \text{Cov}(\Delta_Q,\pi_B) = (1-2\theta_{QB})^2/8. \end{aligned} \quad \text{(H.2)}$$

Plugging the above results into the equation (H.1), 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

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

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 $3\rho(\Delta_i, \Delta_j)/2 - (1-2\theta_{ij})^2 = (1 - 8\theta_{ij} + 24\theta_{ij}^2 - 32\theta_{ij}^3 + 16\theta_{ij}^4)/2 = (1-2\theta_{ij})^4/2$ 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 (3.7) are valid.

APPENDIX I

Consider a sib-pair with trait values y_i and y_j. First, we have the following equation from Haseman (1970) (also see Amos 1994, equation (5) on p537 or Amos et al. 1989, p437)

$$\begin{aligned}\mathrm{Cov}(y_i, y_j|\pi_A, \Delta_A) &= \frac{1}{2}\sigma^2_{Ga} + \frac{1}{4}\sigma^2_{Gd} + \sigma^2_s + (1-\psi_A)\sigma^2_g + \psi_A(\psi_A - 1)\sigma^2_{gd} \\ &\quad +[-(1-2\psi_A)\sigma^2_g - (1-2\psi_A)^2\sigma^2_{gd}]\pi_A + (1-2\psi_A)^2\sigma^2_{gd}\Delta_A.\end{aligned}$$

Comparing the above equation with $\mathrm{Cov}(y_i, y_j|\pi_A, \Delta_A) = \pi_Q\sigma^2_{ga} + \Delta_Q\sigma^2_{gd} + \frac{1}{2}\sigma^2_{Ga} + \frac{1}{4}\sigma^2_{Gd} + \sigma^2_s$, we find

$$\Delta_Q = (1-\psi_A)^2 - [(1-2\psi_A) + (1-2\psi_A)^2]\pi_A + (1-2\psi_A)^2\Delta_A. \tag{I.1}$$

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

$$\mathrm{Cov}(\Delta_Q, \Delta_A) = -[(1-2\psi_A) + (1-2\psi_A)^2]\,\mathrm{Cov}(\pi_A, \Delta_A) + (1-2\psi_A)^2\,\mathrm{Cov}(\Delta_A, \Delta_A).$$

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

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

Similarly, we can show the other equations in (H.2).

APPENDIX J

To calculate the intercept α in (3.7), we consider the joint distribution of π_Q, π_A and π_B for a sib-pair. Assume that there is no interference for disjoint chromosome regions. Then

$$\begin{aligned}
&P(\pi_{ijA} = i_A, \pi_{ijQ} = i_Q, \pi_{ijB} = i_B) \\
&= P(\pi_{ijA} = i_A, \pi_{ijQ} = i_Q) P(\pi_{ijB} = i_B | \pi_{ijA} = i_A, \pi_{ijQ} = i_Q) \qquad \text{(J.1)} \\
&= P(\pi_{ijA} = i_A | \pi_{ijQ} = i_Q) P(\pi_{ijQ} = i_Q) P(\pi_{ijB} = i_B | \pi_{ijQ} = i_Q).
\end{aligned}$$

From Haseman and Elston (1972), Table IV, we construct the joint distribution of π_{ijQ}, π_{ijA} and π_{ijB} by equation (J.1); the results are presented in Table II. Consider a sib-pair with trait values y_i and y_j. Then from Table II we have

$$\begin{aligned}
&\mathrm{Cov}(y_i, y_j | \pi_A = 0, \pi_B = 0) - \Big[\frac{1}{2}\sigma^2_{Ga} + \frac{1}{4}\sigma^4_{Gd} + \sigma^2_s\Big] \\
&= (\sigma^2_{ga} + \sigma^2_{gd}) P(\pi_Q = 1 | \pi_A = 0, \pi_B = 0) + \frac{\sigma^2_{ga}}{2} P(\pi_Q = 1/2 | \pi_A = 0, \pi_B = 0) \\
&= \frac{(1-\psi_A)(1-\psi_B)}{\psi_A\psi_B + (1-\psi_A)(1-\psi_B)}\sigma^2_{ga} + \frac{(1-\psi_A)^2(1-\psi_B)^2}{[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]^2}\sigma^2_{gd}.
\end{aligned}$$

Therefore, we have the intercept α in (3.7) since it is the coefficient of σ^2_{gd} in above equation.

APPENDIX K

For simplicity, let us assume $\sigma^2 = 1$ and define $K = \frac{\sigma^2_{Ga}}{2} + \frac{\sigma^2_{Gd}}{4} + \sigma^2_s$. From Table II and equation (3.10), we may calculate

$$
\begin{aligned}
C_{22} &= \sigma^2_{ga}\frac{\psi_A\psi_B}{\psi_A\psi_B + (1-\psi_A)(1-\psi_B)} + \sigma^2_{gd}\frac{\psi^2_A\psi^2_B}{[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]^2} + K \\
C_{21} &= \frac{\sigma^2_{ga}}{2}\Big[\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}\Big] \\
&\quad + \sigma^2_{gd}\frac{\psi^2_A\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]} + K \\
C_{20} &= \sigma^2_{ga}\frac{\psi_A(1-\psi_B)}{\psi_A(1-\psi_B) + (1-\psi_A)\psi_B} + \sigma^2_{gd}\frac{\psi^2_A(1-\psi_B)^2}{[\psi_A(1-\psi_B) + (1-\psi_A)\psi_B]^2} + K \\
C_{12} &= \frac{\sigma^2_{ga}}{2}\Big[\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}\Big] \\
&\quad + \sigma^2_{gd}\frac{\psi_A(1-\psi_A)\psi^2_B}{[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)][\psi_A(1-\psi_B) + (1-\psi_A)\psi_B]} + K \\
C_{11} &= \frac{\sigma^2_{ga}}{2} + \sigma^2_{gd}\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} + K \\
C_{10} &= \frac{\sigma^2_{ga}}{2}\Big[\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}\Big] \\
&\quad + \sigma^2_{gd}\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]} + K \\
C_{02} &= \sigma^2_{ga}\frac{(1-\psi_A)\psi_B}{\psi_A(1-\psi_B) + (1-\psi_A)\psi_B} + \sigma^2_{gd}\frac{(1-\psi_A)^2\psi^2_B}{[\psi_A(1-\psi_B) + (1-\psi_A)\psi_B]^2} + K \\
C_{01} &= \frac{\sigma^2_{ga}}{2}\Big[\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}\Big] \\
&\quad + \sigma^2_{gd}\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]} + K \\
C_{00} &= \sigma^2_{ga}\frac{(1-\psi_A)(1-\psi_B)}{\psi_A\psi_B + (1-\psi_A)(1-\psi_B)} + \sigma^2_{gd}\frac{(1-\psi_A)^2(1-\psi_B)^2}{[\psi_A\psi_B + (1-\psi_A)(1-\psi_B)]^2} + K.
\end{aligned}
$$

APPENDIX L

For each y_i of the n individuals, $\Sigma_i = \sigma^2$ and $X_i = (\,1 \quad x_{Ai} \quad x_{Bi} \quad z_{Ai} \quad z_{Bi}\,), i = 1, 2, \cdots, n$. From formulas in Fan and Xiong (2002), Appendix A, we show that

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

where V_A and V_D are additive and dominant variance-covariance matrices of (3.3). For each of the m sib-pairs, the variance-covariance matrix $\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}$ and the model matrix $X_i = \begin{pmatrix} 1 & x_{A1}^{(i)} & x_{B1}^{(i)} & z_{A1}^{(i)} & z_{B1}^{(i)} \\ 1 & x_{A2}^{(i)} & x_{B2}^{(i)} & z_{A2}^{(i)} & z_{B2}^{(i)} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}, i = n+1, 2, \cdots, n+m$. 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 $\mathrm{E}\,[X_{i1}^\tau X_{i2}] = \mathrm{E}\,[X_{i2}^\tau X_{i1}] = diag(1, V_A/2, V_D/4)$. By above formulas and the formulas in Fan and Xiong (2002), Appendix A, we have the following

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

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 $X_i = \begin{pmatrix} 1 & x_{A1}^{(i)} & x_{B1}^{(i)} & z_{A1}^{(i)} & z_{B1}^{(i)} \\ 1 & x_{A2}^{(i)} & x_{B2}^{(i)} & z_{A2}^{(i)} & z_{B2}^{(i)} \\ 1 & x_{A3}^{(i)} & x_{B3}^{(i)} & z_{A3}^{(i)} & z_{B3}^{(i)} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \end{pmatrix}, i = n+m+1, 2, \cdots, n+m+k$. 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 $\mathrm{E}\,[X_{ij}^{\tau}X_{ik}] = \mathrm{E}\,[X_{ik}^{\tau}X_{ij}] = diag(1, V_A/2, V_D/4), j, k = 1, 2, 3, j \neq k$. 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

$$\frac{1}{k}\sum_{i=n+m+1}^{n+m+k} X_i^{\tau}\Sigma_i^{-1}X_i \approx \frac{1}{C_3\sigma^2}\Big[C_{31}diag(1, V_A, V_D) + C_{32}diag(1, V_A/2, V_D/4)\Big]. \quad \text{(L.3)}$$

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

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

Then equations (P.1), (P.3) and (L.3) lead to equation (3.8).

APPENDIX M

Taking variance-covariance among x_{ij}, z_{ij}, y_i of regression (4.2) leads to the following variance-covariance equations

$$\mathrm{Cov}\begin{pmatrix} (x_{i1},x_{i1}) & (x_{i2},x_{i1}) & \cdots & (x_{ik},x_{i1}) & (z_{i1},x_{i1}) & \cdots & (z_{ik},x_{i1}) \\ (x_{i1},x_{i2}) & (x_{i2},x_{i2}) & \cdots & (x_{ik},x_{i2}) & (z_{i1},x_{i2}) & \cdots & (z_{ik},x_{i2}) \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ (x_{i1},z_{ik}) & (x_{i2},z_{ik}) & \cdots & (x_{ik},z_{ik}) & (z_{i1},z_{ik}) & \cdots & (z_{ik},z_{ik}) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \delta_1 \\ \vdots \\ \delta_k \end{pmatrix} = \mathrm{Cov}\begin{pmatrix} (y_i,x_{i1}) \\ (y_i,x_{i2}) \\ \vdots \\ (y_i,x_{ik}) \\ (y_i,z_{i1}) \\ \vdots \\ (y_i,z_{ik}) \end{pmatrix} \quad \text{(M.1)}$$

In a similar way as Appendix A, Fan and Xiong (2002), the following expectations, variance and covariances can be derived accordingly: $Ex_{ij} = 0, Ez_{ij} = 0$, $E(x_{ij}^2) = \mathrm{Cov}(x_{ij}, x_{ij}) = 2P_{M_j}P_{m_j}, E(z_{ij}^2) = \mathrm{Cov}(z_{ij}, z_{ij}) = P_{M_j}^2 P_{m_j}^2, E(x_{ij}x_{il}) = \mathrm{Cov}(x_{ij}, x_{il}) = 2D_{M_jM_l}$, $E(z_{ij}z_{il}) = \mathrm{Cov}(z_{ij}, z_{il}) = D_{M_jM_l}^2, E(x_{ij}z_{il}) = \mathrm{Cov}(x_{ij}, z_{il}) = 0$, $\mathrm{Cov}(y_i, x_{ij}) = E(y_i x_{ij}) = 2D_{M_jQ}\alpha_Q$, $\mathrm{Cov}(y_i,\ z_{ij}) = E(y_i z_{ij}) = D_{M_jQ}^2\delta_Q$ for $j, l = 1, \cdots, k, j \neq l$. Plugging the above quantities into (M.1) gives

$$\begin{pmatrix} 2P_{M_1}P_{m_1} & 2D_{M_1M_2} & \cdots & 2D_{M_1M_k} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 2D_{M_1M_k} & 2D_{M_2M_k} & \cdots & 2P_{M_k}P_{m_k} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & P_{M_1}^2P_{m_1}^2 & \cdots & D_{M_1M_k}^2 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & D_{M_1M_k}^2 & \cdots & P_{M_k}^2P_{m_k}^2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ \delta_1 \\ \vdots \\ \delta_k \end{pmatrix} = \begin{pmatrix} 2D_{M_1Q}\alpha_Q \\ \vdots \\ 2D_{M_kQ}\alpha_Q \\ D_{M_1Q}^2\delta_Q \\ \vdots \\ D_{M_kQ}^2\delta_Q \end{pmatrix}.$$

Therefore, the coefficients of (4.3) are being derived.

APPENDIX N

To simplify notations, we omit subscripts ij from π_{ijQ},π_{ijM_1},$\cdots$, π_{ijM_k}, Δ_{ijM_1},$\cdots$, Δ_{ijM_k} in the appendices B and C. Taking variance-covariance among π_Q, π_{M_j}, y_i of equation (4.4) leads to

$$\text{Cov}\begin{pmatrix} (\pi_{M_1},\pi_{M_1}) & (\pi_{M_1},\pi_{M_2}) & \cdots & (\pi_{M_1},\pi_{M_k}) \\ (\pi_{M_1},\pi_{M_2}) & (\pi_{M_2},\pi_{M_2}) & \cdots & (\pi_{M_2},\pi_{M_k}) \\ \vdots & \vdots & \vdots & \vdots \\ (\pi_{M_1},\pi_{M_k}) & (\pi_{M_2},\pi_{M_k}) & \cdots & (\pi_{M_k},\pi_{M_k}) \end{pmatrix}\begin{pmatrix} \beta_{\pi M_1} \\ \beta_{\pi M_2} \\ \vdots \\ \beta_{\pi M_k} \end{pmatrix} = \text{Cov}\begin{pmatrix} (\pi_Q,\pi_{M_1}) \\ (\pi_Q,\pi_{M_2}) \\ \vdots \\ (\pi_Q,\pi_{M_k}) \end{pmatrix} \quad \text{(N.1)}$$

From Elston and Keats (1985) and Almasy and Blangero (1998), we have the following

$$\begin{aligned} \text{Cov}(\pi_{M_i},\pi_{M_i}) &= 1/8, i = 1,\cdots,k, \\ \text{Cov}(\pi_{M_i},\pi_{M_j}) &= (1-2\theta_{M_iM_j})^2/8, i \neq j = 1,\cdots,k, \\ \text{Cov}(\pi_Q,\pi_{M_i}) &= (1-2\theta_{M_iQ})^2/8, i = 1,\cdots,k. \end{aligned}$$

Plugging above quantities into equation(N.1) gives

$$\frac{1}{8}\begin{pmatrix} 1 & (1-2\theta_{M_1M_2})^2 & \cdots & (1-2\theta_{M_1M_k})^2 \\ (1-2\theta_{M_1M_2})^2 & 1 & \cdots & (1-2\theta_{M_2M_k})^2 \\ \vdots & \vdots & \vdots & \vdots \\ (1-2\theta_{M_1M_k})^2 & (1-2\theta_{M_2M_k})^2 & \cdots & 1 \end{pmatrix}\begin{pmatrix} \beta_{\pi M_1} \\ \beta_{\pi M_2} \\ \vdots \\ \beta_{\pi M_k} \end{pmatrix} = \frac{1}{8}\begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ (1-2\theta_{M_2Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \end{pmatrix},$$

which leads to

$$\begin{pmatrix} \beta_{\pi M_1} \\ \beta_{\pi M_2} \\ \vdots \\ \beta_{\pi M_k} \end{pmatrix} = \begin{pmatrix} 1 & (1-2\theta_{M_1M_2})^2 & \cdots & (1-2\theta_{M_1M_k})^2 \\ (1-2\theta_{M_1M_2})^2 & 1 & \cdots & (1-2\theta_{M_2M_k})^2 \\ \vdots & \vdots & \vdots & \vdots \\ (1-2\theta_{M_1M_k})^2 & (1-2\theta_{M_2M_k})^2 & \cdots & 1 \end{pmatrix}^{-1}\begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ (1-2\theta_{M_2Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \end{pmatrix}.$$

APPENDIX O

Taking variance-covariance among $\Delta_Q, \pi_{M_j}, \Delta_{M_l}$ of equation (4.5) leads to

$$\text{Cov}\begin{pmatrix} (\pi_{M_1},\pi_{M_1}) & \cdots & (\pi_{M_k},\pi_{M_1}) & (\Delta_{M_1},\pi_{M_1}) & \cdots & (\Delta_{M_k},\pi_{M_1}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (\pi_{M_1},\pi_{M_k}) & \cdots & (\pi_{M_k},\pi_{M_k}) & (\Delta_{M_1},\pi_{M_k}) & \cdots & (\Delta_{M_k},\pi_{M_k}) \\ (\pi_{M_1},\Delta_{M_1}) & \cdots & (\pi_{M_k},\Delta_{M_1}) & (\Delta_{M_1},\Delta_{M_1}) & \cdots & (\Delta_{M_k},\Delta_{M_1}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (\pi_{M_1},\Delta_{M_k}) & \cdots & (\pi_{M_k},\Delta_{M_k}) & (\Delta_{M_1},\Delta_{M_k}) & \cdots & (\Delta_{M_k},\Delta_{M_k}) \end{pmatrix}\begin{pmatrix} \beta_{M_1} \\ \vdots \\ \beta_{M_k} \\ r_{M_1} \\ \vdots \\ r_{M_k} \end{pmatrix}$$

$$= \text{Cov}\begin{pmatrix} (\Delta_Q,\pi_{M_1}) \\ \vdots \\ (\Delta_Q,\pi_{M_k}) \\ (\Delta_Q,\Delta_{M_1}) \\ \vdots \\ (\Delta_Q,\Delta_{M_k}) \end{pmatrix}. \tag{O.1}$$

As in Appendix N, the following covariances are from Elston and Keats (1985), Almasy and Blangero (1998) and Fan and Jung (2003)

$$\begin{aligned}
\text{Cov}(\Delta_{M_i},\pi_{M_i}) &= 1/8, i = 1,\cdots,k, \\
\text{Cov}(\Delta_{M_i},\pi_{M_j}) &= \text{Cov}(\Delta_{M_j},\pi_{M_i}) = (1-2\theta_{M_iM_j})^2/8, i,j = 1,\cdots,k, i \neq j, \\
\text{Cov}(\Delta_{M_i},\Delta_{M_i}) &= \frac{3}{16}, i = 1,\cdots,k, \\
\text{Cov}(\Delta_{M_i},\Delta_{M_j}) &= \frac{3}{16}\rho(\Delta_{M_i},\Delta_{M_j}), i,j = 1,\cdots,k, i \neq j \\
\text{Cov}(\Delta_Q,\pi_{M_i}) &= (1-2\theta_{M_iQ})^2/8, i = 1,\cdots,k, \\
\text{Cov}(\Delta_Q,\Delta_{M_i}) &= \frac{3}{16}\rho(\Delta_Q,\Delta_{M_i}), i = 1,\cdots,k,
\end{aligned}$$

where $\rho(\Delta_1, \Delta_2) = 1 - \frac{16}{3}\theta_{ij} + \frac{32}{3}\theta_{ij}^2 - \frac{32}{3}\theta_{ij}^3 + \frac{16}{3}\theta_{ij}^4$. Plugging the above results into the equation (O.1), we have a sub-matrix block equation

$$\begin{pmatrix} A & A \\ A & B \end{pmatrix} \begin{pmatrix} \beta_{M_1} \\ \vdots \\ \beta_{M_k} \\ r_{M_1} \\ \vdots \\ r_{M_k} \end{pmatrix} = \begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \\ 3\rho(\Delta_{M_i}, \Delta_Q)/2 \\ \vdots \\ 3\rho(\Delta_{M_k}, \Delta_Q)/2 \end{pmatrix},$$

where

$$A = \begin{pmatrix} 1 & (1-2\theta_{M_1M_2})^2 & \cdots & (1-2\theta_{M_1M_k})^2 \\ (1-2\theta_{M_1M_2})^2 & 1 & \cdots & (1-2\theta_{M_2M_k})^2 \\ \vdots & \vdots & \vdots & \vdots \\ (1-2\theta_{M_1M_k})^2 & (1-2\theta_{M_2M_k})^2 & \cdots & 1 \end{pmatrix},$$

$$B = \frac{3}{2}\begin{pmatrix} 1 & \rho(\Delta_{M_1}, \Delta_{M_2}) & \cdots & \rho(\Delta_{M_1}, \Delta_{M_k}) \\ \rho(\Delta_{M_1}, \Delta_{M_2}) & 1 & \cdots & \rho(\Delta_{M_2}, \Delta_{M_k}) \\ \vdots & \vdots & \vdots & \vdots \\ \rho(\Delta_{M_1}, \Delta_{M_k}) & \rho(\Delta_{M_2}, \Delta_{M_k}) & \cdots & 1 \end{pmatrix}.$$

Therefore, we have from Harville (1997) that

$$\begin{pmatrix} \beta_{M_1} \\ \vdots \\ \beta_{M_k} \\ r_{M_1} \\ \vdots \\ r_{M_k} \end{pmatrix} = \begin{pmatrix} A & A \\ A & B \end{pmatrix}^{-1} \begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \\ 3\rho(\Delta_{M_i}, \Delta_Q)/2 \\ \vdots \\ 3\rho(\Delta_{M_k}, \Delta_Q)/2 \end{pmatrix}$$

$$= \begin{pmatrix} A^{-1}+(B-A)^{-1} & -(B-A)^{-1} \\ -(B-A)^{-1} & (B-A)^{-1} \end{pmatrix} \begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \\ 3\rho(\Delta_{M_i},\Delta_Q)/2 \\ \vdots \\ 3\rho(\Delta_{M_k},\Delta_Q)/2 \end{pmatrix}.$$

The equation $3\rho(\Delta_i,\Delta_j)/2-(1-2\theta_{ij})^2 = (1-8\theta_{ij}+24\theta_{ij}^2-32\theta_{ij}^3+16\theta_{ij}^4)/2 = (1-2\theta_{ij})^4/2$ leads to

$$\begin{pmatrix} r_{M_1} \\ r_{M_2} \\ \vdots \\ r_{M_k} \end{pmatrix} = (B-A)^{-1} \begin{pmatrix} 3\rho(\Delta_{M_1},\Delta_Q)/2-(1-2\theta_{M_1Q})^2 \\ 3\rho(\Delta_{M_2},\Delta_Q)/2-(1-2\theta_{M_2Q})^2 \\ \vdots \\ 3\rho(\Delta_{M_k},\Delta_Q)/2-(1-2\theta_{M_kQ})^2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & (1-2\theta_{M_1M_2})^4 & \cdots & (1-2\theta_{M_1M_k})^4 \\ (1-2\theta_{M_1M_2})^4 & 1 & \cdots & (1-2\theta_{M_2M_k})^4 \\ \vdots & \vdots & \vdots & \vdots \\ (1-2\theta_{M_1M_k})^4 & (1-2\theta_{M_2M_k})^4 & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} (1-2\theta_{M_1Q})^4 \\ (1-2\theta_{M_2Q})^4 \\ \vdots \\ (1-2\theta_{M_kQ})^4 \end{pmatrix}.$$

Moreover, we have

$$\begin{pmatrix} \beta_{M_1} \\ \beta_{M_2} \\ \vdots \\ \beta_{M_k} \end{pmatrix} = A^{-1} \begin{pmatrix} (1-2\theta_{M_1Q})^2 \\ (1-2\theta_{M_2Q})^2 \\ \vdots \\ (1-2\theta_{M_kQ})^2 \end{pmatrix} - (B-A)^{-1} \begin{pmatrix} 3\rho(\Delta_{M_1},\Delta_Q)/2-(1-2\theta_{M_1Q})^2 \\ 3\rho(\Delta_{M_2},\Delta_Q)/2-(1-2\theta_{M_2Q})^2 \\ \vdots \\ 3\rho(\Delta_{M_k},\Delta_Q)/2-(1-2\theta_{M_kQ})^2 \end{pmatrix}$$

$$= \begin{pmatrix} \beta_{\pi M_1} \\ \beta_{\pi M_2} \\ \vdots \\ \beta_{\pi M_k} \end{pmatrix} - \begin{pmatrix} r_{M_1} \\ r_{M_2} \\ \vdots \\ r_{M_k} \end{pmatrix}.$$

APPENDIX P

To derive a_1, a_2, a_3 in approximation (4.6), we assume three sub-samples of a population: n individuals, m trio families each has both parents and a single child, and s nuclear families each has both parents and two offspring.

(a) For each y_i of the n individuals, $\Sigma_i = \sigma^2$ and $X_i = (1, x_{i1}, \cdots, x_{ik}, z_{i1}, \cdots, z_{ik}), i = 1, \cdots, n$. When the sample size n of individuals is large, the large number law leads to

$$\begin{aligned}
\frac{1}{n}X^tX &= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} n & x_{i1} & x_{i2} & \cdots & x_{ik} & z_{i1} & \cdots & z_{ik} \\ x_{i1} & x_{i1}^2 & x_{i2}x_{i1} & \cdots & x_{ik}x_{i1} & z_{i1}x_{i1} & \cdots & z_{ik}x_{i1} \\ x_{i2} & x_{i1}x_{i2} & x_{i2}^2 & \cdots & x_{ik}x_{i2} & z_{i1}x_{i2} & \cdots & z_{ik}x_{i2} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ z_{ik} & x_{i1}z_{ik} & x_{i2}z_{ik} & \cdots & x_{ik}z_{ik} & z_{i1}z_{ik} & \cdots & z_{ik}^2 \end{pmatrix} \\
&\approx \begin{pmatrix} 1 & Ex_{i1} & Ex_{i2} & \cdots & Ex_{ik} & Ez_{i1} & \cdots & Ez_{ik} \\ Ex_{i1} & Ex_{i1}^2 & Ex_{i2}x_{i1} & \cdots & Ex_{ik}x_{i1} & Ez_{i1}x_{i1} & \cdots & Ez_{ik}x_{i1} \\ Ex_{i2} & Ex_{i1}x_{i2} & Ex_{i2}^2 & \cdots & Ex_{ik}x_{i2} & Ez_{i1}x_{i2} & \cdots & Ez_{ik}x_{i2} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ Ez_{ik} & Ex_{i1}z_{ik} & Ex_{i2}z_{ik} & \cdots & Ex_{ik}z_{ik} & Ez_{i1}z_{ik} & \cdots & Ez_{ik}^2 \end{pmatrix} \\
&= diag(1, V_A, V_D).
\end{aligned}$$

Therefore, we have the following approximation

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

where V_A and V_D are additive and dominant variance-covariance matrices defined by (**??**).

(b) For i-th trio family, let $(y_{fi}, y_{mi}, y_{i1})^\tau$ be the trait values, and $X_i = (X_{fi}, X_{mi}, X_{i1})^\tau$ be the related model matrix, $i = n+1, \cdots, n+m$. In the same way as Appendix A of Fan and Xiong (2003), the covariance matrix between parents and their offspring can be shown to be

$$\text{E } X_{fi}^\tau X_{i1} = \text{E } X_{mi}^\tau X_{i1} = \begin{pmatrix} V_A/2 & O_k \\ O_k & O_k \end{pmatrix}, \tag{P.2}$$

where O_k is zero $k \times k$ matrix. For each of the m trio families, the variance-covariance matrix $\Sigma_i = \sigma^2 \begin{pmatrix} 1 & 0 & \rho_0 \\ 0 & 1 & \rho_0 \\ \rho_0 & \rho_0 & 1 \end{pmatrix}$. The inverse matrix of Σ_i is $\Sigma_i^{-1} = \frac{1}{(1-2\rho_0^2)\sigma^2} \begin{pmatrix} 1-\rho_0^2 & \rho_0^2 & -\rho_0 \\ \rho_0^2 & 1-\rho_0^2 & -\rho_0 \\ -\rho_0 & -\rho_0 & 1 \end{pmatrix}$. By above formulae, we can show the following

$$\frac{1}{m} \sum_{i=n+1}^{n+m} X_i^\tau \Sigma_i^{-1} X_i \approx \frac{2}{(1-2\rho_0^2)\sigma^2} \begin{pmatrix} 3-4\rho_0 & 0 & 0 \\ 0 & (3-2\rho_0-2\rho_0^2)V_A & 0 \\ 0 & 0 & (3-2\rho_0^2)V_D \end{pmatrix} \tag{P.3}$$

(c) For the i-th family which composes of both parents and two offspring, let $(y_{fi}, y_{mi}, y_{i1},\ y_{i2})^\tau$ be the trait values, and $X_i = (X_{fi}, X_{mi}, X_{i1}, X_{i2})^\tau$ be the related model matrix, $i = n+m+1, \cdots, n+m+s$. In the same way as Appendix C of Fan and Xiong (2003), it can be shown that

$$\text{E } X_{i1}^\tau X_{i2} = \begin{pmatrix} V_A/2 & O_k \\ O_k & V_D/4 \end{pmatrix}. \tag{P.4}$$

For each of the s families, the inverse variance-covariance matrix

$$\Sigma_i^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 1+2\rho_0 C & 2\rho_0 C & -C & -C \\ 2\rho_0 C & 1+2\rho_0 C & -C & -C \\ -C & -C & \frac{C(1-2\rho_0^2)}{\rho_0(1-\rho_{12})} & -\frac{C(\rho_1 2-2\rho_0^2)}{\rho_0(1-\rho_{12})} \\ -C & -C & -\frac{C(\rho_1 2-2\rho_0^2)}{\rho_0(1-\rho_{12})} & \frac{C(1-2\rho_0^2)}{\rho_0(1-\rho_{12})} \end{pmatrix} \tag{P.5}$$

where $C = \rho_0(1-\rho_{12})/[(1-2\rho_0^2)^2 - (\rho_{12}-2\rho_0^2)^2]$. Using (P.2), (P.4) and (P.5), we can show

$$\frac{1}{s}\sum_{i=n+m+1}^{n+m+s} X_i^\tau \Sigma_i^{-1} X_i \approx diag(d_{11}, d_{22}V_A, d_{44}V_D) \tag{P.6}$$

where the constants are given by $d_{11} = 2[1+4C\rho_0-4C+C/\rho_0], d_{22} = 2+4C(\rho_0-1)+C(2-\rho_{12}-2\rho_0^2)/[\rho_0(1-\rho_{12})], d_{44} = 2(1+2C\rho_0)+C[4(1-2\rho_0^2)-(\rho_{12}-2\rho_0^2)]/[2\rho_0(1-\rho_{12})]$. Combining the n individuals, m trio families, and s families with two offspring, the equations (P.1), (P.3) and (P.6) lead to $\sum_{i=1}^{n+m+s} X_i^\tau \Sigma_i^{-1} X_i \approx diag(a_1, a_2V_A, a_3V_D)/\sigma^2$, where

$$\begin{aligned} a_1 &= n + m(1-2\rho_0^2)^{-1}(3-4\rho_0) + sd_{11}, \\ a_2 &= n + m(1-2\rho_0^2)^{-1}(3-2\rho_0-2\rho_0^2) + sd_{22}, \\ a_3 &= n + m(1-2\rho_0^2)^{-1}(3-2\rho_0^2) + sd_{44}. \end{aligned} \tag{P.7}$$

APPENDIX Q

Using (P.2) and (P.4), we can show approximation (4.7). The constants b_1 and b_2 are given by

$$\begin{aligned} b_1 &= \sum_{j=1}^{l+2} \gamma_{jj} + (\gamma_{13} + \cdots + \gamma_{1,l+2}) + (\gamma_{23} + \cdots + \gamma_{2,l+2}) + \sum_{h=3}^{l+2} \sum_{j=h+1}^{l+2} \gamma_{hj}, \\ b_2 &= \sum_{j=1}^{l+2} \gamma_{jj} + \sum_{h=3}^{l+2} \sum_{j=h+1}^{l+2} \gamma_{hj}/2. \end{aligned} \tag{Q.1}$$

VITA

Jeesun Jung was born in Busan, Korea on September 7, 1972. She is the second daughter of Hongsuk Jung and Kunhae Kim. She graduated from Inje University in Kimhae, Korea in February 1995 with a Bachelor of Science degree in statistics. On August 1998, she received a Master of Art degree in statistics under the supervision of Dr. Sangun Yun from Yonsei University in Seoul, Korea. She completed her Ph.D. in Statistics at Texas A&M University in August 2004. Her speciality is statistical genetics focused on quantitative trait loci mapping.

Her permanent address is :

34-17 Yonji Dong Busanjin Gu

Busan, Republic of Korea