Behavior Genetics, Vol. 29, No. 4, 1999

Distinguishing Population Stratification from Genuine Allelic Effects with Mx: Association of ADH2 with Alcohol Consumption

M. C. Neale,[1] S. S. Cherny,[2,3] P. C. Sham,[3] J. B. Whitfield,[4] A. C. Heath,[5] A. J. Birley,[6] and N. G. Martin[6]

Received 25 Nov. 1998—Final 24 June 1999

A universal problem in genetic association studies is to distinguish associations due to genuine effects of the locus under investigation, or linkage disequilibrium with a nearby locus that has a genuine effect, from associations due to population stratification or other artifacts. Fulker *et al.* (1999) have suggested a test using unselected sib pairs to distinguish these two causes of association. The test is readily implemented within a standard maximum-likelihood framework using the Mx package. The approach is applied to data on *ADH2* genotypes and a measure of alcohol consumption from an Australian DZ twin pair sample. Results indicate that the association of the *ADH2*2* allele with lower alcohol consumption cannot be explained by simple admixture and that there may be genuine allelic effects of the locus on alcohol consumption. Power calculations are provided to show that these results are plausible for the sample size in this study and consider the effects of genetic architecture and sample structure on required sample sizes for the Fulker *et al.* test.

KEY WORDS: Linkage; linkage disequilibrium; genetic association; admixture; population stratification; ADH2; alcohol consumption; Mx; statistical model; identical by descent (IBD); twins.

INTRODUCTION

Linkage and Association

In order to detect and locate genes affecting a phenotype, two broad strategies are commonly employed, linkage and association. Both strategies depend on having a linkage map, usually of phenotypically neutral markers, and both exploit the cosegregation of linked genes. Linkage analysis (Elston, 1998) relies on the fact that relatively widely spaced markers will be cotransmitted within a few generations. Suitable samples for linkage analysis therefore range from pairs of relatives to pedigrees containing several generations. Linkage analysis requires relatively few markers to conduct a genome scan for quantitative trait loci (QTL); 200–400 are usually considered sufficient in the case of humans. The main disadvantages of linkage analysis are that it yields imprecise location information and that, in certain circumstances, it has substantially less statistical power than tests of association.

The association study relies upon a within population association between the alleles at a specific locus and the phenotype under study. Only very closely linked genes may be expected to cosegregate in populations (i.e., be in linkage disequilibrium with the locus), so the location information is very precise. Widely spaced loci, such as those used in linkage studies, would be expected to be in linkage equilibrium with a QTL in most

[1] Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Box 980126, Richmond, Virginia 23298-0126.
[2] Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado 80309-0447.
[3] Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, London, England.
[4] Royal Prince Alfred Hospital, Sydney, Australia.
[5] Department of Psychiatry, Washington University, St. Louis, Missouri 63110.
[6] Queensland Institute of Medical Research, Brisbane, Australia.

0001-8244/99/0700-0233$16.00/0

populations and would therefore show no population association. Closely linked genes, perhaps less than a centimorgan (cM) apart, may be in disequilibrium and may show population associations over a large number of generations (Lander and Botstein, 1989; Risch and Merikangas, 1996; Sham, 1997). We define an "allelic effect" as a direct effect of allele substitution on the expected value of a phenotype, which is of most interest in the search for trait-relevant loci in the genome.

The statistical detection of association is relatively straightforward; at its simplest level it involves looking at mean differences in phenotype between groups with different marker alleles. In general, statistical tests based on means are more powerful than those based on higher-order moments such as covariances (Cohen, 1969), so tests of association are usually more powerful than tests of linkage. In the simplest possible case, where a marker happens to be a polymorphism within a functional gene, association analysis is simply classical Mendelian analysis, although candidate genes may also be assessed through the methods of within-family linkage. A major drawback to association testing is that thousands of markers are needed to conduct a genome scan. This in turn causes the problem of high type I error rates (Risch and Teng, 1998). Currently, association analysis is popular for the purposes of examining candidate loci, for which type I error rates are more easily controlled.

Risch and Merikangas (1996) compared the statistical properties of linkage and association mapping. Both continuous and discrete traits were considered for random samples of sib pairs in a linkage study and case–control samples in an association study. They concluded that, given a suitably dense marker map, association mapping would be the preferred procedure for detecting QTLs in genome scans.

Despite its lower statistical power, linkage analysis does have the advantage that positive results are unlikely to arise from artifacts other than type I error. In contrast, association may appear in a population for many reasons other than a genuine effect of the locus itself or one very closely linked to it. For example, population associations among genes will occur after recent admixture of two or more subpopulations. Indeed, initially, when two populations are considered as one, genes on different chromosomes may be associated due to strong disequilibrium. Only after many generations of random mating within and between the subpopulations will loosely linked genes come into Hardy–Weinberg equilibrium and appear unassociated, leaving only closely linked loci in disequilibrium (Sham, 1997). Furthermore, when both phenotype and genotype frequencies differ between subpopulations, a genotype–phenotype association will occur even if there is no effect of genotype on phenotype whatsoever, i.e., a heritability of zero.

The ideal situation for association mapping is where a single mutant allele predisposing to disease arises in an isolated population and is subjected to many generations of random mating. Under this model, only very closely linked markers will remain associated with the disease gene and detection and location may be powerful and accurate. For such populations and traits, classical genetic analysis is straightforward if individuals can be unambiguously genotyped. The mean effect of a genotype on a trait can be measured in a population and its contribution to disease liability may be readily assessed.

In the case of complex traits which may be influenced by many genes, the causes of disequilibrium may include population admixture and interfamilial heterogeneity. One way to address problems of this sort is to consider association within families. The haplotype relative risk (HRR), transmission disequilibrium (TDT), and sib TDT are examples of tests that statistically control for population stratification by measuring the probability that an allele is transmitted from a parent to an affected offspring (Allison, 1997; Ewens and Spielman, 1998; Spielman and Ewens, 1996). Allison's test involves continuous phenotypes but requires genotypic information from the parents. Fulker *et al.* (1999) described a combined linkage and association test for quantitative traits that does not require data from parents. In this paper we describe the implementation of this test using Mx software (Neale, 1997) and illustrate it with data on alcohol dehydrogenase genotypes (*ADH2*) and alcohol consumption.

METHODS

Likelihood Method

The likelihood method (Edwards, 1972; Fisher, 1925) has a number of statistical advantages. Parameter estimates are asymptotically unbiased and have the minimum variance of all such estimates. Confidence intervals may be obtained on parameter estimates by examining the change in likelihood as a parameter is varied from its maximum-likelihood estimate, a procedure which is automated in the Mx program (Neale and Miller, 1997). Similarly, the difference in support between a model and a submodel may be assessed using a likelihood-ratio test. In general, denoting the

log-likelihood maximized with respect to a vector θ of t parameters by $\ln L_1$, and the log-likelihood with a subset of $j \leq t$ parameters fixed at hypothesized values by $\ln L_0$, the statistic

$$T = 2(\ln L_1 - \ln L_0) \quad (1)$$

is asympotically distributed as χ^2 with j degrees of freedom.

Likelihood methods demand a parametric model for the data. The usual assumption for random variables is that the data are sampled from a multivariate normal distribution. Any large allelic effect will disrupt such normality. Therefore, we seek a method which is multivariate normal *conditional* on the particular alleles that an individual has at the locus in question. In addition, other covariates may be included that further improve the multivariate normality of the conditional distribution. This much weaker assumption of conditional normality (normality of the residual variation given the genotype and covariates) is more likely to be satisfied in practice.

The log-likelihood of a column vector of sib-pair trait scores, $\mathbf{x} = (x_1, x_2)'$, may be written

$$\ln L = C - .5\left\{\ln|\Sigma_i| + (\mathbf{x} - \boldsymbol{\mu}_i)'\Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right\} \quad (2)$$

where Σ_i is the predicted covariance matrix of sib pair i, $|\Sigma_i|$ is the determinant of Σ_i, the prime denotes transpose, and C is a constant term. The predicted mean vector of the sib pair, $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})'$, is a function of the following parameters: the population mean m, the pair mean s_i, the pair difference d_i, and the regression weights $\boldsymbol{\beta}$ of the phenotypes on a vector of covariates $\mathbf{k}_{ij}$ measured on sibling j in pair i.

$$\mu_{i1} = m + s_i + \frac{d_i}{2} + \boldsymbol{\beta}\mathbf{k}_{i1} \quad (3)$$

$$\mu_{i2} = m + s_i - \frac{d_i}{2} + \boldsymbol{\beta}\mathbf{k}_{i2} \quad (4)$$

It is especially important to recognize that the predicted pair means and differences, s_i and d_i, depend on the observed genotypes of the specific pair in question. Also, the pair mean will be a function of observed covariates $\mathbf{k}_{ij}$. Any model of Σ_i will be of residual variation not accounted for by the measured genotype and the covariates. We now consider models for s_i, d_i, and Σ_i.

Allelic Effects

The theory behind the joint linkage and association test is described in detail by Fulker *et al.* (1999). Here we restate the principles in brief. Consider an additive QTL with alleles A_1 and A_2, which occur at frequencies p and q. Let the effects of the three genotypes A_2A_2, A_1A_2, and A_1A_1 be $-a$, 0, and a, respectively. Under this model the nine possible combinations of sibling pair genotypes have the pair means and pair differences shown in Table I. Their predicted frequencies in a random mating population are also shown in Table I.

Test of Association

To test for association it is necessary to examine whether the means of siblings vary as a function of their genotype at a candidate locus. This test is prone to spurious associations due to population stratification. However, family members such as full sib pairs originate from the same stratum of the population; in terms of

Table I. Expected Sib-Pair Means and Differences and Their Frequencies for a Single Additive Two-Allele Locus

Genotype		Additive effects				
Sib 1	Sib 2	Sib 1	Sib 2	Mean	Difference/2	Frequency
A_1A_1	A_1A_1	a	a	a	0	$p^4 + p^3q + (p^2q^2/4)$
A_1A_1	A_1A_2	a	0	$a/2$	$a/2$	$p^3q + (p^2q^2/2)$
A_1A_1	A_2A_2	a	$-a$	0	a	$p^2q^2/4$
A_1A_2	A_1A_1	0	a	$a/2$	$-a/2$	$p^3q + (p^2q^2/2)$
A_1A_2	A_1A_2	0	0	0	0	$p^3q + 3p^2q^2 + pq^3$
A_1A_2	A_2A_2	0	$-a$	$-a/2$	$a/2$	$(p^2q^2/2) + pq^3$
A_2A_2	A_1A_1	$-a$	a	0	$-a$	$p^2q^2/4$
A_2A_2	A_1A_2	$-a$	0	$-a/2$	$-a/2$	$(p^2q^2/2) + pq^3$
A_2A_2	A_2A_2	$-a$	$-a$	$-a$	0	$(p^2q^2/4) + pq^2 + q^4$

the classical twin study, population stratification acts like a shared environmental factor. With data from relatives such as sib pairs, the allelic effect a may be partitioned into effects between families, a_b, and effects within families, a_w (Mather and Jinks, 1982). The between-family component a_b affects only pair means, while the within-family component a_w affects only intrapair differences. For a pair of sibs with pair mean s and pair difference d, the predicted means of the sibs are simply $\mu_1 = s + (d/2)$ for Sib 1 and $\mu_2 = s - (d/2)$ for Sib 2. The pair means, pair differences, and means for Sib 1 and Sib 2 for each of the nine possible combinations of sib pair genotypes are listed in Table II.

Any genuine allelic effect will contribute to both within- and between-pair components, and to the same extent. Thus a test for stratification is to compute T between the model with separate free parameters a_b and a_w and the model with $a_b = a_w$.

A robust test for association may be obtained by computing T between a model with a_w free and a model with a_w set to 0, while a_b is free in both models. This test is free of the effects of population stratification. An Mx script that fits this model is given in Appendix 1.

An alternative test is obtained by relaxing the constraints on pair means in Table II, so that there are five distinct a_b deviations relative to the A_2A_2,A_2A_2 pair mean, as shown in Table III. The statistical test for a single value of a_b has 5 degrees of freedom, comparing the fit of the model with a_1 to a_5 free with the fit of one in which all five deviations are fixed to zero. While theoretically having lower power,[7] the advantage of this test over the more restrictive equal a_b test is that nonadditive allelic interactions (dominance) are not assumed to be absent.

Technically, the robust tests of association presented here are tests of association and linkage, because they require linkage disequilibrium between the locus being tested and the disease locus. The test detects $\delta(1 - 2\theta)$, where δ is the disequilibrium coefficient and θ is the recombination fraction. If $\delta = 0$ is zero, there is no association and no effect will be detected. Similarly, if $\theta = .5$ there is no linkage and no effect will be detected. Essentially, $\theta < .5$ and $\delta > 0$ are tested simultaneously. This in contrast to nonrobust tests such as those conducted with unrelated individuals, which are subject to the effects of population stratification and which may detect association in the absence of linkage.

Test of Linkage via Identical by Descent (IBD) Status

It is possible to test for what we term "IBD linkage" via measured probabilities that relatives share zero, one, or two, alleles IBD at a putative disease locus. In the model described above, Σ_i is the covariance matrix of sib pair i, which, for a test of IBD linkage, may be parameterized as

$$\Sigma_i = \begin{pmatrix} \sigma_q^2 + \sigma_r^2 + \sigma_e^2 & \hat{\pi}_i\sigma_q^2 + \sigma_r^2 \\ \hat{\pi}_i\sigma_q^2 + \sigma_r^2 & \sigma_q^2 + \sigma_r^2 + \sigma_e^2 \end{pmatrix} \quad (5)$$

where σ_q^2 is the estimated variance of a QTL detected via IBD linkage, σ_r^2 is an additional component of variance contributing to sibling resemblance over and above that which is due to the QTL, and σ_e^2 is nonshared

[7] Except for the biologically unusual case of extreme overdominance (Georges and Cockett, 1996).

Table II. Partitioning the Additive Effect of a Locus into Between (a_b)- and Within (a_w)-Pair Components

Genotype		Contribution of locus		Predicted mean	
Sib 1	Sib 2	Mean	Difference/2	Sib 1	Sib 2
A_1A_1	A_1A_1	a_b	0	a_b	a_b
A_1A_1	A_1A_2	$a_b/2$	$a_w/2$	$(a_b/2) + (a_w/2)$	$(a_b/2) - (a_w/2)$
A_1A_1	A_2A_2	0	a_w	a_w	$-a_w$
A_1A_2	A_1A_1	$a_b/2$	$-a_w/2$	$(a_b/2) - (a_w/2)$	$(a_b/2) + (a_w/2)$
A_1A_2	A_1A_2	0	0	0	0
A_1A_2	A_2A_2	$-a_b/2$	$a_w/2$	$(-a_b/2) + (a_w/2)$	$(-a_b/2) - (a_w/2)$
A_2A_2	A_1A_1	0	$-a_w$	$-a_w$	a_w
A_2A_2	A_1A_2	$-a_b/2$	$-a_w/2$	$(-a_b/2) - (a_w/2)$	$(-a_b/2) + (a_w/2)$
A_2A_2	A_2A_2	$-a_b$	0	$-a_b$	$-a_b$

Table III. Partitioning Additive Effect into Between- and Within-Pair Components: Saturated Between-Pair Component

Genotype					
Sib 1	Sib 2	Mean	Difference/2	Sib 1	Sib 2
A_1A_1	A_1A_1	a_1	0	a_1	a_1
A_1A_1	A_1A_2	a_2	$a_w/2$	$a_2 + (a_w/2)$	$a_2 - (a_w/2)$
A_1A_1	A_2A_2	a_3	a_w	$a_3 + a_w$	$a_3 - a_w$
A_1A_2	A_1A_1	a_2	$-a_w/2$	$a_2 - (a_w/2)$	$a_2 + (a_w/2)$
A_1A_2	A_1A_2	a_4	0	a_4	a_4
A_1A_2	A_2A_2	a_5	$(a_w/2)$	$a_5 + (a_w/2)$	$a_5 - (a_w/2)$
A_2A_2	A_1A_1	a_3	$-a_w$	$a_3 - a_w$	$a_3 + a_w$
A_2A_2	A_1A_2	a_5	$-a_w/2$	$a_5 - (a_w/2)$	$a_5 + (a_w/2)$
A_2A_2	A_2A_2	0	0	0	0

genetic and environmental variance contributing to sibling differences. The $\hat{\pi}_i$ are the estimated proportions of alleles shared identical by descent (IBD) by sib pair i, which may be computed using software such as Mapmaker/Sibs or Genehunter2 (Kruglyak and Lander, 1995). Again, it is important to note that the predicted covariance matrix may be different for each pair of siblings in the sample. This test is known as the variance component test of linkage (Amos, 1994; Elston, 1998).

A test of IBD linkage is given by computing T between a model with σ_q^2 free and one where $\sigma_q^2 = 0$. However, if the variance component σ_q^2 has a lower bound of zero, the test statistic is asymptotically distributed as a 50:50 mixture of χ^2 of 0 and χ^2 with 1 degree of freedom. This provides a test of linkage between the putative QTL and the trait. Its distribution is a mixture because if there is no QTL effect the similarity of IBD 2 pairs has a 50% chance of being less than that of the IBD 0 pairs.

The key aspect of both the $\hat{\pi}$ and the mixture tests of IBD linkage is that they *do not depend on linkage disequilibrium* between the marker and the disease locus. They only require that the recombination fraction θ is less than one-half.

Joint Test of IBD Linkage and Association

Finally, we note that it is possible to conduct the IBD linkage and the within-family association tests simultaneously. However, in the application reported here, we focus on the association component and simply parameterize the sibling covariance matrix with a lower triangular decomposition:

$$\sum = \mathbf{LL}' \tag{6}$$

which does not vary between pairs. Since the test of phenotypic mean differences between genotypes is two-tailed (i.e., differences in any direction are relevant), the difference between twice log-likelihoods is not a mixture distribution but can be interpreted directly as a χ^2 statistic. By omitting this component the test focuses on association due to linkage disequilibrium and does not include the effects of cotransmitted loci that are not in linkage disequilibrium.

APPLICATION TO ALCOHOL CONSUMPTION AND *ADH2*

Alcohol Consumption and *ADH2*

The biochemical and molecular basis of the polymorphism at the alcohol dehydrogenase 2 (*ADH2*) locus is relatively well characterized. The enzymes encoded by the *ADH2*1* and the *ADH2*2* alleles differ by a single amino acid (Hurley *et al.*, 1994). These two enzymes differ in the rate at which they catalyze the breakdown of alcohol into acetaldehyde (Yin *et al.*, 1984), with *ADH2*2* acting more rapidly. There are striking ethnic differences in the frequency of the *ADH2*2* allele; for example, in Europeans the *ADH2*2* allele is about .05, compared to approximately .60 in Asians (Goedde *et al.*, 1992). Alcohol dehydrogenase carries out the first step in the metabolic pathway of alcohol. The second step is carried out by the enzyme aldehyde dehydrogenase, ALDH2, which is associated with the flushing response to alcohol, which is also relatively frequent in Asian populations. Only variation in ADH2 is considered here.

A number of studies have reported associations between alleles at the *ADH2* locus and alcohol consumption or dependence (Higuchi *et al.*, 1994; Maezawa *et al.*, 1995; Nakamura *et al.*, Matsuo *et al.*, 1995; Neumark *et al.*, 1998; Thomasson *et al.*, 1991; Whitfield, 1997; Whitfield *et al.*, 1998). In both Asians and Europeans, an increased *ADH2*2* allele copy number is associated with lower alcohol intake and dependency. While the effect of the ADH2 locus on the enzyme activity is most likely causal, this association between alcohol consumption and the ADH2 genotype is not well understood. It could be a genuine allelic effect of the *ADH2* locus, or linkage disequilibrium with a nearby locus, or it might be due to population stratification.

Most present-day populations contain some racial admixture. Given the differences in allele frequency between populations, there is a prima facie case for controlling for the effects of population stratification when

examining *ADH2* and alcohol consumption. Even in samples consisting of, e.g., Caucasians only, admixture should still be controlled because population stratification could arise from other sources. Therefore, data on alcohol consumption and *ADH2* genotypes are ideal to illustrate the use of Mx to partition population stratification effects from those of allelic effects or linkage disequilibrium. This is especially important given the previous findings of Whitfield *et al.* (1998), using a subset of the data used here, that showed an intrapair but not a between-pair effect of *ADH2*2* reducing alcohol consumption.

Subjects and Methods

The ADH2 polymorphism was typed using the methods of Xu *et al.* (1988), in Australian dizygotic (*DZ*) twins who took part in the Alcohol Challenge Twin Survey (ACTS) of 1979–1981 (Martin *et al.*, 1985). The twins were adult volunteers but no other inclusionary or exclusionary criteria were applied. In total, the ACTS study had 118 DZ pairs comprising 37 male, 42 female, and 39 of opposite sex. Subjects were followed up in 1990–1992 and DNA samples were obtained from 77 of the DZ pairs who completed a questionnaire on alcohol consumption. At that time their average age was 34 years (range, 29–45 years). Seven of the 77 pairs were discordant for *ADH2* genotype, one being *ADH2*11* and the other *ADH2*12*. There were also 3 pairs concordant for *ADH2*12* and the remaining 67 pairs were both *ADH2*11*. In addition to these pairs there were 20 on whom data on alcohol consumption or genotype were missing for one member of the pair. These data were included in the analysis to help reduce selection biases (Little and Rubin, 1987; Neale and Eaves, 1993).

Alcohol consumption was estimated as the product of the weekly frequency of drinking and the number of drinks on each occasion and is log transformed [$\log(x + 1)$] for analysis [variable FC-2 of Whitfield *et al.* (1998)]. A plot of the data for *ADH2*11* and *ADH2*12* individuals is shown in Fig. 1. Visual inspection indicates a relatively normal-shaped unimodal distribution for both *ADH2*11* and *ADH2*12* groups, with a somewhat lower mean for the *ADH2*12* group.

Basic summary statistics for subjects with the two genotypes are shown in Table IV. They confirm the impressions gleaned from visual inspection of the data and indicate little skewness or kurtosis for the transformed alcohol consumption measure.

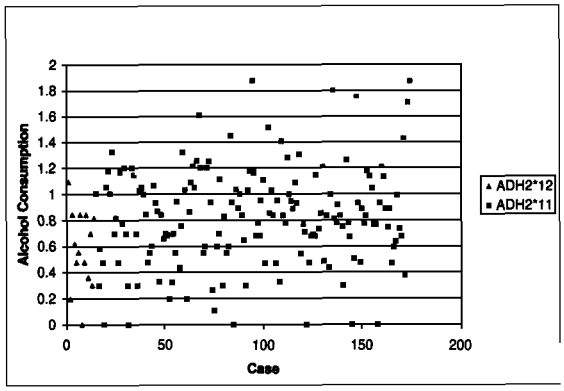

Fig. 1. Plot of total alcohol consumption [$\log_1 0(x + 1)$, where x is number of drinks per week] of subjects with *ADH2*12* (cases 1–14) and *ADH2*11* (cases 15–174) genotypes.

Statistical Model

Since DZ cotwins share, on average, 50% of their genes and some environmental factors, they cannot be considered statistically independent observations. A convenient way to take into account this dependence is by maximum-likelihood estimation using the program Mx (Neale, 1997). The raw observations of phenotypes, together with the covariate sex and *ADH2* genotype, are analyzed jointly. Parameters that influence the means are estimated simultaneously with parameters that predict the covariance matrix of the twins' alcohol consumption scores. In principle, we could estimate the probability that siblings share zero, one, or two alleles IBD at the *ADH2* locus by using the *ADH2* genotypes and other available flanking markers and using the joint test described above. However, because our focus is genetic association, here we make no attempt to model the IBD linkage component of the covariance matrix.

Results

We begin by fitting a six-parameter model involving three parameters for the 2×2 symmetric covariance

Table IV. Basic Summary Statistics for Alcohol Consumption of Subjects with *ADH2*11* or *ADH2*12* Genotypes

Genotype	*N*	Mean	SD	Skewness	Kurtosis
*ADH2*11*	160	.832	.378	.139	.421
*ADH2*12*	14	.581	.301	.266	−.394

matrix Σ, a constant term (which is the expected value for females who are *ADH2*11*), a sex effect (the male deviation), and a deviation for the *ADH2*2* allele. Results of fitting this model (Model 1) and two submodels (Models 2 and 3) are listed in Table V. Twice the difference in log-likelihood between a model in which a parameter is allowed to vary and one in which it is fixed as zero is given as T, which can be used to assess statistical significance. The fifth column in Table V shows T for various pairs of models. Comparing Models 1 and 2, the sex deviation appears highly significant (T_1 = 11.53). Parameter estimates reveal that males have a consumption score about 0.198 log unit (1.6 drinks per week) higher than that of females. The allelic deviation parameter estimate is also highly significant; Model 3 shows a substantial loss of fit compared to Model 1 ($T_1 = 8.38, p = 0.006$). The direction of the effect is such that the *ADH2*2* allele is associated with a decrease in alcohol consumption of 0.314 log unit (i.e., about 2 drinks per week).

To assess the relative impact of stratification and association, the model was parameterized to estimate the allelic deviations corresponding to a_b and a_w in place of a single deviation (Table V). The maximum-likelihood parameter estimates under this model (Model 4) were $a_b = 0.237$ and $a_w = 0.315$, respectively. To test whether the allelic deviations were due to population stratification, the parameter a_w was fixed at zero (Model 5). Doing so led to a significant loss of fit (T_1 = 4.59, p = .032), which suggests that there is a genuine association between *ADH2* polymorphism and alcohol intake or consumption, which is not entirely explained by population stratification.

For a more robust test of association, the constraints on twin pair family means were relaxed and the means of each genotypically distinct type of family pair were estimated as five parameters (Model 6). This model fits only very slightly better than Model 4, which indicates that the hypothesis of only additive effects of the locus is not rejected by the data. The significant increase in the χ^2 value (Model 7) when a_w was dropped from this model (χ^2_1 = 4.61, p = .032) was almost exactly the same as when additivity was assumed. Both results imply that the association between *ADH2* polymorphism and alcohol intake or consumption is not due entirely to population stratification. The association could be due either to an allelic effect of the locus on alcohol consumption, or to linkage disequilibrium between the *ADH2* polymorphism and a closely linked functional variant of significant effect.

Power calculations for this study and for a variety of types of genetic architecture are given in Appendix 2. These power calculations reveal a number of interesting features about the power of this method and, in general, indicate that the results found here are plausible for a data set of this size and the relative infrequency of the *ADH2*2* allele.

DISCUSSION

The joint test of linkage and association suggested by Fulker *et al.* (1999) is a valuable addition to the statistical tools for the analysis of complex traits. Separation of the effects of simple association and linkage into within- and between-sib pair components provides methods analogous to the sib pair TDT test used for

Table V. Results of Modeling Allelic Effects at the *ADH2* Locus on Alcohol Consumption in 77 DZ Twin Pairs and 20 DZ Twin Pairs with Missing Alcohol Data in Terms of Either Global Allelic Deviations or Between- and Within-Pair Allelic Effects (a_b, a_w)

Model	Description	−2LL	df	T^a	df	vs	Model
1	Global deviation	135.40	944	—	—		—
2	Drop sex deviation	146.93	945	11.53	1		1
3	Drop allelic deviation	143.76	945	8.36	1		1
4	Full ($a_b + a_w$)	135.15	943	—	—		—
5	Drop a_w	139.74	944	4.59	1		4
6	Redefine a_b as 5 deviations	133.99	939	1.16[b]	4		4
7	Drop a_w	138.60	938	4.61	1		6

[a] The likelihood-ratio χ^2 statistic for the comparison between models.
[b] The difference statistic is computed as Model 4 χ^2 − Model 6 χ^2, as Model 6 has more parameters.

qualitative traits (Spielman and Ewens, 1996; Ewens and Spielman, 1998). This development is especially valuable because tests based on continuous variables are generally more powerful than those based on ordinal or binary measures. Many major public health problems such as depression, substance abuse, and cardiovascular disease can be assessed with continuous variables that index liability. These index variables would prove highly suitable for this type of analysis (Neale, 1999).

Modeling sibling pair resemblance as a function of $\hat{\pi}$, as in Eq. (5), is a simple approach which can be generalized to larger sibships very easily. This method can be extended to other types of relative such as avuncular or cousin pairs, identical twins, more distant relatives, or adoptees, but there are limitations. In reality, sibling pairs consist of three distinct types of pair according to whether they share zero, one, or two alleles IBD at the locus of interest. A finite mixture distribution model (Everitt, 1981; Neale, 1999) is more appropriate for this situation. This theoretically superior approach has the disadvantage that it is computationally intensive for large sibships.

A fundamental assumption of the method used here is that the data are multivariate normal, conditional on the genotypes of the sib pairs and on any covariates. Statistical tests based on means are often robust to violations of normality, whereas tests on variances are less so (Box, 1953). Therefore we would expect violations of conditional normality to be have less impact on type I error rates for tests of association than for tests of IBD linkage or for the joint test. Allison *et al.* (1999) investigated the robustness of variance-components tests of IBD linkage quite thoroughly and found that it is robust to some types of nonnormality and not to others. Kurtosis may have more adverse effects than skewness, so simple visual inspection may not be sufficient when IBD linkage or joint tests are being conducted.

Implementation of the Fulker *et al.* method in the Mx (Neale, 1997) statistical modeling package has several advantages. First, the program is freely available on the Internet, (http://views.vcu.edu/mx). Second, no programming is required of the user because the script language is relatively straightforward and is well known to many users in the fields of genetic epidemiology and structural equation modeling. Third, the program is flexible enough to allow for the joint modeling of the effects of covariates, such as age and sex. In the example used here, the effect of age was not significant and was therefore not included in the final model. Modeling of simple covariates such as age is merely a starting point for more complex and potentially more powerful multivariate analyses. In many domains it may be prudent to attempt to find association between allelic variants at a candidate locus and the best theoretical measure of liability as assessed by multiple traits. In this case, modeling allelic effects on the means of a latent factor would be appropriate. The structural equation modeling features of Mx make it ideal for extensions of this type.

The treatment of association used here involves a diallelic locus, which is a limitation because most loci exhibit more than two variants. In some cases it may be possible to recode alleles into two classes so that the two allele method can be used. Other cases will require more complex modeling of within pair differences and may yield problems with multiple testing or small sample sizes of genotypes. Occasionally, the investigator may be able to organize the polymorphisms at a locus, e.g., by number of CAG repeats. Zhu *et al.* (1999) used this approach in a test of association between the length of the D9S942 polymorphism and a measure of total body mole count. Size ordering would lend itself to dichotomizing alleles into two categories but would typically incur loss of statistical power. A more general treatment for the multiple allele case is under development.

APPENDIX 1: Mx SCRIPT FOR SIMULATED DATA

The main idea behind the model is that sibling means are to be modeled as a function of the pair mean parameter a_b and the pair difference parameter a_w. The particular function of these parameters differs according to the type of the twin pair (A_1A_1,A_1A_1; A_1A_1, A_1A_2, etc.) A preparatory step in the analysis is therefore to classify the sibling pairs into the nine possible types shown in Table II. These types form one of three variables in the raw data file, which has the following structure:

Alci	Alc2	Type
2.32	1.45	2
−1.50	0.76	8
1.27	1.98	5
etc.		

The Type variable will be used to compute the predicted mean of the sibling pair; it is not going to be analyzed as an observed variable. The **definition** statement in the script accomplishes this step. It is then necessary to set up matrices containing the coefficients of the a_b and a_w parameters for the nine sibling types. These are set in Matrices **V** and **X**. The matrix product

[V|X]W yields a 9 × 1 vector of the predicted means for Sibling 1, and [V: −X]W yields the corresponding vector for Sibling 2. Concatenating these matrices horizontally yields a 9 × 2 matrix of predicted means for both siblings, with the mean vector for a sibling pair of type *i* in row *i* of this matrix. Planting the Type variable in matrix **I** and forming the matrix **E** containing [*i*1*i*2], we can use the **\part** function in Mx to obtain the appropriate submatrix of predicted means for the sibling-pair type in question.

The example in this paper focuses on the modeling of the means of the sibling pairs and leaves the covariances to be estimated freely. Maximum-likelihood estimation is more robust if the predicted covariance matrix is constrained to be positive definite, so we parameterize the covariance matrix with a lower triangular decomposition $\Sigma = \mathbf{A} * \mathbf{A}'$.

```
! Mx script for association test in the presence of possible
! stratification.
! Comments begin with !

#ngroups 1

! Begin the job and data group

Simulated data, real allele effect
  Data NInput=3                  ! 3 input variables
  Rectangular File=realpub2.rec ! read data from file
  Labels p1 p2 type

! type is coded: 1: AAAA 2: AAAa 3: AAaa
!               4: AaAA 5: AaAa 6: Aaaa
!               7: aaAA 8: aaAa 9: aaaa

Definition_variables type;
  ! only p1 and p2 will be analyzed, type defines the model

Begin Matrices;

  A Lower 2 2 Free   ! Residual Covariance, Cholesky

  I Full 1 1                    ! To store Type
  J Unit 1 1                    ! 1
  K Full 1 1                    ! 2
  U Unit 1 2                    ! 1 1

  N Full 1 1 Free  ! Grand mean parameter
  W Full 2 1 Free               ! ab and aw parameters
  V Full 9 1                    ! between pair coefficients
  X Full 9 1                    ! within pair coefficients

End Matrices;

Matrix A 1 0 1
Matrix I 1
Matrix K 2
Matrix V 1 .5 0 .5 0 –.5 0 –.5 –1
Matrix X 0 .5 1  .5 0 .5 –1 –.5 0

Specify I type
Bound –5 5 N 1 1

Begin Algebra;

  E = I|J|I|K;                   ! to select row of V|X
  Z = \part((V|X),E)*W | \part((V|–X),E)*W;
                                 ! selects row of V|X and V|–X

End Algebra;

Means (N|N) + Z;     ! Grand mean in N plus deviations in Z

Covariance A*A';

  Option nd = 4       ! request 4 decimal places in output
  Option RS Multiple ! request residuals, multiple fit
  Option issat ! this is saturated model for submodel comparison

End

  Save simrealpub.mxs
  Drop w 1 2 1        ! fit model without a_b

End

  Drop w 1 1 1        ! fit model without a_b and without a_w

End

  Free w 1 2 1        ! fit model with only a_b

End
```

The Mx script that fits the model of age and sex effects to ADH2 data is somewhat more complicated due to the additional definition variables and matrices with free parameters used to model the age and genotype effects. A script for this model is available at the Mx website: http://views.vcu.edu/mx.

APPENDIX 2: POWER CALCULATIONS FOR THE FULKER METHOD

The power of the association test is quite striking, compared to tests of IBD linkage or for heritability, both of which involve second-order statistics. Here we consider a few simple cases to illustrate the power of the method. Let the frequencies of allele A_1 be p and of allele A_2 be q. Under Hardy–Weinberg equilibrium (i.e., assuming assortative mating, migration, mutation, and selection to be absent), the contribution of additive effects of this locus to the variance is

$$V_A = 2pq(a + d(q - p))^2$$

and the contribution of dominance effects is

$$V_D = 4p^2q^2d^2$$

where a is half the distance between the homozygote means and d is the deviation of the heterozygote mean from the midpoint of the homozygotes (Mather and

Jinks, 1982). Table VI shows the variance components for various values of p, q, d and a. The effect sizes are thus $2a$ for the additive component and d for the dominance component. By choosing values of 0, .2, and .5 we have selected effect sizes that are deemed zero, between small and moderate, and between moderate and large in Cohen's (1969) terminology.

Several features are worthy of note. First, an additive deviation generates more variance than a dominance deviation, all other things being equal. Second, comparing the line for DZ and DZ + MZ (monozygotic), when there is no dominance MZ twins are uninformative for the a_w test, resulting in double the required sample size. However, dominance variation makes both MZ and DZ twins informative for both the within- and the between-family components, although MZ twins are much less informative than DZ. This result may be an artifact of testing only for additive effects of the locus. For the test of both components, MZ and DZ twins are both informative, although required sample sizes increase by approximately 10%. Third, in almost all cases, the B + W test is much less powerful than the W test. The exception is the case of extreme overdominance where the additive deviation is zero, in which case there is no information from the between-family component, as the sample size for both tests is equal (e.g., 446 for both the W and the B + W tests when $d = .2$ and $a = .0$). Fourth, unequal allele frequencies lead to a loss of power in all cases, with no dominance being most adversely affected. Fifth, and most pertinent to the results presented for *ADH2* in this paper, 103 pairs would be required to detect allelic effects of moderate to large size (a = .5) even when allele frequencies are unequal. This is more than the number available in this study, indicating that the effect reported for *ADH2* is large. Indeed this is so, as the variance of alcohol consumption is estimated to be approximately .38, and the estimate of $a_b = a_w = .25$, which is approximately .66 standard deviation, or a Cohen difference of 1.32 standard deviations—a large effect. The allele frequency of *ADH2*1* is approximately .95, so the results in Table V are consistent with the power calculations.

Table VI. Number of Pairs Required to Achieve 80% Power to Reject at the .05 Level the False Hypothesis of No Allele Effect[a]

Allele frequency, $p = .5$								
Dominance deviation	$d = .0$	$d = .0$	$d = .2$	$d = .2$	$d = .2$	$d = .5$	$d = .5$	$d = .5$
Additive deviation	$a = .2$	$a = .5$	$a = .0$	$a = .2$	$a = .5$	$a = .0$	$a = .2$	$a = .5$
V_A	.0200	.1250	.0000	.0200	.1250	.0000	.0200	.1250
V_D	.0000	.0000	.0100	.0100	.0100	.0625	.0625	.0625
DZ (S)	567	103	392	239	86	68	58	47
DZ (NS)	218	41	392	146	39	68	42	30
DZ + MZ (S)	1130	203	446	326	145	77	66	61
DZ + MZ (NS)	240	45	446	162	43	77	44	34
Allele frequency, $p = .1$								
Dominance deviation	$d = .0$	$d = .0$	$d = .2$	$d = .2$	$d = .2$	$d = .5$	$d = .5$	$d = .5$
Additive deviation	$a = .2$	$a = .5$	$a = .0$	$a = .2$	$a = .5$	$a = .0$	$a = .2$	$a = .5$
V_A	.0072	.0450	.0046	.0233	.0784	.0288	.0648	.1458
V_D	.0000	.0000	.0013	.0013	.0013	.0081	.0081	.0081
DZ (S)	1559	273	1354	433	160	240	77	88
DZ (NS)	601	109	712	183	65	125	47	39
DZ + MZ (S)	3114	543	2052	783	305	378	93	159
DZ + MZ (NS)	661	120	788	201	72	139	50	43

[a] The first two rows show the additive (V_A) and dominance (V_D) variance generated by a diallelic locus with allele frequencies p and $q = 1 - p$ and additive and dominance deviations of a and d, respectively. DZ indicates a design of DZ pairs (or siblings) only; DZ + MZ represents a design with 50% MZ and 50% DZ pairs. (S) is the test for the within-family component only (controlling for stratification); (NS) is the less robust test for both between- and within-family effects. Residual background factors generate an additional variance of 1.0 and a covariance of .3 between the members of a pair, regardless of zygosity.

A script to compute the power to detect association with the Fulker method is available at the Mx website: http://views.vcu.edu/mx/examples/abaw.

ACKNOWLEDGMENTS

This work was supported by NHMRC (Australia) grants to N. G. M. and J. B. W.; NIAAA Grants AA07535 AA10249, and AA07728 to A. C. H.; NIDA Grant DA–11015 to D. W. F. and S. S. C.; EY–12562 and MRC (England) grants to D. W. F., P. C. S., and S. S. C.; NIMH Grant MH-01458; and a grant from Gemini Holdings PLC, Cambridge UK, to M. C. N. The authors are grateful for comments from two anonymous reviewers and the editor M. Miller which greatly improved this paper.

REFERENCES

Allison, D. (1997). Transmission-disequilibrium tests for quantitative traits. *Am. J. Hum. Genet.* **60**:676–690.

Allison, D., Neale, M., Zannolli, R., Schork, N., Amos, C., and Blangero, J. (1999). Testing the robustness of a variance-component quantitative trait loci (QTL) mapping procedure to non-normality. *Am. J. Hum. Genet.* **61**:531–544.

Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**: 535–543.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrica* **40**:318–335.

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences,* Academic Press, New York.

Edwards, A. W. F. (1972). *Likelihood,* Cambridge, London.

Elston, R. C. (1998). Linkage and association. *Genet. Epidemiol.* **15**:565–576.

Everitt, B. S. (1981). A monte carlo investigation of the likelihood ratio test for the number of components in a mixture of two normal distributions. *Multivar. Behav. Res.* **16**:171–180.

Ewens, W. J., and Spielman, R. S. (1998). A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test. *Am. J. Hum. Genet.* **62**:450–458.

Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Cambr. Phil. Soc.* **22**:700–725.

Fulker, D. W., Cherny, S. S., Sham, P. C., and Hewitt, J. K. (1999). Combined linkage and association sib pair analysis for quantitative traits. *Am. J. Hum. Genet.* **64**:259–267.

Georges, M., and Cockett, N. (1996). The ovine callipyge locus: A paradigm illustrating the importance of non-mendelian genetics in livestock. *Reprod. Nutr. Dev.* **36**(6):651–657.

Goedde, H. W., Agarwal, D. P., Fritze, D., Meier-Takmann, D., Singh, S., Beckmann, G., Bhatia, K., *et al.* (1992). Distribution of ADH2 and ALDH2 genotypes in different populations.. *Hum. Genet.* **88**:344–346.

Higuchi, S., Matsushita, S., Imazeki, H., Kinoshita, T., Takagi, S., and Kono, H. (1994). Aldehyde dehydrogenase genotypes in japanese alcoholic. *Lancet* **343**:741–742.

Hurley, T. D., Bosron, W. F., and Stone, C. L. (1994). Structures of three human β-alcohol dehydrogenase variants, correlations with functional differences. *J. Mol. Biol.* **239**:415–429.

Kruglyak, L., and Lander, E. S. (1995). Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57**:439–454.

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

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data;* Wiley, New York.

Maezawa, Y., Yamauchi, M., Toda, G., Suzuki, H., and Sakurai, S. (1995). Alcohol-metabolizing enzyme polymorphisms and alcoholism in japan. *Alcohol. Clin. Exp. Res.* **19**:951–954.

Martin, N. G., Perl, J., Oakeshott, J. G., Gibson, J., Starmer, G. A., and Wilks, A. (1985). A twin study of ethanol metabolism. *Behav. Genet.* **15**:93–109.

Mather, K., and Jinks, J. L. (1982). *Biometrical Genetics,* 3rd ed., Chapman and Hall, London.

Nakamura, K., Suwaki, H., Matsuo, Y., Ichikawa, Y., Miyatake, R., and Iwahashi, K. (1995). Association between alcoholics and the genotypes of ALDH2, ADH2, ADH3 as well as p-4502e1. *Arukoru Kenkyuto Yakubutsu Ison* **30**:33–42.

Neale, M. C. (1997). *Mx: Statistical Modeling,* 4th ed.

Neale, M. C. (1999). QTL mapping with sib-pairs: The flexibility of Mx. In Spector, T., Snieder, H., and MacGregor, A. (eds.), *Advances in Twin and Sib Pair Analysis,* Greenwich Medical Media, London (in press).

Neale, M. C., and Eaves, L. J. (1993). Estimating and controlling for the effects of volunteer bias with pairs of relatives. *Behav. Genet.* **23**:271–277.

Neale, M. C., and Miller, M. M. (1997). The use of likelihood-based confidence intervals in genetic models. *Behav. Genet.* **27**: 113–120.

Neumark, Y. D., Friedlander, Y., Thomasson, H. R., and Li, T. K. (1998). Association of the adh2*2 allele with reduced ethanol consumption in Jewish men in Israel: A pilot study. *J. Stud. Alcohol* **59**:133–139.

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273**:1516–1517.

Risch, N., and Teng, J. (1998). The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. I DNA pooling. *Genome Res.* **8**:1273–1288.

Sham; P. C. (1997). Transmission/disequilibrium tests for multiallelic loci. *Am. J. Hum. Genet.* **61**:774–778.

Spielman, R., and Ewens, W. (1996). The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**:983–989.

Thomasson, H. R., Edenberg, H. J., Crabb, D. W., Mai, X. L., Jerome, R. E., Li, T. K., Wang, S. P., Lin, Y. T., Lu, R. B., and Yin, S. J. (1991). Alcohol and aldehyde dehydrogenase genotypes and alcoholism in chinese men. *Am. J. Hum. Genet.* **8**:677–681.

Whitfield, J. B. (1997). Meta-analysis of the effects of alcohol dehydrogenase genotype on alcohol dependence and alcoholic liver disease. *Alcohol Alcohol.* **32**: 613–619.

Whitfield, J. B., Nightingale, B. N., Bucholz, K. K., Madden, P. A. F., Heath, A. C., and Martin, N. G. (1998). ADH genotypes and alcohol use and dependence in europeans. *Alcohol. Clin. Exp. Res.* **22**:1463–1469.

Xu, Y., Carr, L. G., Bosron, W. F., Li, T.-K., and Edenberg, H. J. (1988). Genotyping of human alcohol dehydrogenases at the ADH2 and ADH3 loci following DNA sequence amplification. *Genomics* **2**:209–214.

Yin, S., Bosron, W. F., Magnes, L. J., and Li, T. (1984). Human liver alcohol dehydrogenase: Purification and kinetic characteristion of the b2b2, b2 b1 ab2 and b2c oriental isoenzymes. *Biochemistry* **23**:5847–5853.

Zhu, G., Duffy, D. L., Eldridge, A., Grace, M., Mayne, C., O'Gorman, L., Aitken, J. F., Neale, M. C., Hayward, N. K., Green, A. C., and Martin, N. G. (1999). Mole density is linked to the familial melanoma gene CDKN2A. *Am. J. Hum. Genet.* **65**:483–492.

Edited by Mike Miller