Am. J. Hum. Genet. 64:259–267, 1999

Combined Linkage and Association Sib-Pair Analysis for Quantitative Traits

D. W. Fulker,[1,2] S. S. Cherny,[1,2] P. C. Sham,[2] and J. K. Hewitt[1]

[1]Institute for Behavioral Genetics, University of Colorado, Boulder; and [2]Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, University of London, London

Summary

An extension to current maximum-likelihood variance-components procedures for mapping quantitative-trait loci in sib pairs that allows a simultaneous test of allelic association is proposed. The method involves modeling of the allelic means for a test of association, with simultaneous modeling of the sib-pair covariance structure for a test of linkage. By partitioning of the mean effect of a locus into between- and within-sibship components, the method controls for spurious associations due to population stratification and admixture. The power and efficacy of the method are illustrated through simulation of various models of both real and spurious association.

Introduction

There is an urgent need for the development of methods for the analysis of the genetic architecture of complex traits. In particular, methodological developments are needed for those aspects of risk that are quantitative rather than qualitative in nature. Recently, several such methods have been developed, for sib-pair linkage analysis, that exploit the power of procedures involving maximum likelihood, variance-components analysis, and statistical selection (Amos 1994; Xu and Atchley 1995; Fulker and Cherny 1996; Almasy and Blangero 1998), although comparable methodological advances have been made in the area of association or disequilibrium mapping only for the case of discontinuous, qualitative traits (Spielman et al. 1993; Curtis and Sham 1995; Sham and Curtis 1995*a,* 1995*b;* Curtis 1997; Sham 1997; Boehnke and Langefeld 1998; Lazzeroni and Lange 1998). In part, this emphasis on qualitative traits has been due to their perceived importance within the framework of clinical diagnosis. However, there is increasing recognition that for many traits of clinical interest, such as alcoholism, depression, diabetes, obesity, or hypertension, quantitative phenotypes may be more informative than diagnostic categories for genetic analysis.

Most notable of the various methodological advances made in the area of association or disequilibrium mapping for qualitative traits are those techniques based on the use of parental control groups, such as the transmission/disequilibrium test (TDT; Spielman et al. 1993), the haplotype–relative risk approach (Terwilliger and Ott 1992), and, more recently, the development of similar procedures that use siblings (Boehnke and Langefeld 1998; Spielman and Ewens 1998). What is lacking is an attempt to exploit fully the available statistical procedures for the analysis of continuous traits, involving maximum likelihood and statistical selection, which have proved to be so effective in sib-pair and related linkage procedures. Some progress in this direction has recently been made by Allison (1997) and by Rabinowitz (1997), who both extended the TDT to deal with continuous, quantitative measures, although these approaches are not based on a unified statistical model that can be readily integrated with linkage tests for quantitative-trait loci (QTLs).

In this article, we present a systematic approach to the use of sib pairs for the simultaneous analysis of both association and linkage for quantitative traits. The novel joint analysis of both linkage and association for continuous traits that we describe is made possible by a statistical approach unified by the use of maximum likelihood and of a common biometrical model for the simultaneous analysis of means and covariance matrices. The utility of sib pairs, for QTL linkage analysis, is well established and is based on the use of identity (identical)–by-descent (IBD) relationships among genotypes (Haseman and Elston 1972). The use of sib pairs for association analysis and disequilibrium mapping allows for control of proband selection, in a manner analogous to the use of parents in the TDT (Spielman et al. 1993). The use of siblings has the additional practical advantages of (1) avoiding the difficulties of recruiting parents for the study of late-onset conditions and (2) making

Received July 6, 1998; accepted for publication October 21, 1998; electronically published December 18, 1998.

Address for correspondence and reprints: Dr. Stacey S. Cherny, Institute for Behavioral Genetics, Campus Box 447, University of Colorado, Boulder, CO 80309-0447. E-mail: Stacey.Cherny@Colorado.edu

0002-9297/99/6401-0032$02.00

use of phenotypic measures of siblings who are unlikely to be of greatly differing ages.

The Likelihood Approach

The methods that we suggest in this article involve maximum-likelihood modeling of the raw sibship data. In previous work, we (Fulker and Cherny 1996) and others have suggested modeling of the full sibship covariance structure by maximizing the natural log of the following likelihood of the data:

$$L = \prod_{i=1}^{M} (2\pi)^{-k_i/2} |\mathbf{\Sigma}_i|^{-1/2} e^{-1/2[(\mathbf{y}_i-\boldsymbol{\mu}_i)'\mathbf{\Sigma}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i)]} , \qquad (1)$$

with respect to $\mathbf{\Sigma}_i$ and $\boldsymbol{\mu}$, where k is the number of variables (siblings in the single-phenotype case) measured in family i, $\mathbf{\Sigma}_i$ is the expected covariance matrix among siblings in family i, $\mathbf{y}_i$ is a vector of observed scores obtained for siblings in family i, $\boldsymbol{\mu}_i$ is the vector of expected means for family i, and M is the number of families. In this general expression for the likelihood of multivariate normal data, the elements of the expected covariance matrix and/or mean vector can be estimated directly, or, more usefully, these elements can be made a function of theoretical parameters of interest. These theoretical parameters can be tested for statistical significance by fitting the model with the parameter or parameters of interest and obtaining the natural log of the likelihood of the data, $\ln(L_1)$, and by refitting without those parameters and obtaining $\ln(L_0)$, the natural log of the likelihood of the data under the null hypothesis that the particular parameters are zero. In large samples, $2[\ln(L_1) - \ln(L_0)]$ is asymptotically distributed as a χ^2 statistic, with df equal to the number of parameters being tested. This approach is general for the modeling of quantitative phenotypes obtained from sibships or extended families.

In the case of testing for linkage at a particular position on a chromosome, a sib pair's expected covariance matrix, $\mathbf{\Sigma}_i$, would be given by:

$$\mathbf{\Sigma}_i = \begin{pmatrix} \sigma_q^2 + \sigma_c^2 + \sigma_e^2 & \hat{\pi}_i\sigma_q^2 + \sigma_c^2 \\ \hat{\pi}_i\sigma_q^2 + \sigma_c^2 & \sigma_q^2 + \sigma_c^2 + \sigma_e^2 \end{pmatrix} .$$

In this expression, σ_q^2 estimates the variance explained by the putative QTL, σ_c^2 estimates the residual sibling resemblance (which contains both the environmental variance shared by siblings and half the additive polygenic variance), and σ_e^2 estimates the variance not shared by siblings in a family. The approach uses $\hat{\pi}$, the estimated proportion of alleles shared IBD by any pair of siblings, at a particular chromosomal location. Kruglyak and Lander (1995) outline an approach for the direct use of the probabilities that a pair of siblings share 0, 1, or 2 alleles IBD, with the likelihood of the data computed as a weighted sum of the likelihood assuming 0, 1, or 2 alleles shared. Fulker and Cherny (1996) further extended this method to the use of the full sib-pair covariance structure. Use of any of these likelihood functions, which include provisions for the estimation of mean effects, in addition to effects on covariance structure, suggests a straightforward extension for the modeling of the allelic effects of a candidate locus or of a locus suspected to be in disequilibrium with a trait locus.

The Biometrical Model

Consider a putative 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. There are nine possible combinations of sib-pair genotypes (if the order of sibs 1 and 2 is considered; otherwise, only six unique genotypic combinations are possible), which, under random mating, are characterized by pair means and pair differences, as shown in table 1.

For a pair of siblings, we can model the expected mean vector in the likelihood function described above, as a function of an overall mean m, the pair mean s_m, and the pair difference s_d, as follows: $\mu_1 = m + s_m + (s_d/2)$ and $\mu_2 = m + s_m - (s_d/2)$, with the expectations for these sib-pair means and differences obtained from table 1. We then can test association by a 1-df χ^2 test of a. However, this test may be prone to spurious associations, owing to population stratification. Since population stratification will influence pair means but not pair differences, one possible way to avoid spurious associations is to allow the gene effect a to be different for the pair means and pair differences. We denote the gene effect on pair means and pair differences as a_b and a_w, respectively. The model for sib 1 and sib 2 can then be seen in table 2. The information in table 2 provides a

Table 1

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

Genotype		Additive Effect			
Sib 1	Sib 2	Sib 1	Sib 2	Mean	Difference/2
A_1A_1	A_1A_1	a	a	a	0
A_1A_1	A_1A_2	a	0	$a/2$	$a/2$
A_1A_1	A_2A_2	a	$-a$	0	a
A_1A_2	A_1A_1	0	a	$a/2$	$-a/2$
A_1A_2	A_1A_2	0	0	0	0
A_1A_2	A_2A_2	0	$-a$	$-a/2$	$a/2$
A_2A_2	A_1A_1	$-a$	a	0	$-a$
A_2A_2	A_1A_2	$-a$	0	$-a/2$	$-a/2$
A_2A_2	A_2A_2	$-a$	$-a$	$-a$	0

Table 2

Partitioning of Additive Effect into Between- and Within-Pairs Components

Genotype		Additive Effect			
Sib 1	Sib 2	Sib 1	Sib 2	Mean	Difference/2
A_1A_1	A_1A_1	a_b	a_b	a_b	0
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	a_w	$-a_w$	0	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	$-a_w$	a_w	0	$-a_w$
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	$-a_b$	$-a_b$	$-a_b$	0

general expression for the exponential part of the likelihood function in equation (1), as given by

$$-\tfrac{1}{2}(\boldsymbol{y}_i - \boldsymbol{m} - \mathbf{X}_{b_i}\boldsymbol{a}_b - \mathbf{X}_{w_i}\boldsymbol{a}_w) \times \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{m} - \mathbf{X}_{b_i}\boldsymbol{a}_b - \mathbf{X}_{w_i}\boldsymbol{a}_w) \ ,$$

where $\mathbf{X}_{b_i}$ and $\mathbf{X}_{w_i}$ are diagonal matrices for sib pair i specified by the coefficients implied in the third and fourth columns of table 2, $\boldsymbol{a}_b$ and $\boldsymbol{a}_w$ are vectors containing the a_b and a_w parameters, respectively, for each member of a sibship, and $\boldsymbol{m}$ is a vector containing the overall phenotypic mean for each member of a sibship (set equal for all members).

A more robust test for association is obtained by computing the 1-df likelihood ratio χ^2 between a model with a_w free and that with a_w set to 0, while a_b is free in both models. In the case of a locus that is dominant, the above models are readily extended to allow a between- and within-pairs dominance parameter. Furthermore, the modeling of multiple alleles is a straightforward extension, although, with the inclusion of dominance parameters in such a model, the number of df for the test of association increases dramatically as the number of alleles increases.

In the case of population admixture, the model presented in table 2 is not strictly correct. Even under an additive model, the pair (or sibship) means do not follow this simple relationship of additive allelic effects. This problem can be overcome by relaxing the constraint on pair means and modeling the mean of each genotypic combination as a separate parameter (and then dropping the grand mean from the model). However, as we illustrate below, use of the simple additive model does not introduce noticeable bias in parameter estimates for the cases that we have explored.

The Joint Analysis

Because it is based on procedures for estimation of maximum-likelihood variance components, this new test of association includes all the advantages of those procedures. One major advantage is that linkage can (and should) be modeled simultaneously with the association parameters. Linkage is modeled in the covariance structure, as illustrated above, while the association parameters, along with other covariates (if desired), are modeled on the means. All parameters would be estimated as a full model, which would be compared with various submodels, allowing individual tests of association and linkage. A simple test of the within-pairs association parameter would yield a robust test of association while controlling for stratification. Testing linkage while simultaneously modeling association would provide a test of whether the putative QTL locus is a candidate or whether it is merely in disequilibrium with a trait locus. If significant linkage is detected while modeling association, one can conclude that the putative locus is not the functional gene but, rather, is a locus in disequilibrium with a trait locus. The method also provides a test for the presence of stratification. If the a_b parameter is not found to be equal to the a_w parameter, one can conclude that at least part of the association observed is a result of population stratification. However, if a_w alone is still significant, one still can conclude that a true association has been found. Because of the variance-components model-fitting framework in which we have cast this method, many other possible hypotheses also can be tested.

Simulations

Procedures

To illustrate the application and usefulness of the combined approach to the modeling of linkage and association in sibship data, we performed a series of simulations involving a dense map of diallelic markers and a diallelic trait locus. In all cases, we simulated a marker map of four diallelic anonymous markers, with equally frequent alleles, that are equally spaced 2 cM apart, thereby spanning a small chromosomal segment 6 cM

in length. We chose such a dense map because this would be typical of a case that might be constructed to finely map a region that suggests the presence of a QTL. Furthermore, a dense map of diallelic markers that can be genotyped efficiently and for relatively little cost, by use of genotyping chips promised to be available soon, would suggest that the approach we are advocating can be routinely used in all sib-pair genome scans.

All simulations that we present involved a diallelic QTL accounting for 20% of the phenotypic variance ($h_q^2 = .2$) and a sample of 1,000 sib pairs with parental genotypes available. The proportion of total variance shared by siblings (but not including QTL variance), c^2, varied between 0 and .4, and, therefore, the proportion of variance due to random environmental influences, e^2, varied between .8 and .4. For all situations, we simulated 100 samples and present χ^2 statistics that were averaged across all 100 simulations. These average χ^2 statistics, minus their associated df, can be considered noncentrality parameters for use in power calculations. We explored a situation that involved disequilibrium between one of the anonymous markers and the gene influencing the phenotype, with disequilibrium parameters (D) ranging from small (.025) to complete (.25). In all these cases, the trait locus was positioned halfway along the 6-cM chromosomal segment, at 3 cM, and was placed in disequilibrium with the second marker (located at 2 cM), which was 1 cM away from the QTL. In the presence of linkage equilibrium, the probability of allele 1 of a marker coupling with allele 1 of a QTL is .5, in the case for which both the marker and the QTL each have two alleles of equal frequency, which is the basic case from which we were working. In simulating disequilibrium, we altered, from its equilibrium value of .5, the probability of allele 1 of the marker coupling with allele 1 of the QTL. The disequilibrium was generated in the parental genotypes, which subsequently had undergone a single round of random mating to produce the offspring that were used for analysis. We also explored the case in which the QTL genotype itself was available for analysis, which is equivalent to our case of complete disequilibrium, except that the disequilibrium was not diluted by a round of random mating. In the tables presented below, this is referred to as the "candidate-gene case."

The second broad type of situation that we explored was one of initial admixture of two sib-pair populations, each containing 500 pairs. We distorted the QTL allele frequencies from an equal .5:.5 mixture of the two alleles, covering the range .55:.45 to .99:.01 in the first population and using the reverse frequencies for the second population. To create disequilibrium, we distorted, from .5:.5, the allelic frequencies of the second anonymous diallelic marker, in a manner similar to but in the opposite direction of that used for the QTL allele frequencies. We explored all these cases of initial admixture for the case of a linked QTL at position 3 cM on the chromosomal segment and for the case of the QTL being unlinked on another chromosome.

For each of the simulations, we performed a number of statistical tests. We tested for association while simultaneously modeling linkage, by testing the between- and within-pairs components individually and in combination. When modeling linkage, we computed IBD along the chromosome, using the Kruglyak and Lander (1995) multipoint method employed in Mapmaker/SIBS, and then we computed $\hat{\pi}$ from the IBD probabilities and used the variance-components procedure outlined, which uses $\hat{\pi}$. As discussed above, the between-pairs component should appear in both true and spurious cases of association. However, the within-pairs component should be present only in the case of a true association and not in a case due to admixture or stratification. Next, we performed tests of linkage, both while modeling association and while not modeling association. Finally, we performed a combined test of linkage and association.

Results

In tables 3 and 4 we present average χ^2 statistics obtained from the simulations involving the smallest sibling correlation, .1, that we explored and in which there was no additional sibling resemblance beyond that due to the QTL. Table 3 includes results from simulations in which the association simulated was a true association. In the candidate-gene case, there was both a large between-pairs and a large within-pairs association effect, as expected. We note that the between-pairs effect is far more statistically significant and that the χ^2 statistics for the between- and within-pairs components roughly sum to the 2-df χ^2 obtained when the two components were tested simultaneously. As can be seen from the χ^2 obtained when linkage was tested without simultaneous modeling of association, the linkage effect is substantial in this case—again, as expected, since the QTL is one of the markers and the other diallelic markers are quite close, which is very informative for linkage. The test of linkage in the presence of association parameters yielded a χ^2 of ~.5, which is the χ^2 expected under the null hypothesis, when the statistic is, as in this case, a 50:50 mixture of χ^2 and a point mass of 0. The linkage effect drops to 0, since all sibling resemblance or lack of resemblance can be accounted for by the within-pairs–component mean model, leaving no covariance to be explained by the linkage model.

The results for complete disequilibrium, $D = .25$, were essentially the same as those for the candidate-gene case, as expected under the model. As the amount of simulated disequilibrium diminished, the association χ^2

Table 3

Average χ^2 Statistics Obtained from the Combined Linkage and Association Test, under Conditions of Real Association: No Additional Sibling Resemblance

Test	χ^2 for Candidate-Gene Case	χ^2 for D =				
		.25	.20	.10	.05	.025
Association:						
Between pairs	320.66	313.82	190.28	45.96	12.87	4.29
Within pairs	122.31	115.46	70.31	17.43	5.16	2.18
Total	435.16	421.55	257.93	63.07	17.82	6.29
Linkage:						
Without association	6.02	6.14	5.49	4.86	4.84	4.80
With association	.22	.20	1.05	3.50	4.42	4.65
Linkage and association	440.19	427.23	263.06	67.69	22.42	10.90

Note.—Under all conditions, 100 samples of 1,000 sib pairs were simulated. A locus accounting for 20% of the phenotypic variance was simulated with no additional cause of sibling covariance, yielding a total sibling correlation of .10. Between-pairs association was modeled as a single additive deviation. All the χ^2 statistics have 1 df, except those obtained from testing total association and combined linkage and association, which have 2 df and 3 df, respectively. See text for further details.

statistics decreased. In fact, the χ^2 statistics (minus their associated df) were proportional to D^2, as seen in table 3, since the observed additive effect of the marker is directly proportional to D. In the more general case, when the additive system at the QTL is considered, the average effect (as defined by Falconer [1989]) of the neutral marker A, resulting from its disequilibrium D with functional QTL B, will be equal to aD/D_{max}, where D_{max} (the maximum value of the disequilibrium parameter, for given marker allele frequencies and any QTL allele frequencies) is $p_A q_A$, the product of the allele frequencies of the marker. This result is true irrespective of the allele frequencies at the QTL. Thus, for a diallelic marker with equal allele frequencies, this average effect is $4Da$, as can be seen from the estimates in table 3.

As can be seen in table 3, the test of linkage in the presence of association becomes more powerful as values of D decrease, since less of the sibling resemblance due to the QTL locus is being explained by the association parameters, leaving linkage to account for the remaining resemblance, given that the effect of the locus is, in all the cases, of the same magnitude. In effect, the procedure partials out the effects of association, from the test of linkage.

Table 4 presents results from simulations of spurious association due to admixture of two populations of equal size. In the case of admixture with a linked locus, power to detect the between-pairs–association effect diminished as a function of the magnitude of the allele-frequency difference between the two populations, as expected. However, since the within-pairs component yields the test of association while controlling for admixture and stratification, the expected χ^2 is constant across varying degrees of admixture and near its expected value of unity. We note that, for the case of extreme admixture, in which the frequency of the decreaser allele (q) is equal to .99 for 50% of the total population and $q = .01$ for the other 50% of the population, the average χ^2 is noticeably less than its expected value of unity. This is due to essentially no within-family variability in genotype, since members within a given family are almost always all the same genotype. This results in the within-pairs parameter being empirically underidentified. In practice, however, such an extreme situation can be easily recognized by the marked departure from Hardy-Weinberg equilibrium. In all cases of admixture, the two tests of linkage, one with modeling of association and the other ignoring association, both yield similar χ^2 statistics, since there is no within-pairs component removed from the sib-pair resemblance used to test for linkage.

The second set of simulations in table 4 deal with admixture, as in the previous set, but the QTL is unlinked on a different chromosome. This appears to result in test statistics for the association tests similar to those obtained in the linked case, but slightly smaller, and the within-pairs–component χ^2 is almost exactly its expected value of unity, in all four cases. All the linkage-test statistics are around their expected value of .5.

In all the simulations presented in both tables 3 and 4, the between- and within-pairs components approximately add up to the total-association test statistics, implying that these two components are essentially orthogonal, and the χ^2 obtained from simultaneous testing of linkage and association is approximately the sum of the association χ^2 statistics and the χ^2 from the linkage test in the presence of association, further implying that association and linkage are orthogonal tests when modeled together.

Tables 5 and 6 present the results for a set of simu-

Table 4

Average χ^2 Statistics Obtained from the Combined Linkage and Association Test, under Conditions of Spurious Association and under the Null Hypothesis: No Additional Sibling Resemblance

	χ^2, as a Function of Admixture Ratio								χ^2 under the Null Hypothesis
	Linked Trait Locus				Unlinked Trait Locus				
Test	.99:.01	.7:.3	.6:.4	.55:.45	.99:.01	.7:.3	.6:.4	.55:.45	
Association:									
Between pairs	2,961.20	45.17	4.84	1.52	2,932.20	42.48	3.48	1.07	1.00
Within pairs	.19	1.52	1.50	1.40	.07	.98	.94	.96	.79
Total	2,961.20	46.46	6.09	2.70	2,932.20	43.48	4.42	2.04	1.80
Linkage:									
Without association	3.83	4.80	4.82	4.75	.70	.53	.58	.58	.58
With association	4.57	5.21	4.77	4.68	.84	.60	.57	.57	.57
Linkage and association	2,964.79	51.02	10.66	7.22	2,932.86	44.02	5.00	2.62	2.39

Note.—Under all conditions, 1,000 sib pairs were simulated. A locus accounting for 20% of the phenotypic variance was simulated with no additional cause of sibling covariance, yielding a total sibling correlation of .10. Between-pairs association was modeled as a single additive deviation. All the χ^2 statistics have 1 df, except those obtained from testing total association and combined linkage and association, which have 2 df and 3 df, respectively. The cases of admixture involved the mixture of two populations of equal size: for one population, the ratio of p to q for the two alleles of the diallelic marker locus is given in the column headings, and, for the other population, the ratio is reversed. In these two populations, the trait locus has allele frequencies opposite to those of the marker locus. See text for further details.

lations that are parallel to those in tables 3 and 4, but an additional shared-environment component of covariance was introduced, which accounted for 40% of the total phenotypic variance, resulting in a total sibling correlation of .5. The main difference between the results in tables 5 and 6 and those in tables 3 and 4 is that the discrepancy between the between-pairs association χ^2 and the within-pairs association χ^2 is reduced, to the extent that the between-pairs and within-pairs χ^2 statistics are almost equal across tables 3 and 5. However, the combined between- and within-pairs test yielded a χ^2 similar to that obtained when there was no additional sibling resemblance. In addition, as expected, the tests of linkage were more powerful in the presence of additional sibling resemblance, when this additional shared variance reduced the proportion of random environmental variance in the presence of a constant proportion of QTL variance.

For both sibling correlations, as presented in the last column of tables 4 and 6, we simulated the null hypothesis of a QTL accounting for 20% of the variance, but on a chromosome on which no markers were genotyped. Results indicated that each of the between- and within-pairs components were approximately distributed as a 1-df χ^2; that the test of the between- and within-pairs components together was a 2-df test; that the two types of linkage tests were a 50:50 mixture of χ^2 and a point mass of 0, which yielded an expected χ^2 of .5; and that the simultaneous test of linkage and association resulted in a χ^2 of ~2.5, the sum of the expected χ^2 statistics of each of the individual tests.

Table 7 presents mean estimates of the QTL effect size, a, which, in all cases, was simulated to equal .5. First, estimates of a_b and a_w for the case of $c^2 = .4$ were essentially the same as those for the case of $c^2 = 0$. In the candidate-gene case, estimates of a_b and a_w were both quite near their simulated value of .5. In the case of complete disequilibrium, $D = .25$, a_b was still near .5, whereas a_w was slightly smaller. This is because the disequilibrium was introduced into the parental chromosomes, which subsequently undergo a round of recombination in the formation of gametes. As mentioned above, for cases of incomplete disequilibrium, the effect size a is proportional to the amount of disequilibrium. Our estimates of a_b are, in fact, very close to their expected values. Again, estimates of a_w were slightly smaller, owing to the method by which disequilibrium was introduced. As expected, estimates of a_b decreased from ~.5 for extreme admixture to ~0 when allele frequencies were not very different between the two subpopulations, whereas estimates of a_w were essentially 0 throughout, confirming that the within-pairs component is not present when association is only spurious and due to admixture.

Discussion

The proposed variance-components model offers a method of combined QTL linkage and association analysis for sib-pair data. Most notably, the model partitions association into between- and within-pairs components, and a robust test of association is constructed on the basis of the within-pairs component. In this respect, this test is similar to the TDT (Spielman et al. 1993) for diseases and to recent adaptations of the TDT to quantitative traits (Allison 1997; Rabinowitz 1997). Our new approach, however, is based on a variance-components model, which has many advantages. First, the robust test of association can be combined with variance-components methods of QTL linkage analysis (Fulker and

Table 5

Average χ^2 Statistics Obtained from the Combined Linkage and Association Test, under Conditions of Real Association: Sibling Correlation of .5

Test	χ^2 for Candidate-Gene Case	χ^2 for D =				
		.25	.20	.10	.05	.025
Association:						
Between pairs	232.17	221.79	136.99	34.12	9.95	3.62
Within pairs	213.54	201.42	122.32	29.30	8.07	2.95
Total	443.69	420.78	258.23	63.13	17.71	6.29
Linkage:						
Without association	13.02	13.12	11.79	10.55	10.44	10.42
With association	.51	.54	2.45	7.70	9.60	10.12
Linkage and association	454.77	433.08	269.38	73.31	27.81	16.42

Note.—Under all conditions, 1,000 sib pairs were simulated. A locus accounting for 20% of the phenotypic variance was simulated, with an additional c^2 component of .4. Between-pairs association was modeled as a single additive deviation. All the χ^2 statistics have 1 df, except those obtained from testing total association and combined linkage and association, which have 2 df and 3 df, respectively. See text for further details.

Cherny 1996). This combined test should increase the power of detecting a QTL when the marker locus is not the QTL itself but is in linkage disequilibrium with the QTL. Furthermore, if there is appreciable linkage evidence prior to the incorporation of association, then the extent that this linkage evidence is diminished by the inclusion of association into the model can be used to provide an indication of the strength of the linkage disequilibrium between the QTL and the marker. In effect, modeling of association parameters in the mean structure partials out any variance accounted for by the alleles themselves, from the simultaneous test of linkage in the covariance structure. In cases for which the linkage evidence vanishes entirely, the marker may be the QTL itself or at least may be in very strong linkage disequilibrium with the QTL. This phenomenon was demonstrated clearly in our simulations: as the strength of linkage disequilibrium increased, the evidence of linkage diminished, when association was included in the model.

Another interesting observation from our simulations is that the evidence of linkage increased with an increase in disequilibrium between the QTL and the marker, when association was omitted from the model. Our explanation of this phenomenon is that, when both the QTL and the marker are moderately polymorphic, a greater degree of linkage disequilibrium would increase the probability that a parent is heterozygous at both loci and therefore is informative for linkage. If this explanation is correct, then the effect should diminish as the heterozygosity of the markers increases or, in a multipoint analysis, as the number and density of markers increases. Our simulations have confirmed that this indeed is the case.

Our simulations appear to have demonstrated that a single fixed-effect parameter (a_b) is sufficient for modeling of the between-pairs component, in order for the within-pairs test to be robust to population stratification. Although this could have been due to our choice of admixture parameters for the simulations (i.e., marker allele frequencies and admixture proportion), an additional reason could be the inclusion of a parameter for residual sibling resemblance, σ_c^2, which represents a random effect on sib-pair means. Further work is necessary, to establish whether a_b and σ_c^2 are always sufficient or whether the saturated model for pair means is sometimes necessary for the within-pairs test to be robust to population stratification.

In our simulations, as the residual sibling correlation increased (we also simulated an intermediate sibling correlation, which showed intermediate results and, therefore, was not presented) with a corresponding reduction in random environmental variance, the average χ^2 of the between-pairs component diminished, whereas the average χ^2 of the within-pairs component increased. Residual sibling resemblance introduces variation between sib-pair means, in addition to that introduced by the QTL. Therefore, increasing residual sibling correlation may lead to increased confounding between a_b and σ_c^2, resulting in a reduction of the power to detect the between-pairs component. In our simulations, this was reflected as an increase in the sampling variance of a_b, as the residual sibling correlation increased. On the other hand, as we increased the residual sibling correlation, the nonshared component of variance, σ_e^2, was correspondingly reduced (since total variance was fixed to unity), thereby increasing the power to detect the within-pairs component. In our simulations, this was reflected as a decrease in the sampling variance of a_w, as the residual sibling correlation increased. Overall, a substantial residual sibling correlation is favorable for the robust, within-pairs test of association, since it provides a source of variation that can be explained, thereby reducing error variance.

Consideration of the quantitative, rather than the

Table 6

Average χ^2 Statistics Obtained from the Combined Linkage and Association Test, under Conditions of Spurious Association and under the Null Hypothesis: Sibling Correlation of .5

	χ^2, as a Function of Admixture Ratio								χ^2 under the Null Hypothesis
	Linked Trait Locus				Unlinked Trait Locus				
Test	.99:.01	.7:.3	.6:.4	.55:.45	.99:.01	.7:.3	.6:.4	.55:.45	
Association:									
Between pairs	2,624.55	34.78	4.28	1.68	2,642.19	32.08	2.71	1.01	.94
Within pairs	.62	1.84	1.78	1.62	.02	1.08	.97	1.01	.80
Total	2,624.55	36.14	5.54	2.86	2,642.19	33.14	3.66	2.02	1.78
Linkage:									
Without association	10.44	10.00	10.36	10.40	.86	.56	.57	.54	.63
With association	10.78	10.17	10.27	10.30	.79	.57	.55	.52	.62
Linkage and association	2,634.39	45.66	15.38	12.82	2,643.02	33.70	4.21	2.55	2.44

Note.—Under all conditions, 1,000 sib pairs were simulated. A locus accounting for 20% of the phenotypic variance was simulated, with an additional c^2 component of .4. Between-pairs association was modeled as a single additive deviation. All the χ^2 statistics have 1 df, except those obtained from testing total association and combined linkage and association, which have 2 df and 3 df, respectively. The cases of admixture involved the mixture of two populations of equal size: for one population, the ratio of p to q for the two alleles of the diallelic marker locus is given in the column headings, and, for the other population, the ratio is reversed. In these two populations, the trait locus has allele frequencies opposite to those of the marker locus. See text for further details.

qualitative, effects of association through disequilibrium shows very clearly that effect sizes are always underestimated (within sampling error) and that they systematically depend on the level of disequilibrium. For example, for weak disequilibrium (e.g., $D = .05$, for a diallelic marker with equal allele frequencies) the maximum effect associated with the marker is one-fifth the QTL effect. This implies that effect sizes determined from association studies potentially are much smaller than the actual effect the functional gene may have on the phenotype of interest. For instance, in a recent report of an association between the insulin-like growth factor–2 receptor and cognitive ability, Chorney et al. (1998) estimated that this locus accounts for 2% of the phenotypic variation. However, given that the disequilibrium between the functional gene influencing cognitive ability and the gene with which it is associated is not complete, the variance explained by the as-yet-unknown gene could be substantially larger.

Although we have illustrated the proposed simultaneous test of association and linkage for sibships of size two and a single diallelic candidate or marker locus, the method can be generalized to larger sibships and sibships of variable size within a study and, in addition, to candidate or marker loci with multiple alleles that each convey different quantitative effects on the phenotype or marker loci with multiple alleles that are differentially associated with the alleles at a trait locus. In brief, to accommodate sibships of variable size, the sibship means are modeled as in the sib-pair case for the between-pairs effect, and individual sibling deviations from the sibship mean are modeled as for the within-pairs effect (in place of half the difference score). For multiple alleles and dominance, a between- and within-sibship parameter would be specified for each allele, under the constraint that they sum to zero, and a between- and within-sibship dominance parameter would be specified for each heterozygote type. However, this would result in a test with multiple df, thereby reducing power. Alternatively, if, for example, allele size (i.e., the number of repeats) has a potential linear or quadratic effect on a phenotype, such effects also could be accommodated within the framework we propose, in a powerful manner. However, the statistical properties of such situations should be explored thoroughly, both analytically and via simulation,

Table 7

Average Parameter Estimates Obtained from All Simulations

	$c^2 = 0$		$c^2 = .4$	
	a_b	a_w	a_b	a_w
Candidate-gene case	.504	.499	.506	.499
D =:				
.25	.499	.486	.498	.486
.20	.400	.388	.400	.388
.10	.202	.193	.201	.194
.05	.103	.096	.103	.096
.025	.052	.047	.052	.048
Admixture ratio (linked):				
.99:.01	.490	−.001	.489	−.000
.7:.3	.172	.004	.172	.003
.6:.4	.055	.003	.055	.001
.55:.45	.017	.002	.017	.000
Admixture ratio (unlinked):				
.99:.01	.490	−.002	.490	−.003
.7:.3	.168	.003	.168	.002
.6:.4	.047	.000	.045	−.000
.55:.45	.009	.001	.008	.000
Null hypothesis	.000	−.008	.000	−.007

before they are applied, which is beyond the scope of this article.

In this article, we have discussed the foundations of a variance-components approach to combined linkage and association analysis of quantitative traits. We also have established, by means of simulations, some basic properties of the methodology. The proposed variance-components model provides a flexible and powerful framework for further generalizations and extensions. For example, the model can be generalized to multiple phenotypes and can be made to incorporate measured covariates as well as gene-environment interactions. Furthermore, the analysis of selected samples can be accommodated in the maximum-likelihood framework simply by incorporating an ascertainment correction in the likelihood function or by imposing appropriate constraints on the parameters. These further developments should lead to a set of powerful tools for the detection of QTLs and the dissection of complex quantitative traits in humans.

Acknowledgments

This research was supported, in part, by National Institutes of Health grants AA-07330, AA-10556, DA-11015, EY-12562, MH-43899, MH-53480, and MH-53668 and by a grant from the Medical Research Council of Great Britain.

References

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

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62: 1198–1211

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

Boehnke M, Langefeld CD (1998) Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet 62:950–961

Chorney MJ, Chorney K, Seese N, Owen MJ, Daniels J, McGuffin P, Thompson LA, et al (1998) A quantitative trait locus associated with cognitive ability in children. Psychol Sci 9:159–166

Curtis D (1997) Use of siblings as controls in case-control association studies. Ann Hum Genet 61:319–333

Curtis D, Sham PC (1995) A note on the application of the transmission disequilibrium test when a parent is missing. Am J Hum Genet 56:811–812

Falconer DS (1989) Introduction to quantitative genetics, 3d ed. Longman Scientific & Technical, London

Fulker DW, Cherny SS (1996) An improved multipoint sib-pair analysis of quantitative traits. Behav Genet 26:527–532

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

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

Lazzeroni LC, Lange K (1998) A conditional inference framework for extending the transmission/disequilibrium test. Hum Hered 48:67–81

Rabinowitz D (1997) A transmission disequilibrium test for quantitative trait loci. Hum Hered 47:342–350

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

Sham PC, Curtis D (1995*a*) Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann Hum Genet 59:97–105

——— (1995*b*) An extended transmission/disequilibrium test (TDT) for multiallelic marker loci. Ann Hum Genet 59: 323–336

Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–458

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Terwilliger J, Ott J (1992) A haplotype-based "haplotype relative risk" approach to detecting allelic associations. Hum Hered 42:337–346

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