Genetic Epidemiology 32: 132–142 (2008)

A Genomic Imprinting Test for Ordinal Traits in Pedigree Data

Rui Feng[1]* and Heping Zhang[2]

[1]*Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama*
[2]*Department of Epidemiology and Public Health, Yale University, New Haven, Connecticut*

Genomic imprinting can lead maternally and paternally derived alleles with identical nucleotide sequences to function differently and has been found to affect the complex inheritance of a variety of human disorders. Statistical methods that differentiate the parent-of-origin effects on human diseases are available for binary traits and continuous traits. However, numerous common diseases are measured on discrete ordinal scales. Imprinting may also contribute to the complex genetic basis of these traits. In a previous study, we proposed a latent variable model and developed computationally efficient score statistic to test linkage of ordinal traits for any size pedigree while adjusting for non-genetic covariates. In this study, we extend the latent variable model to incorporate parent-of-origin information and further develop a score statistic for testing the imprinting effect in linkage analysis. We evaluated the properties of our test statistic using simulations. We then applied our method to the Collaborative Study on the Genetics of Alcoholism and found a novel locus on chromosome 18 that shows a strong signal for imprinting. In addition, we identified two loci on chromosomes 3 and 4 significantly ($p < 0.0001$) linked with alcoholism. *Genet. Epidemiol.* 32:132–142, 2008.

*Correspondence to: Rui Feng, 1665 University Blvd, RPHB 327, Birmingham, AL 35294. E-mail: rfeng@ms.soph.uab.edu
Received 13 February 2007; Revised 11 June 2007; Accepted 3 September 2007
Published online 5 October 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20270

INTRODUCTION

Genomic imprinting can lead maternally and paternally derived alleles with identical nucleotide sequences to function differently. For example, the expression of one set of alleles can be completely or partially silenced if it is derived from the mother and not from the father because of differential epigenetic marks (such as methylation and acetylation) imposed on male and female gametes [Adams et al., 2000]. Imprinting plays a critical role in gene expression, mammalian development, and human disease [Everman and Cassidy, 2000]. In the past decade a wide range of disorders, including neonatal diabetes, Silver-Russell syndrome, Beckwith Wiedemann syndrome, Angelman Syndrome, Prader-Willi Syndrome, Alzheimer's disease, autism, diabetes, and schizophrenia, were found to be related to maternal imprinting on chromosomes 6, 7, 11, 14, and 15 [Huxtable et al., 2000; Francks et al., 2003; Bartlett and Vieland, 2005; Luedi et al., 2005; Bassett et al., 2006]. Beckwith-Wiedemann syndrome, a developmental disorder, is associated with two maternally expressed growth-regulatory genes on chromosome 11 [Rump et al., 2005]. An allele on a paternally expressed region at chromosome 6 is responsible for transient neonatal diabetes [Temple and Shield, 2002; Rump et al., 2005].

Statistical methods that can differentiate the parent-of-origin effects on human diseases are available for binary traits [Weinberg et al., 1998; Weinberg, 1999; Strauch et al., 2000] and continuous traits [Knapp and Strauch, 2004; Shete et al., 2003; Whittaker et al., 2003]. However, numerous diseases, such as asthma, cancer, and most psychiatric disorders and neurodegenerations, are measured on discrete ordinal scales. Genomic imprinting may also contribute to the complex genetic basis of these traits [Steinke et al., 2003]. Furthermore, most methods to test imprinting are limited to sibs, relative pairs, or case-parent triads [Weinberg, 1999; Hanson et al., 2001; Karason et al., 2003; Knapp and Strauch, 2004; Vincent et al., 2006]. Using large pedigree information directly can lead to more efficient and more powerful methods than dividing pedigrees into sibs [Wijsman and Amos, 1997; Shete et al., 2003], but only a few can be applied for extended pedigrees [Strauch et al., 2000; Shete et al., 2003].

In a previous study, we developed a latent variable model and used a likelihood ratio test to map genes for ordinal traits from pedigree data that showed increased power compared with methods using dichotomized traits [Feng et al., 2004]. However, computation of the likelihood ratio in relatively large pedigrees is overwhelming and often infeasible. In this study, we extend the model to accommodate different parent-of-origin effects. To alleviate the computation burden and address the complication introduced by imprinting in general pedigree data, we use a score statistic based on the identity by

decent (IBD) sharing information within pedigrees. The score test improves the speed dramatically. Although we expect the likelihood ratio test to be more powerful than the score test, computational complexity makes the likelihood ratio statistic difficult to obtain and is the reason why the score test outperformed the likelihood ratio test numerically in our previous study [Feng and Zhang, 2007]. After presenting the score test for detecting imprinting, we evaluate the method using both simulations and an application to data from the Collaborative Study on the Genetics of Alcoholism (COGA).

METHOD

The regression models for linkage analysis establish the association of a trait and the inheritance pattern at genetic markers. The inheritance pattern of a pedigree at a locus is inferred from the observed marker information through the so-called inheritance vectors [Kruglyak et al., 1996]. In this section, we first explain the parent-of-origin specific inheritance vectors and define separate latent variables depending on the parental origin of alleles. We then introduce our regression model that links the ordinal trait with the genetic susceptibility transmitted according to the inferred pattern at a putative locus. Finally, based on the regression model, we derive a computationally efficient score statistic to detect imprinting.

PARENT-OF-ORIGIN SPECIFIC INHERITANCE VECTORS AND LATENT VARIABLES

For a given a pedigree, the inheritance vector of non-founders (subjects whose parents are included in the pedigree) at a given locus q describes the ancestral origin of the DNA inherited by every non-founder at that locus [Lander and Green, 1987]. For instance, in a nuclear family with two parents and n siblings, we may arbitrarily index the two alleles of the father at q as 1 and 2 and similarly two alleles of the mother as 3 and 4. For the jth sibling, define $v_{p,j}=1$ or 2 according to which paternal allele is transmitted and $v_{m,j}=3$ or 4 according to which maternal allele is transmitted. Then the inheritance vector of the n siblings, $v(q)=(v_{p,1}, v_{m,1}, v_{p,2}, v_{m,2}, \ldots, v_{p,n}, v_{m,n})'$, completely specifies which of the four distinct paternal and maternal alleles are transmitted to every sibling. For a more complex pedigree with f founders and n non-founders, we can index the alleles of the f founders as (1, 2), (3, 4), (5, 6), …, (2f–1, 2f) and define the inheritance vector for the n non-founders similarly.

For the ith family, we assume there exist two types of latent random variables U^i_g and U^i_e that represent, respectively, (a) the genetic susceptibility at a marker of interest and (b) the residual genetic or environmental factors in a family that are unobserved or difficult to assess through the observed data.

U_g depends on the inheritance vector in the pedigree as follows. Let $U^i_{g,1},\ldots,U^i_{g,2f_i}$ be the genetic susceptibility associated with the $2f_i$ alleles at a disease susceptibility locus (DSL) on $2f_i$ pairs of chromosomes of all f_i founders in the ith family. In a simple pedigree with two founders, $U^i_{g,1}$ and $U^i_{g,2}$ represent the genetic susceptibility associated with the two alleles of the father at the DSL and $U^i_{g,3}$ and $U^i_{g,4}$ represent the genetic susceptibility associated with the two alleles of the mother at the DSL. For the jth founder in the ith family, his or her latent variables are $U^i_{g,2j}$ and $U^i_{g,2j-1}$, where $2j$ and $2j-1$ are arbitrarily labeled, reflecting the genetic susceptibility from two alleles with unknown parental origins. For the jth non-founder in the ith family, his or her latent variables are $U^i_{g,v_{p,j}}$ and $U^i_{g,v_{m,j}}$, reflecting the genetic susceptibility due to the maternally derived allele and the paternally derived allele at the locus, respectively. We use $U^i_{gp,j}$ and $U^i_{gm,j}$ to denote the genetic susceptibility due to the paternal and maternal alleles for each individual, respectively, i.e., $U^i_{gp,j}=U^i_{g,v_{p,j}}$ and $U^i_{gm,j}=U^i_{g,v_{m,j}}$ for non-founders.

A PROPORTIONAL-ODDS MODEL FOR ORDINAL TRAITS

We consider a trait Y taking an ordinal value from 0, 1, …, $K(K\geq 1)$. Let $\mathbf{x}$ be a p-vector of covariates that is also available for every study subject. All U^i_gs of the founders and the U^i_es are assumed to be independently and identically distributed across families. Conditional on all latent variables, denoted by U^i, and inheritance vectors v^i within the ith family, the traits of all non-founders are independent and follow the distribution

$$\begin{aligned}\log it(P\{Y^i_j\leq k|U^i,v^i\}) &= \boldsymbol{x}^i_j\boldsymbol{\beta}+\alpha_k+U^i_{gp,j}\gamma_p+U^i_{gm,j}\gamma_m\\ &+U^i_e\gamma_e, \text{for } i=1,\ldots,n, j=1,\ldots,n_i, k=0,1,\ldots,K-1,\end{aligned} \quad (1)$$

where n is the total number of pedigrees, n_i is the number of individuals in the ith family, $\boldsymbol{\beta}$ is a p-vector of parameters reflecting the covariate effects on the trait, and α_ks are the trait-level-dependent intercepts reflecting the differences between cumulative probabilities $P(Y\leq k)$. We must have $\alpha_0\leq\alpha_1\leq\ldots\leq\alpha_{K-1}$ so that the category probabilities $P(Y^i_j=k)$ are non-negative [McCullagh and Nelder, 1989]. Parameters γ_p and γ_m indicate two corresponding genetic effects due to the paternal and maternal alleles, and γ_e indicates the contribution

due to non-observed shared genetic and non-genetic factors.

Here, we consider the additive genetic susceptibility due to the gene at DSL; the dominant or recessive susceptibility can be incorporated as well. If the allele frequencies at DSL are the same for both male and female founders in the population, we can assume standardized mean and variance for U, i.e., $E[U^i_{gp,j}] = E[U^i_{gm,j}] = 0$ and $\mathrm{Var}[U^i_{gp,j}] = Var[U^i_{gm,j}] = 1$ for $j = 1, 2, \ldots, 2f_i$ without loss of generality. We also assume that U_es follow the standard normal distribution.

SCORE TEST FOR IMPRINTING

Let $\boldsymbol{\gamma} = (\gamma_p, \gamma_m)'$ and $\omega = (\boldsymbol{\beta}, \alpha_0, \ldots, \alpha_{K-1}, \gamma_e)$. We test the null hypothesis that there is no linkage and no imprinting, i.e., γ_p (paternal imprinting) $= \gamma_m$ (maternal imprinting) $= 0$. Let S_p and S_m be the score functions for testing the paternal imprinting and maternal imprinting, respectively. We obtained the score vector regarding to $\boldsymbol{\gamma}$ as follows (see appendix for details):

$$\boldsymbol{S} = \begin{pmatrix} S_p \\ S_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{l=1}^{n_i} \frac{\int \ddot{P}^i_{jl}(\omega,\mathbf{0})P^i(\omega,\mathbf{0})dF(U^i_e)q^i_{p,jl}}{2\int P^i(\omega,\mathbf{0})dF(U^i_e)} \\ \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{l=1}^{n_i} \frac{\int \ddot{P}^i_{jl}(\omega,\mathbf{0})P^i(\omega,\mathbf{0})dF(U^i_e)q^i_{m,jl}}{2\int P^i(\omega,\mathbf{0})dF(U^i_e)} \end{pmatrix}, \quad (2)$$

where

$$\begin{aligned}\ddot{P}^i_{jl}(\omega,\mathbf{0}) =& I(Y^i_j = r, Y^i_l = t)\left(1 - \pi^i_{r,j}(\omega,\mathbf{0}) - \pi^i_{r-1,j}(\omega,\mathbf{0})\right) \\ & \left(1 - \pi^i_{t,l}(\omega,\mathbf{0}) - \pi^i_{t-1,l}(\omega,\mathbf{0})\right), \text{ for } j \neq l \\ \ddot{P}^i_{jj}(\omega,\mathbf{0}) =& -I(Y^i_j = r)\Big[\pi^i_{r,j}(\omega,\mathbf{0})\left(1 - \pi^i_{r,j}(\omega,\mathbf{0})\right) \\ & + \pi^i_{r-1,j}(\omega,\mathbf{0})\left(1 - \pi^i_{r-1,j}(\omega,\mathbf{0})\right)\Big],\end{aligned}$$

$$\begin{aligned}\pi^i_{r,j}(\omega,\mathbf{0}) =& \frac{\exp(x^i_j\boldsymbol{\beta} + \alpha_r + U^i_e\gamma_e)}{1 + \exp(x^i_j\boldsymbol{\beta} + \alpha_r + U^i_e\gamma_e)} \\ & \text{for } r = 0, 1, \ldots, K-1,\ \pi^i_{-1,j} = 0, \pi^i_{K,j} = 1.\end{aligned}$$

$F(\cdot)$ is the cumulative distribution function of U_g, and $q^i_{p,jl}$ and $q^i_{m,jl}$ are the number of paternal alleles and maternal alleles shared by the jth and lth individuals, respectively. The two elements of $\mathbf{S}$, S_p and S_m, can indicate the evidence for imprinting—paternal or maternal depending on whether S_p or S_m is greater.

The asymptotic variance of $\mathbf{S}$ under the null hypothesis can be calculated directly [Cox and Hinkley, 1974].

$$\boldsymbol{I_S} = \boldsymbol{I}_{\gamma\gamma} - \boldsymbol{I}_{\gamma\omega}\boldsymbol{I}^{-1}_{\omega\omega}\boldsymbol{I}^{T}_{\gamma\omega}, \quad (3)$$

where $\boldsymbol{I}_{\gamma\gamma} = \sum_{i=1}^{N}\left(\frac{\partial l^i}{\partial\gamma}\right)^T\left(\frac{\partial l^i}{\partial\gamma}\right)$, $I_{\gamma\omega} = \sum_{i=1}^{N}\left(\frac{\partial l^i}{\partial\gamma}\right)^T\left(\frac{\partial l^i}{\partial\omega}\right)$, $I_{\omega\omega} = \sum_{i=1}^{N}\left(\frac{\partial l^i}{\partial\omega}\right)^T\left(\frac{\partial l^i}{\partial\omega}\right)$, and the expectation is taking over the distribution of Y.

Let $\hat{\mathbf{S}}$ and $\hat{\mathbf{I}}_S$ be the values of $\mathbf{S}$ and $\mathbf{I}_S$, respectively, calculated at the corresponding estimates $\hat{\omega}$ under the null hypothesis. The score statistic then equals $S_1 = \hat{\mathbf{S}}^T\hat{\mathbf{I}}^{-1}_{\mathbf{S}}\hat{\mathbf{S}}$ for testing both linkage and imprinting.

Without differentiating the effects due to different parental origins, we let $\gamma_0 = \gamma_p = \gamma_m$ in model (1) and the new model is $\log it(P\{Y^i_j \le k|U^i, v^i\}) = x^i_j\boldsymbol{\beta} + \alpha_k + (U^i_{gp,j} + U^i_{gm,j})\gamma_0 + U^i_e\gamma_e$, for $i = 1, \ldots, n, \quad j = 1, \ldots, n_i$, $k = 0, 1, \ldots, K-1$. The null hypothesis for testing evidence of linkage becomes $\gamma_0 = 0$. The score function in this setting is

$$S = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{l=1}^{n_i}\frac{\int \ddot{P}^i_{jl}(\omega,\mathbf{0})P^i(\omega,\mathbf{0})dF(U^i_e)q^i_{jl}}{2\int P^i(\omega,\mathbf{0})dF(U^i_e)}, \quad (4)$$

where q^i_{jl} is the expected number of general IBD alleles shared by the jth and lth members (without differentiating the parental origins) in the ith pedigree, given the marker information [Feng and Zhang, 2007]. The score statistic then equals $S_0 = \hat{S}^T\hat{I}^{-1}_S\hat{S}$, where $\hat{S}$ and $\hat{I}_S$ are values of the S and the asymptotic variance of S, respectively, calculated at the corresponding estimates $\hat{\omega}$ under the null hypothesis. Under the null hypothesis that there is no linkage, the asymptotic distribution of the score function S_0 is χ^2_1 [Casella and Berger, 2002; Feng and Zhang, 2007].

The difference between two scores S_1 and S_0 indicates the extra contribution due to parent-of-origin effect and a therefore-defined score statistic, $S_{im} = S_1 - S_0$, will suggest evidence of imprinting, which follows χ^2_1 asymptotically. The score statistic S_{im} is very computationally efficient and can test imprinting for any sized pedigrees while adjusting for non-genetic covariates.

SIMULATION

We performed a series of simulations to examine the empirical distribution of our score test statistics. First, we verified whether the empirical distribution of the test statistic is similar to the theoretical asymptotic distribution under the null hypothesis. Second, we examined the power of detecting imprinting when an imprinted gene is linked to markers.

EMPIRICAL DISTRIBUTION OF SCORE STATISTIC UNDER THE NULL HYPOTHESIS

The simulation was replicated 10,000 times, resulting in 10,000 data sets. For each data set, we generated 100 pedigrees with two parents and three offspring in each pedigree. A latent variable U_e^i was generated from $N(0,1)$ that is shared by all family members, and a random noise e_j^i was also generated from $N(0,1)$ for each individual.

For each founder in a pedigree, 20 highly polymorphic markers with 10 equally likely alleles [Speer et al., 1995], spaced 5 cM apart, were generated on one chromosome. Recombination fractions were converted to map distances without interference and there was no linkage disequilibrium among markers. After the genotypes were generated for the founders, the genotypes of non-founders were generated subsequently based on the recombination fractions.

We considered two possible scenarios under the null hypothesis: (1) no linkage (and thus no imprinting effect) and (2) a linkage but no imprinting. For the first scenario, all markers were generated independently. A liability variable Z_j^i for the jth person in the ith family was defined as $U_e^i + e_j^i$. And for the second scenario, we assumed a di-allelic disease locus between the 10th and 11th markers and the disease-causing allele, D, with frequency $P = 0.3$. The genetic contributions U_g of allele D and d in founders were set equal to 1 and -1, respectively. The genetic contributions from the same alleles were the same regardless of parental origin. The liability variable Z_j^i was defined as $U_e^i + 2(N_D - 1) + e_j^i$, where N_D is the number of allele D at the disease locus for each individual. For both settings the ordinal response, Y, equaled 0 if $Z < 0$, 1 if $0 \leq Z < 1$, or 2 if $Z \geq 1$.

Fig. 1 and 2 give the QQ plots of the empirical score distribution at a locus and the asymptotic score distribution χ_1^2. The nearly straight lines in both figures show that the empirical distribution approximates the theoretical distribution reasonably well for a modest sample size (100 pedigrees and a total of 500 subjects). The line in Fig. 1 has a slight departure from the diagonal but suggests non-inflated type I errors with a slightly smaller slope. At the nominal level of 0.05, the empirical type I errors are 0.055 and 0.044, respectively, for the two experiments.

As phenotype and genotypes are often missing from family data, we examined type I errors when the data are missing. We simulated 10,000 datasets under three scenarios: 10% of parental genotypes are randomly missing, 20% of parental genotypes are randomly missing, and both phenotype and genotypes of one randomly chosen parent in each family are missing. The empirical score distributions are also close to the asymptotic distribution for the different scenarios. When there is no linkage and no imprinting, the type I errors are 0.055, 0.055, and 0.061 for three different scenarios at the nominal level of 0.05, respectively; when there is linkage but no imprinting, the type I errors are 0.032, 0.032, and 0.049, respectively. If all the genotypes of both parents in a family are missing, we cannot infer the parental origin of the offspring's alleles and thus the family is not informative for testing imprinting. So when a small proportion of genotypes of parents are missing, type I error rates are reserved; especially for the test of imprinting in the presence of linkage, the type I errors are slightly conservative (the empirical confidence intervals are (0.029, 0.035) and (0.029, 0.035) based on $100 \times 10{,}000$ samples). But when both phenotype and genotypes of at least one parent are missing, type I error rates are slightly inflated. How to adjust type I errors for missing data deserves further investigations.

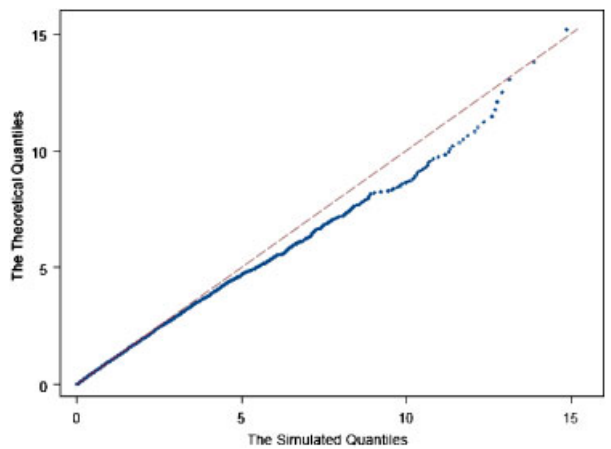

Fig. 1. The QQ plot of the score statistic in the absence of linkage based on 10,000 simulations.

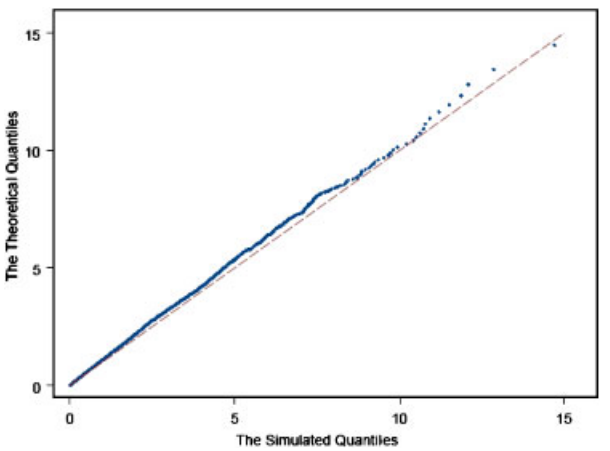

Fig. 2. The QQ plot of the score statistic in the presence of linkage based on 10,000 simulations.

POWER OF SCORE TEST TO DETECT IMPRINTING UNDER ALTERNATIVE HYPOTHESIS

Data sets were simulated similarly to the scenario of linkage but no imprinting in the previous section. However, the genetic contributions from the paternal and maternal alleles with the same DNA composition are different. We first considered complete imprinting in which one parental allele is completely silenced. Let a and $-a$ denote the genetic contributions from paternally derived alleles D and d, respectively; but there is no genetic contribution from maternally derived alleles, i.e., maternal imprinting. The liability variable Z_j^i was defined as $U_e^i + (2n_D - 1)a + e_j^i$, where n_D is the number of paternal allele D. The ordinal response, Y, equaled 0 if $Z<0$, or 1 if $0 \le Z<1$, or 2 if $Z \ge 1$. If we assume equal allele frequencies for both genders and allele D has a frequency of p, the genetic heritability, defined as proportion of phenotype variance explained by the major genetic variance, is $\frac{4p(1-p)a^2}{4p(1-p)a^2+\mathrm{Var}(U_e)+\mathrm{Var}(e)}$. We fixed the total variance at 2, let $\mathrm{Var}(U_e) = \mathrm{Var}(e)$, and let heritability vary from 0.05 to 0.4 at the 0.05 interval. For each set of parameters and two different numbers of pedigrees (100 and 200), we simulated 1,000 datasets. Table II displays the empirical power, defined as the proportion of trials with score statistics larger than $\chi^2_{1,\alpha}$ at the 10th or 11th marker (these are flanking markers for the true locus). A significance level of 0.0025 is the Bonferroni-adjusted significance level for 20 markers.

We also repeated the experiments in the presence of partial imprinting, where alleles from both parents are active but differ in effect depending on origin. The genetic contributions from paternally derived alleles D and d are a and $-a$, respectively; but the genetic contribution from both maternally derived alleles D and d are $a/2$ and $-a/2$. The liability variable Z_j^i was $U_e^i + (2n_D - 1)a + (2m_D - 1)a/2 + e_j^i$, where n_D and m_D are the number of paternal and maternal allele D, respectively. The genetic heritability from both parents, defined as proportion of phenotype variance explained by the major genetic variance, is then $\frac{5p(1-p)a^2}{5p(1-p)a^2+\mathrm{Var}(U_e)+\mathrm{Var}(e)}$. We fixed the total variance at 2, let $\mathrm{Var}(U_e) = \mathrm{Var}(e)$, and let heritability vary from 0.05 to 0.4 at the 0.05 interval.

In Tables I and II, the power of detecting imprinting increases with the increased difference between the contributions from paternal and maternal alleles as we expected. It is interesting that the power of detecting imprinting is very low for low heritability under complete imprinting, but improves greatly under partial imprinting. For both scenarios, the power of detecting linkage increases with the increased total heritability and the gain is faster with less imprinting signal or decreased difference between parental and maternal alleles. With the same heritability, the power of detecting linkage is higher with partial imprinting than with the complete imprinting. This is consistent with previous findings that imprinting can strongly reduce the power to detect linkage when using classic approaches [Hanson et al., 2001; Shete et al., 2003].

APPLICATION TO ALCOHOLISM STUDY

BACKGROUND

COGA aims to identify genes that affect the alcohol dependence and related phenotypes [Edenberg, 2002]. The COGA data set includes 143 families with a total of 1,614 members (including non-genotyped founders). Family size varies from five to 32 members and generations range from two to

TABLE I. Power of detecting linkage and imprinting under complete imprinting

	$N = 100$				$N = 200$			
	Detect imprinting		Detect linkage		Detect imprinting		Detect linkage	
Additive h^2 [a]	$\alpha = 0.5$	$\alpha = .0025$	$\alpha = 0.5$	$\alpha = .0025$	$\alpha = 0.5$	$\alpha = .0025$	$\alpha = 0.5$	$\alpha = .0025$
0.05 (0.05)	0.094	0.005	0.212	0.048	0.107	0.009	0.222	0.049
0.10 (0.10)	0.167	0.016	0.288	0.071	0.273	0.041	0.372	0.120
0.15 (0.15)	0.321	0.061	0.413	0.135	0.517	0.148	0.630	0.289
0.20 (0.20)	0.526	0.146	0.607	0.282	0.797	0.359	0.840	0.559
0.25 (0.25)	0.730	0.296	0.782	0.459	0.959	0.674	0.959	0.799
0.30 (0.30)	0.883	0.520	0.925	0.689	0.995	0.916	0.995	0.955
0.35 (0.35)	0.956	0.756	0.982	0.874	0.999	0.989	1.000	0.994
0.40 (0.40)	0.995	0.903	0.998	0.951	1.000	1.000	1.000	1.000

[a]Difference between contributions from paternal and maternal alleles.

five. Among all individuals, 1,388 members have some discrete and quantitative phenotypes, covariates, and microsatellite genotypes for a 10 cM genome scan. The total number of genotyped microsatellite markers is 315.

Among the multiple alcohol-related phenotypes available, the phenotype ALDX1 is of particular interest. ALDX1 is defined as the severity of the alcohol dependence according to the DSM-III-R [1987] criteria, based on thorough evaluation of various symptoms including craving, binge eating, desire to stop drinking, giving up activities, blackouts due to drinking, physical health problems, and emotional/psychological problems from drinking. Various linkage analyses for alcohol dependence have been performed using the microsatellite genotypes either in the whole genome [Wiener et al., 2005; Williams et al., 2005; Zhu et al., 2005] or on selected chromosomes [Bartlett and Vieland, 2005; Dunn et al., 2005; Martin et al., 2005; Zhao, 2005]. Besides regular linkage evidence without differentiation of parent-of-origin effect [see Wilcox et al., 2005 for summaries], paternal imprinting (i.e., maternal expression) was consistently found at regions linked to alcoholism on chromosomes 10, 12, and 21 as well as maternal imprinting at a locus on chromosome 11 by study groups at Genetic Workshop 14 [Strauch and Baur, 2005]. However, all previous analyses treated ALDX1 as a binary outcome, whereas the original ALDX1 is a four-level ordinal-scaled variable (pure unaffected, never drunk alcohol, unaffected with some symptoms, and affected).

DATA ANALYSIS

Genotyping errors can lead to misleading inferences about inheritance patterns in pedigrees and may greatly affect the results of linkage analysis. Before the analysis we used Merlin [Abecasis et al., 2002] to detect genotype errors and coded them as missing. The error detection method has better accuracy for large pedigrees.

We adjusted our genome-wide analysis for age at interview and sex, which were found significant in predicting the alcohol-related phenotype. The appropriate adjustment was found to increase power of detecting linkage relative to the unadjusted analyses [Doan et al., 2005].

In the first step, we used Merlin to calculate the extended IBD distribution of all the pedigrees. Merlin can not calculate the IBD for nine extended pedigrees because of the memory limits. Thus, we broke each of the 5 largest pedigrees into two smaller pedigrees and removed three–eight uninformative members in the left pedigrees so that the largest bit (a measure of the pedigree complexity, which equals twice the number of non-founders minus the number of founders) of a pedigree is 24. The final dataset for our analyses included 148 families with a total of 1,593 members. All the individuals including those with missing phenotypes or genotypes were included to derive the maximum inheritance information.

In the second step, the individuals with missing response or covariates (age at interview and sex) were excluded. We excluded the 29 subjects who had never drunk alcohol due to the concern that they might not have been exposed to alcohol as were the others. The trait (Y) values are 0, 1, and 2 for purely unaffected, unaffected with some symptoms, and affected, respectively. We fitted the data using our model and performed the score test for imprinting.

RESULTS

Fig. 3 displays the LOD value, defined as $S_{\text{im}}/2\ln 10$, along the genome. The solid curve is the LOD from the score test of linkage and the dashed curve is from the test of imprinting.

TABLE II. Power of detecting linkage and imprinting under partial imprinting

	$N = 100$				$N = 200$			
	Detect imprinting		Detect linkage		Detect imprinting		Detect linkage	
Additive h^2 [a]	$\alpha = 0.5$	$\alpha = .0025$	$\alpha = 0.5$	$\alpha = .0025$	$\alpha = 0.5$	$\alpha = .0025$	$\alpha = 0.5$	$\alpha = .0025$
0.05 (0.03)	0.127	0.011	0.250	0.060	0.207	0.024	0.308	0.088
0.10 (0.06)	0.355	0.067	0.437	0.160	0.564	0.173	0.659	0.314
0.15 (0.09)	0.623	0.222	0.691	0.346	0.875	0.510	0.900	0.678
0.20 (0.12)	0.835	0.445	0.897	0.621	0.995	0.865	0.991	0.918
0.25 (0.15)	0.961	0.720	0.967	0.851	1.000	0.985	1.000	0.989
0.30 (0.18)	0.992	0.902	0.994	0.950	1.000	1.000	1.000	1.000
0.35 (0.21)	1.000	0.972	1.000	0.989	1.000	1.000	1.000	1.000
0.40 (0.24)	1.000	0.993	1.000	0.994	1.000	1.000	1.000	1.000

[a]Difference between contributions from paternal and maternal alleles.

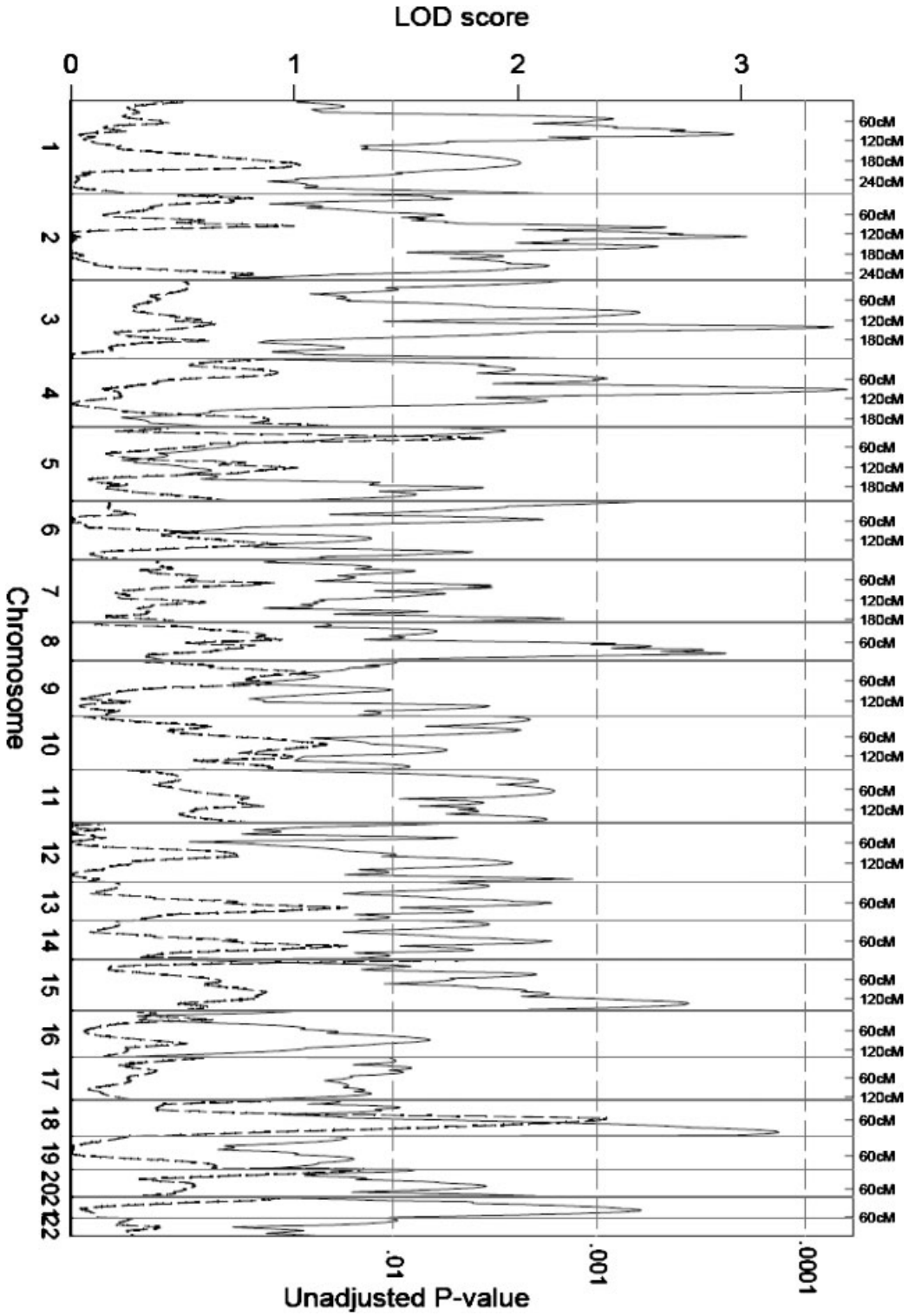

Fig. 3. The LODs on all autosomes. The dashed curve is for test of imprinting and the solid curve is for test of linkage.

The highest peak from the score test for imprinting is near the marker D18A535 on chromosome 18 (LOD = 2.40, position 55 cM, $p = 0.0009$). S_p and S_m components in the S_1 that measure the effects from parental and maternal alleles are 29.615442 and −1.892955, respectively, showing a strong paternal

effect (i.e., maternal imprinting). No study has detected a locus in this region linked to alcoholism, but the signal in this novel region is stronger than previously reported signals in other regions. Several previous studies reported the difference in LOD due to imprinting for alcohol-related phenotypes, the largest difference being 2.06 on chromosome 12 for another phenotype ALDX2 [Bautista et al., 2005].

The score test for linkage identified two regions, near marker ATA34G06 on chromosome 3 (LOD = 3.41, position 141 cM, $p = 0.00007$) and near marker D4S1559 on chromosome 4 (LOD = 3.47, position 92 cM, $p = 0.00006$). Reck et al. [2005] found a significant linkage at ATA34G06 using the SAGE lodpal program [Elston et al., 2002]. Marker D4S1559 is next to marker ADH3, whose linkage to alcoholism has been supported by other human genetic research and experiments on animals [Phillips et al., 1994; Thomasson et al., 1991; Markel et al., 1996; Shen et al., 1997; Reich et al., 1998].

DISCUSSION

We proposed a latent variable, proportional odds model to differentiate the parent-of-origin effect in the linkage analysis for ordinal traits. We developed a score statistic to take advantages of the IBD sharing information between relatives, which greatly alleviates the computational burden for large pedigrees. Our method was implemented in a C program and will be available from http://peace.med.yale.edu/pub/LOT. Because binary traits are only special ordinal traits with two categories, our method can be readily applied for binary traits.

Using the proposed model, we scanned the microsatellite markers on 22 autosomes for alcohol dependence in the COGA dataset. Our score test suggests evidence of maternal imprinting at a novel region near D18A535 on chromosome 18.

Our method depends on and hence is limited by existing methods or software such as Merlin and Simwalk2 [Sobel and Lange, 1996] to compute the parent-of-origin-specific IBD. Simwalk can calculate IBD for large pedigrees using the Markov chain Monte Carlo algorithm but requires intensive computation time. For example, it may take months to complete the multipoint IBD computation according to the previous experiments [Kim et al., 2005; Wilcox et al., 2005]. Therefore, we used Merlin for the parent-of-origin- -specific allele-sharing IBD calculation even though dividing or trimming large pedigrees may lose some potentially useful inheritance information. Further investigation is warranted upon the availability of updated algorithm and software for calculating parent-of-origin specific IBD probabilities.

ACKNOWLEDGMENTS

Data were provided by the Collaborative Study on the Genetics of Alcoholism (U10AA008401). The authors thank Drs. Raymond Crowe and Jean MacCluer for facilitating the use of the COGA data. The authors would also like to thank Ms. Judith Dickson for reviewing the manuscript.

REFERENCES

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

Adams S, Vinkenoog R, Spielman M, Dickinson HG, Scott RJ. 2000. Parent-of-origin effects on seed development in Arabidopsis thaliana require DNA methylation. Development 127:2493–2502.

Bartlett CW, Vieland VJ. 2005. Two novel quantitative trait linkage analysis statistics based on the posterior probability of linkage: application to the COGA families. BMC Genet 6(Suppl 1):S121.

Bassett SS, Avramopoulos D, Perry RT, Wiener H, Watson B Jr, Go RC, Fallin MD. 2006. Further evidence of a maternal parent-of-origin effect on chromosome 10 in late-onset Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 141:537–540.

Bautista JF, Quade SR, Parrado AR, Goddard KA. 2005. Linkage analysis of alcoholism-related electrophysiological phenotypes: genome scans with microsatellites compared to single-nucleotide polymorphisms. BMC Genet 6(Suppl 1):S156.

Casella G, Berger RL. 2002. Statistical inference. Australia, Pacific Grove, CA: Thomson Learning.

Cox R, Hinkley DV. 1974. Theoretical statistics. London: Chapman & Hall

Doan BQ, Frangakis CE, Shugart YY, Bailey-Wilson JE. 2005. Application of the propensity score in a covariate-based linkage analysis of the collaborative study on the genetics of alcoholism. BMC Genet 6(Suppl 1):S33.

Dunn G, Hinrichs AL, Bertelsen S, Jin CH, Kauwe JS, Suarez BK, Bierut LJ. 2005. Microsatellites versus single-nucleotide polymorphisms in linkage analysis for quantitative and qualitative measures. BMC Genet 6(Suppl 1):S122.

Edenberg HJ. 2002. The collaborative study on the genetics of alcoholism: an update. Alcohol Res Health 26:214–218.

Elston R, Bailey-Wilson J, Bonney G, Tran L, Keats B, Wilson A. 2002. Statistical analysis for genetic epidemiology (S.A.G.E.). Version 5.0: Cleveland, OH: Rammelkamp Center for Education and Research, Metro Health Campus.

Everman DB, Cassidy SB. 2000. Genetics of childhood disorders: XII. Genomic imprinting: breaking the rules. J Am Acad Child Adolesc Psychiatry 39:386–389.

Feng R, Leckman JF, Zhang H. 2004. Linkage analysis of ordinal traits for pedigree data. Proc Natl Acad Sci USA 101: 16739–16744.

Feng R, Zhang H. 2007. A score test for linkage analysis of ordinal traits based on IBD sharing. Biostatistics. Epub, May 22, 2007, doi: 10.1093/biostatistics/kxm016.

Francks C, DeLisi LE, Shaw SH, Fisher SE, Richardson AJ, Stein JF, Monaco AP. 2003. Parent-of-origin effects on handedness and schizophrenia susceptibility on chromosome 2p12-q11. Hum Mol Genet 12:3225–3230.

Hanson RL, Kobes S, Lindsay RS, Knowler WC. 2001. Assessment of parent-of-origin effects in linkage analysis of quantitative traits. Am J Hum Genet 68:951–962.

Huxtable SJ, Saker PJ, Haddad L, Walker M, Frayling TM, Levy JC, Hitman GA, O'Rahilly S, Hattersley AT, McCarthy MI. 2000. Analysis of parent-offspring trios provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally transmitted class III variable number tandem repeat alleles. Diabetes 49:126–130.

Karason A, Gudjonsson JE, Upmanyu R, Antonsdottir AA, Hauksson VB, Runasdottir EH, Jonsson HH, Gudbjartsson DF, Frigge ML, Kong A, Stefansson K, Valdimarsson H, Gulcher JR. 2003. A susceptibility gene for psoriatic arthritis maps to chromosome 16q: evidence for imprinting. Am J Hum Genet 72:125–131.

Kim H, Hutter CM, Monks SA, Edwards KL. 2005. Comparison of single-nucleotide polymorphisms and microsatellites in detecting quantitative trait loci for alcoholism: the collaborative study on the genetics of alcoholism. BMC Genet 6(Suppl 1):S5.

Knapp M, Strauch K. 2004. Affected-sib-pair test for linkage based on constraints for identical-by-descent distributions corresponding to disease models with imprinting. Genet Epidemiol 26:273–285.

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

Lander ES, Green P. 1987. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367.

Luedi PP, Hartemink AJ, Jirtle RL. 2005. Genome-wide prediction of imprinted murine genes. Genome Res 15:875–884.

Markel PD, Fulker DW, Bennett B, Corley RP, DeFries JC, Erwin VG, Johnson TE. 1996. Quantitative trait loci for ethanol sensitivity in the LS x SS recombinant inbred strains interval mapping. Behav Genet 26:447–458.

Martin LJ, Avery CL, Williams JT, North KE. 2005. Quantitative trait locus-specific genotype x alcoholism interaction on linkage for evoked electroencephalogram oscillations. BMC Genet 6(Suppl 1):S123.

McCullagh P, Nelder JA. 1989. Generalized linear models. London, New York: Chapman and Hall.

Phillips TJ, Crabbe JC, Metten P, Belknap JK. 1994. Localization of genes affecting alcohol drinking in mice. Alcohol Clin Exp Res 18:931–941.

Reck BH, Mukhopadhyay N, Tsai HJ, Weeks DE. 2005. Analysis of alcohol dependence phenotype in the COGA families using covariates to detect linkage. BMC Genet 6(Suppl 1):S143.

Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, Porjesz B, Li TK, Conneally PM, Nurnberger JI Jr, Tischfield JA, Crowe RR, Cloninger CR, Wu W, Shears S, Carr K, Crose C, Willig C, Begleiter H. 1998. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet 81:207–215.

Rump P, Zeegers MP, van Essen AJ. 2005. Tumor risk in Beckwith-Wiedemann syndrome: a review and meta-analysis. Am J Med Genet A 136:95–104.

Shen YC, Fan JH, Edenberg HJ, Li TK, Cui YH, Wang YF, Tian CH, Zhou CF, Zhou RL, Wang J, Zhao ZL, Xia GY. 1997. Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism. Alcohol Clin Exp Res 21:1272–1277.

Shete S, Zhou X, Amos CI. 2003. Genomic imprinting and linkage test for quantitative-trait Loci in extended pedigrees. Am J Hum Genet 73:933–938.

Sobel E, Lange K. 1996. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337.

Speer MC, Terwilliger JD, Ott J. 1995. Data simulation for GAW9 problems 1 and 2. Genet Epidemiol 12:561–564.

Steinke JW, Borish L, Rosenwasser LJ. 2003. Genetics of hypersensitivity. J Allergy Clin Immunol 111(Suppl 2):S495–S501.

Strauch K, Baur MP. 2005. Parent-of-origin, imprinting, mitochondrial, and x-linked effects in traits related to alcohol dependence: presentation Group 18 of Genetic Analysis Workshop 14. Genet Epidemiol 29(Suppl 1):S125–S132.

Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP. 2000. Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. Am J Hum Genet 66:1945–1957.

Temple IK, Shield JP. 2002. Transient neonatal diabetes, a disorder of imprinting. J Med Genet 39:872–875.

Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK, Wang SP, Lin YT, Lu RB, Yin SJ. 1991. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. Am J Hum Genet 48:677–681.

Vincent Q, Alcais A, Alter A, Schurr E, Abel L. 2006. Quantifying genomic imprinting in the presence of linkage. Biometrics 62:1071–1080.

Weinberg CR. 1999. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. Am J Hum Genet 65:229–235.

Weinberg CR, Wilcox AJ, Lie RT. 1998. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 62:969–978.

Whittaker JC, Gharani N, Hindmarsh P, McCarthy MI. 2003. Estimation and testing of parent-of-origin effects for quantitative traits. Am J Hum Genet 72:1035–1039.

Wiener HW, Go RC, Tiwari H, George V, Page GP. 2005. COGA phenotypes and linkages on chromosome 2. BMC Genet 6(Suppl 1):S125.

Wijsman EM, Amos CI. 1997. Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. Genet Epidemiol 14:719–735.

Wilcox MA, Pugh EW, Zhang H, Zhong X, Levinson DF, Kennedy GC, Wijsman EM. 2005. Comparison of single-nucleotide polymorphisms and microsatellite markers for linkage analysis in the COGA and simulated data sets for Genetic Analysis Workshop 14: Presentation Groups 1, 2, and 3. Genet Epidemiol 29(Suppl 1):S7–S28.

Williams AH, Brown WM, Langefeld CD. 2005. Multilocus and interaction-based genome scan for alcoholism risk factors in Caucasian Americans: the COGA study. BMC Genet 6(Suppl 1):S37.

Zhao JH. 2005. Mixed-effects Cox models of alcohol dependence in extended families. BMC Genet 6(Suppl 1):S127.

Zhu X, Cooper R, Kan D, Cao G, Wu X. 2005. A genome-wide linkage and association study using COGA data. BMC Genet 6(Suppl 1):S128.

APPENDIX

Let $\delta_p = \text{sign}(\gamma_p)\gamma_p^2$ and $\delta_m = \text{sign}(\gamma_m)\gamma_m^2$ in (1). Then $\gamma_p = \text{sign}(\delta_p)\sqrt{|\delta_p|}$, $\gamma_m = \text{sign}(\delta_m)\sqrt{|\delta_m|}$ and testing the hypothesis $\gamma_p = \gamma_m = 0$ in (1) is equivalent to testing $\delta_p = \delta_m = 0$ in the following model:

$$\begin{aligned} &\log\text{it}(P\{Y_j^i \le k|U^i, v^i\}) = \boldsymbol{x}_j^i\boldsymbol{\beta} + \alpha_k + \text{sign}(\delta_p)\sqrt{|\delta_p|}U_{gp,j}^i + \text{sign}(\delta_m)\sqrt{|\delta_m|}U_{gm,j}^i + \gamma_e U_e^i, \\ &\quad \text{for } i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \quad k = 0, 1, \ldots, K-1. \end{aligned} \tag{5}$$

The log-likelihood function for the ith family is the following function of ω, γ_p and γ_m,

$$l^i(\omega, \gamma_p, \gamma_p) = \log \int\int \sum_{v^i \in V} P^i(\omega, \gamma_p, \gamma_m) p(v^i) \prod_{j=1}^{2f_i} dG(U_{g,j}^i) dF(U_e^i),$$

where V is the set of all possible inheritance vectors for the ith family, $F(\cdot)$ and $G(\cdot)$ are the cumulative distribution function of U_g and U_e,

$$P^i(\omega, \gamma_p, \gamma_m) = \prod_{j=1}^{n_i} I(Y_j^i = k)\left[\pi_{k,j}^i(\omega, \gamma_p, \gamma_m) - \pi_{k-1,j}^i(\omega, \gamma_p, \gamma_m)\right],$$

$$\pi_{k,j}^i(\omega, \gamma_p, \gamma_m) = \frac{\exp(\boldsymbol{x}_j^i\boldsymbol{\beta} + \alpha_k + \gamma_p U_{gp,j}^i + \gamma_m U_{gm,j}^i + \gamma_e U_e^i)}{1 + \exp(\boldsymbol{x}_j^i\boldsymbol{\beta} + \alpha_k + \gamma_p U_{gp,j}^i + \gamma_m U_{gm,j}^i + \gamma_e U_e^i)} \quad \text{for} \quad k = 0, 1, \ldots, K-1,$$

$$\pi_{-1,j}^i(\omega, \gamma_p, \gamma_m) = 0 \quad \text{and } \pi_{K,j}^i(\omega, \gamma_p, \gamma_m) = 1.$$

Let $\eta_j^i = sign(\delta_p)\sqrt{|\delta_p|}U_{gp,j}^i$and $\eta^i = (\eta_1^i, \ldots, \eta_{n_i}^i)$. If we assume the contributions from the paternal and maternal alleles are independent, a straightforward calculation can show that

$$\lim_{\delta_p \to 0} \int \frac{\partial P^i(\omega, \eta^i, \gamma_m)}{\partial \eta_j^i} U_{gp,j}^i dG(U_g^i) = \lim_{\delta_p \to 0} \frac{\partial P^i(\omega, \eta^i, \gamma_m)}{\partial \eta_j^i} E(U_{gp,j}^i) = 0 \tag{6}$$

and that

$$\begin{aligned} &\lim_{\delta_p \to 0} \frac{\partial \int P^i(\omega, \delta_p, \delta_m = 0) dG(U_g^i)}{\partial \delta_p} = \lim_{\delta_p \to 0} \sum_{j=1}^{n_i} \int \frac{\partial P^i(\omega, \eta^i, 0)}{\partial \eta_j^i} \frac{\partial \eta_j^i}{\partial \delta_p} dG(U_g^i) \\ &= \sum_{j=1}^{n_i} \lim_{\delta_p \to 0} \frac{\int \frac{\partial P^i(\omega, \eta^i, \mathbf{0})}{\partial \eta_j^i} U_{gp,j}^i dG(U_g^i)}{2\sqrt{|\delta_p|}} \\ &= \sum_{j=1}^{n_i} \lim_{\delta_p \to 0} \frac{\int \frac{\partial P^i(\omega, \eta^i, \mathbf{0})}{\partial \eta_j^i \partial \delta_p} U_{gp,j}^i dG(U_g^i)}{2\frac{\partial \sqrt{|\delta_p|}}{\partial \delta_p}} \qquad \text{[by (5) and the L'Hôpital's rule]} \\ &= \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \lim_{\delta_p \to 0} \frac{\int \sqrt{|\delta_p|} \frac{\partial P^i(\omega, \eta^i, \mathbf{0})}{\partial \eta_j^i \partial \eta_l^i} U_{gp,j}^i U_{gp,l}^i dG(U_g^i)}{2\sqrt{|\delta_p|}} = \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \frac{\partial^2 P^i(\omega, 0, 0)}{\partial \eta_j^i \partial \eta_l^i} \frac{E(U_{gp,j}^i U_{gp,l}^i)}{2}. \end{aligned} \tag{7}$$

Similarly, let $\xi_j^i = \text{sign}(\delta_m)\sqrt{|\delta_m|}U_{gm,j}^i$, we can have

$$\lim_{\delta_m \to 0} \frac{\partial \int P^i(\omega, \delta_p = 0, \delta_m) dG(U_g^i)}{\partial \delta_m} = \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \frac{\partial^2 P^i(\omega, 0, 0)}{\partial \xi_j^i \partial \xi_l^i} \frac{E(U_{gm,j}^i U_{gm,l}^i)}{2}.$$

.

Also, for non-inbred individuals j and l within the ith pedigree, we have

$$\begin{aligned}&\sum_{v\in V} E(U^i_{gp,j}U^i_{gp,l})p(v) = \sum_{v\in V}\mathrm{cov}(U^i_{gp,j},U^i_{gp,l})p(v) = \sum_{v\in V}\mathrm{cov}(U^i_{g,v_{2j-1}},U^i_{g,v_{2l-1}})p(v)\\ &= \sum_{v\in V} I(v^i_{2j-1}=v^i_{2l-1})p(v) = q^i_{p,jl} \quad \text{and} \quad \sum_{v\in V} E(U^i_{gm,j}U^i_{gm,l})p(v) = q^i_{m,jl},\end{aligned} \tag{8}$$

where $q^i_{p,jl}$ and $q^i_{m,jl}$ is the number of paternal alleles and maternal alleles shared by the jth and lth individuals at the testing locus, respectively.
Thus, the score vector for testing the hypothesis concerning δ_p and δ_m is as follows:

$$\begin{aligned}S = \begin{pmatrix}S_p\\ S_m\end{pmatrix} &= \begin{pmatrix}\sum_{i=1}^N \lim_{\delta_p\to 0}\frac{\partial l^i(\omega,\delta_p,\mathbf{0})}{\partial\delta_p}\\ \sum_{i=1}^N \lim_{\delta_m\to 0}\frac{\partial l^i(\omega,\mathbf{0},\delta_m)}{\partial\delta_m}\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^N \frac{\sum_{v\in V}\int\int \lim_{\delta_p\to 0}\frac{\partial P^i(\omega,\delta_p,0)}{\partial\delta_p}dG(U^i_g)dF(U^i_e)p(v)}{\int P^i(\omega,\mathbf{0})dF(U^i_e)}\\ \sum_{i=1}^N \frac{\sum_{v\in V}\int\int \lim_{\delta_m\to 0}\frac{\partial P^i(\omega,\mathbf{0},\delta_m)}{\partial\delta_m}dG(U^i_g)dF(U^i_e)p(v)}{\int P^i(\omega,\mathbf{0})dF(U^i_e)}\end{pmatrix}\\ &= \begin{pmatrix}\sum_{i=1}^N\sum_{j=1}^{n_i}\sum_{l=1}^{n_i}\frac{\int\frac{\partial^2 P^i(\omega,\mathbf{0},\mathbf{0})}{\partial\eta^i_j\partial\eta^i_l}dF(U^i_e)\sum_{v\in V}\frac{E(U^i_{gp,j}U^i_{gp,l})}{2}p(v)}{\int P^i(\omega,\mathbf{0})dF(U^i_e)}\\ \sum_{i=1}^N\sum_{j=1}^{n_i}\sum_{l=1}^{n_i}\frac{\int\frac{\partial^2 P^i(\omega,\mathbf{0},\mathbf{0})}{\partial\xi^i_j\partial\xi^i_l}dF(U^i_e)\sum_{v\in V}\frac{E(U^i_{gm,j}U^i_{gm,l})}{2}p(v)}{\int P^i(\omega,\mathbf{0})dF(U^i_e)}\end{pmatrix} \quad \text{[by (7)]}\\ &= \begin{pmatrix}\sum_{i=1}^N\sum_{j=1}^{n_i}\sum_{l=1}^{n_i}\frac{\int\frac{\partial^2 P^i(\omega,\mathbf{0})}{\partial\eta^i_j\partial\eta^i_l}P^i(\omega,\mathbf{0})dF(U^i_e)q^i_{p,jl}}{2\int P^i(\omega,\mathbf{0})dF(U^i_e)}\\ \sum_{i=1}^N\sum_{j=1}^{n_i}\sum_{l=1}^{n_i}\frac{\int\frac{\partial^2 P^i(\omega,\mathbf{0})}{\partial\xi^i_j\partial\xi^i_l}P^i(\omega,\mathbf{0})dF(U^i_e)q^i_{m,jl}}{2\int P^i(\omega,\mathbf{0})dF(U^i_e)}\end{pmatrix}. \quad \text{[by (8)]}\end{aligned} \tag{9}$$

Furthermore, we have

$$\frac{\partial^2 P^i(\omega,\mathbf{0})}{\partial\eta^i_j\partial\eta^i_l} = \frac{\partial^2 P^i(\omega,\mathbf{0})}{\partial\xi^i_j\partial\xi^i_l} = P^i(\omega,\mathbf{0})\left[\frac{\partial\log P^i(\omega,\mathbf{0})}{\partial\eta^i_j}\frac{\partial\log P^i(\omega,\mathbf{0})}{\partial\eta^i_l} + \frac{\partial^2\log P^i(\omega,\mathbf{0})}{\partial\eta^i_j\partial\eta^i_l}\right], \tag{10}$$

where

$$\begin{aligned}&\frac{\partial\log P^i(\omega,\mathbf{0})}{\partial\eta^i_j}\Big|_{\delta_p=\delta_m=0} = I(Y^i_j=r)(1-\pi^i_{r,j}(\omega,\mathbf{0})-\pi^i_{r-1,j}(\omega,0)),\\ &\frac{\partial^2\log P^i(\omega,\mathbf{0})}{\partial\eta^i_j\partial\eta^i_j}\Big|_{\delta_p=\delta_m=0} = -I(Y^i_j=r)[\pi^i_{r,j}(\omega,0)(1-\pi^i_{r,j}(\omega,0))+\pi^i_{r-1,j}(\omega,0)(1-\pi^i_{r-1,j}(\omega,0))],\\ &\frac{\partial^2\log P^i(\omega,\mathbf{0})}{\partial\eta^i_j\partial\eta^i_l}\Big|_{\delta_p=\delta_m=0} = 0, \quad \text{for} \quad j\neq l.\end{aligned}$$

Substituting $\partial^2 P^i(\omega,\mathbf{0})/\partial\eta^i_j\partial\eta^i_l$ in (9) with (10), we have the score vector as expressed in (2). The important ingredients, numbers of paternal and maternal alleles sharing IBD between two relatives, can be obtained through the genetic marker information.